

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. IPR2025-00768
U.S. Patent No. 10,805,111

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 10,805,111**

TABLE OF CONTENTS

I.	INTRODUCTION -----	1
II.	BACKGROUND AND STATE OF THE ART -----	1
III.	THE '111 PATENT -----	3
	A. Overview -----	3
	B. Claims -----	3
	C. Prosecution -----	5
	D. Priority -----	5
IV.	LEVEL OF ORDINARY SKILL IN THE ART -----	5
V.	CLAIM CONSTRUCTION -----	6
VI.	STATEMENT OF PRECISE RELIEF REQUESTED -----	6
	A. Grounds -----	6
	B. Status of References as Prior Art -----	7
VII.	GROUND 1A: CLAIMS 1-2, 4-5, AND 16 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN AND DUNCAN -----	8
	A. Claim 1 -----	8
	1. Preamble -----	8
	2. 1[a]: Network Accessible Library -----	8
	3. 1[b]: Downloading Image Associated with Time Information -----	9
	4. 1[c]: Assembling a First Page -----	11

5.	1[d]: Assigning Time Information to First Page-----	15
a.	Assembling a Page-----	16
b.	Changing the Presentation Speed -----	18
c.	Translating Synchronization Times into Presentation Times-----	20
6.	1[e]: Downloading Audio-----	21
7.	1[f]: Simultaneously Rendering Images and Audio -----	23
B.	Claim 2-----	24
C.	Claim 4-----	25
D.	Claim 5-----	25
E.	Claim 16 -----	27
VIII.	GROUND 1B: CLAIM 3 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND TEMPLEMAN -----	27
IX.	GROUND 1C: CLAIM 6 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND OPTIONALLY HAY -----	30
X.	GROUND 1D: CLAIM 7 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND OPTIONALLY HAY AND/OR HECKERMAN -----	33
XI.	GROUND 1E: CLAIMS 8-11 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND DTB-----	36
A.	Claim 8-----	36
B.	Claim 9-----	39

C.	Claims 10 and 11 -----	40
XII.	GROUND 1F: CLAIMS 12-15 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND DR. SEUSS -----	43
A.	Claim 12 -----	43
B.	Claim 13 -----	47
1.	13[a]: Rendering Text Passages-----	47
2.	13[b]: Assigning Time Information -----	48
C.	Claim 14 -----	49
D.	Claim 15 -----	50
XIII.	GROUND 2A: CLAIMS 1-2, 4-5, 8-11, AND 16 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, AND DUNCAN -----	51
A.	Claim 1-----	51
1.	Preamble -----	51
2.	1[a]: Network Accessible Library -----	52
3.	1[b]: Downloading Images Associated with Time Information -----	55
4.	1[c]: Assembling a First Page -----	57
5.	1[d]: Assigning Time Information to First Page-----	60
a.	Assigning Time Information by Assembling a Page -----	61
b.	Assigning Time Information by Changing the Presentation Speed-----	62

6.	1[e]: Downloading Audio-----	62
7.	1[f]: Simultaneously Rendering Images and Audio-----	63
B.	Claim 2-----	63
C.	Claim 4-----	64
D.	Claim 5-----	64
E.	Claims 8-11 -----	65
F.	Claim 16 -----	65
XIV.	GROUND 2B: CLAIM 3 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, DUNCAN, AND TEMPLEMAN -----	65
XV.	GROUND 2C: CLAIM 6 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, DUNCAN, AND HAY -----	65
XVI.	GROUND 2D: CLAIM 7 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, DUNCAN, HAY, AND PRABHAKARAN -----	67
XVII.	GROUND 2E: CLAIMS 12-15 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, DUNCAN, AND DR. SEUSS -----	68
XVIII.	SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS -----	69
XIX.	DISCRETIONARY DENIAL UNDER §314(A) IS NOT APPROPRIATE -----	69
A.	Factor 1: Potential Stay -----	69
B.	Factor 2: Proximity of Trial to FWD-----	70
C.	Factor 3: Investment in Parallel Proceeding -----	71

D.	Factor 4: Overlapping Issues-----	72
E.	Factor 5: The Parties -----	72
F.	Factor 6: Other Circumstances-----	72
XX.	DISCRETIONARY DENIAL UNDER §325(D) IS NOT APPROPRIATE -----	73
XXI.	CONCLUSION-----	73
XXII.	MANDATORY NOTICES, GROUNDS FOR STANDING, AND FEE PAYMENT -----	74
A.	Real Parties-In-Interest (37 C.F.R. §42.8(b)(1))-----	74
B.	Related Matters (37 C.F.R. §42.8(b)(2)) -----	74
C.	Lead and Backup Counsel (37 C.F.R. §42.8(b)(3))-----	74
D.	Service Information (37 C.F.R. §42.8(b)(4)) -----	76
E.	Grounds for Standing (37 C.F.R. §42.104(a)) -----	76
F.	Payment of Fees (37 C.F.R. §42.15(a))-----	76

TABLE OF AUTHORITIES

<i>Cases:</i>	<i>Page(s):</i>
<i>Amazon.com, Inc. v. Nokia Technologies OY</i> , IPR2024-01140, Paper 9 (P.T.A.B. Feb. 12, 2025)-----	70, 72
<i>Apple v. Fintiv</i> , IPR2020-00019, Paper 11 (P.T.A.B. Mar. 20, 2020)-----	69, 72
<i>Aptiv Services US, LLC v. Microchip Technology, Inc.</i> , IPR2024-00646, Paper 11 (P.T.A.B. Sept. 25, 2024) -----	70
<i>Cal. Inst. of Tech. v. Broadcom Ltd.</i> , 25 F.4th 976 (Fed. Cir. 2022)-----	72
<i>Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc.</i> , 109 F.3d 1394 (9th Cir. 1997)-----	44, 46
<i>Ericsson Inc. v. XR Communications LLC</i> , IPR2024-00613, Paper 9 (P.T.A.B. Oct. 9, 2024) -----	70, 71
<i>Google LLC v. Jawbone Innovations, LLC</i> , IPR2022-00630, Paper 10 (P.T.A.B. Sept. 13, 2022) -----	72
<i>In re GPAC Inc.</i> , 57 F.3d 1573 (Fed. Cir. 1995)-----	5
<i>Kom Software, Inc. v. NetApp, Inc.</i> , No. 2021-1075, 2021 WL 5985360 (Fed. Cir. Dec. 17, 2021)-----	23
<i>KSR Int’l Co. v. Teleflex Inc.</i> , 550 U.S. 398 (2007) -----	<i>passim</i>
<i>Leapfrog Enters., Inc. v. Fisher-Price, Inc.</i> , 485 F.3d 1157 (Fed. Cir. 2007) -----	69
<i>In re Marco Guldenaar Holding B.V.</i> , 911 F.3d 1157 (Fed. Cir. 2018) -----	43, 44
<i>Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co. Ltd.</i> , 868 F.3d 1013 (Fed. Cir. 2017) -----	6

<i>Quasar Sci. LLC v. Colt Int’l Clothing, Inc.,</i> IPR2023-00611, Paper 10 (P.T.A.B. Oct. 10, 2023) -----	73
<i>Samsung Electronics Co. v. Empire Technology Development LLC,</i> IPR2024-00896, Paper 15 (P.T.A.B. Dec. 13, 2024)-----	71
<i>SAP Am., Inc. v. InvestPic, LLC,</i> 898 F.3d 1161 (Fed. Cir. 2018) -----	43
<i>Sec. First Innovations, LLC v. Google LLC,</i> No. 2:23-cv-00097, 2024 WL 234720 (E.D. Va. Jan. 22, 2024)-----	69
<i>Sharpe Innovations, Inc. v. T-Mobile USA, Inc.,</i> No. 2:17-cv-00351, 2018 WL 11198604 (E.D. Va. Jan. 10, 2018)-----	69
<i>Shenzen Chic Elecs. v. Pilot, Inc.,</i> IPR2023-00810, Paper 12 (P.T.A.B. Nov. 8, 2023) -----	73
<i>Sotera Wireless, Inc. v. Masimo Corp.,</i> IPR2020-01019, Paper 12 (P.T.A.B. Dec. 1, 2020) -----	72
<i>Unwired Planet, LLC v. Google Inc.,</i> 841 F.3d 995 (Fed. Cir. 2016) -----	7
<i>Vivid Techs., Inc. v. Am. Sci. Eng’g, Inc.,</i> 200 F.3d 795 (Fed. Cir. 1999)-----	6
<i>Statutes and Rules:</i>	
35 U.S.C. §102 -----	7
35 U.S.C. §103 -----	6
35 U.S.C. §112 -----	6
35 U.S.C. §325 -----	73

TABLE OF EXHIBITS

Exhibit No.	Description
1001	U.S. Patent No. 10,805,111 (“the ’111 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”)
<i>Exhibit Number 1005 Not Used</i>	
1006	Yoshimura et al., <i>Content Delivery Network Architecture for Mobile Streaming Service Enabled by SMIL Modification</i> , 86 IEICE TRANSACTIONS ON COMM’N 1778 (2003) (“Yoshimura”)
<i>Exhibit Numbers 1007-1025 Not Used</i>	
1026	B. PRABHAKARAN, MULTIMEDIA SYSTEMS AND TECHNIQUES, 177-216 (Borko Furht ed., 1996) (“Prabhakaran”)
1027	U.S. Patent Publication No. 2003/0013073 (“Duncan”)
1028	U.S. Patent Publication No. 2002/0184189 (“Hay”)
1029	Robert Steele, et al., <i>Accessing SMIL-based Dynamically Adaptable Multimedia Presentations from Mobile Devices</i> , in INT’L CONF. ON INFO. TECH.: CODING AND COMPUTING (2004) (“Steele”)
1030	U.S. Patent No. 6,260,011 (“Heckerman”)
1031	THEODOR S. GEISEL, THE CAT IN THE HAT (1957) (“Dr. Seuss”)

Exhibit No.	Description
1032	U.S. Patent No. 6,108,703 (“Leighton”)
1033	U.S. Patent No. 5,922,045 (“Hanson”)
<i>Exhibit Numbers 1034-1052 Not Used</i>	
1053	U.S. Patent No. 5,845,303 (“Templeman”)
1054	U.S. Patent No. 6,789,229 (“Dunietz”)
1055	U.S. Patent No. 4,723,209 (“Hernandez”)
1056	U.S. Patent No. RE37,258 (“Patel”)
1057	U.S. Patent No. 5,798,841 (“Takahashi”)
1058	U.S. Patent No. 5,021,989 (“Fujisawa”)
1059	PCT Patent Publication No. WO2001/01373A2 (“Hendricks”)
1060	PCT Patent Publication No. WO2002/08948A2 (“Sull”)
1061	U.S. Patent No. 6,622,171 (“Gupta”)
1062	U.S. Patent No. 6,208,995 (“Himmel”)
1063	P. Delacourt & C.J. Wellekens, <i>DISTBIC: A speaker-based segmentation for audio data indexing</i> , 32 SPEECH COMMUNICATION 111 (2000) (“Delacourt”)
1064	Excerpt from WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY (2002)

Exhibit No.	Description
1065	Excerpt from RANDOM HOUSE WEBSTER’S COLLEGE DICTIONARY (2nd ed. 2001)
1066	Excerpt from B. PRABHAKARAN, MULTIMEDIA DATABASE MANAGEMENT SYSTEMS (1997) (“Prabhakaran II”)
1067	U.S. Patent Publication No. 2004/0024898 (“Wan”)
1068	Joshua Hodas et al., <i>NOVeLLA: A Multi-Modal Electronic-Book Reader With Visual and Auditory Interfaces</i> , 4 INT’L J. SPEECH TECH., 269 (2001) (“Hodas”)
1069	PCT Patent Publication No. WO2002/080524A2 (“Dimitrova”)
1070	PATRICK SCHMITZ, THE SMIL 2.0 TIMING AND SYNCHRONIZATION MODEL, TECH. RPT. MSR-TR-2001-01 (2001) (“Schmitz”)
1071	U.S. Patent Publication No. 2005/0245243A1 (“Zuniga”)
1072	Excerpt from File History for European Patent Publication No. EP1961154
1073	Advisory Action dated Oct. 12, 2010, U.S. Patent App. No. 12/096933
1074	Final Office Action dated July 28, 2010, U.S. Patent App. No. 12/096933
1075	Barry Arons, <i>SpeechSkimmer: A System for Interactively Skimming Recorded Speech</i> , 4 ACM TRANSACTIONS ON COMPUTER-HUMAN INTERACTION 3 (1997) (“Arons”)
<i>Exhibit Numbers 1076-1078 Not Used</i>	

Amazon.com, Inc. v. Audio Pod IP, LLC
IPR Petition – U.S. Pat. No. 10,805,111

Exhibit No.	Description
1079	THEODOR S. GEISEL, THE CAT IN THE HAT AND OTHER DR. SEUSS FAVORITES (narrated by Kelsey Grammer et al., Listening Library 2003) (CD)
<i>Exhibit Numbers 1080-1094 Not Used</i>	
1095	File History for U.S. Patent No. 10,805,111
1096	CV of Professor Ketan Mayer-Patel, Ph.D.
1097	Declaration of Sylvia D. Hall-Ellis, Ph.D.

Petitioners Amazon.com, Inc., Amazon.com Services LLC, Amazon Web Services, Inc., and Audible, Inc. (“Petitioners” or “Amazon”) respectfully request *inter partes* review of claims 1-16 of U.S. Patent No. 10,805,111 (“the ’111 patent”), which Audio Pod IP, LLC (“Patent Owner” or “PO”) purportedly owns.

I. INTRODUCTION

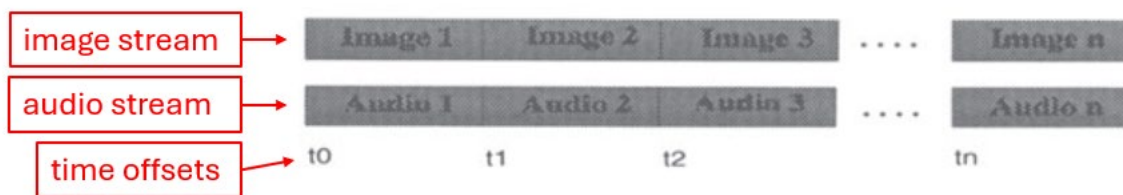
The challenged claims are directed to a client device that renders images and corresponding audio, which are obtained from a network accessible library. The broadest claims recite the well-known steps of downloading images and audio clips, assembling a page from the images, and rendering the page while playing the audio. By the time of the ’111 patent’s earliest possible priority date in December 2005, such methods were widely known.

The primary references in this Petition were never submitted to, or considered by, the Examiner, who allowed the ’111 patent claims without a single Office Action. Although the claims avoided meaningful examination during the application process, they are clearly unpatentable and should be canceled.

II. BACKGROUND AND STATE OF THE ART

A multimedia presentation comprises different media streams such as text, images, audio, and video. (EX-1026, 177; EX-1002 ¶32.) Multimedia synchronization refers to the task of coordinating objects from different media streams in the time domain. (*Id.*)

In 1996, nearly a decade before the earliest possible priority date of the '111 patent, Prabhakaran disclosed a method that included downloading corresponding image and audio streams from a multimedia database, assembling images for display, and rendering the images with their corresponding audio. (EX-1026, 177-78; EX-1002 ¶33.) Synchronization was achieved using pre-defined temporal relations, which specified the time offsets of the audio stream corresponding to each image:



(EX-1026, 177-78, Fig. 1¹ (excerpt); EX-1002 ¶33.)

In 2002, the National Information Standards Organization approved a standard for creating digital talking books, which involved the synchronized rendering of text, images, and audio. (EX-1003, vii; EX-1002 ¶34.) Using that standard, users could access digital talking books via the Internet and control the font size and audio playback speed on their devices. (EX-1003, vii, 21, 24-25, 63; EX-1002 ¶34.)

In 2003, Duncan disclosed an electronic book device that downloaded a media stream from a database. (EX-1027 ¶¶[0004], [0023], Fig. 2; EX-1002 ¶35.) The

¹ Figures herein may be colored and/or annotated for clarity.

device displayed a page and simultaneously read that page aloud before automatically turning the page to continue reading. (EX-1027, Abstract, ¶¶[0013], [0033]; EX-1002 ¶35.) The device stored “the start and end position of a page of text being displayed” to determine when a page should be displayed during the audio playback. (EX-1027 ¶[0028]; EX-1002 ¶35.) The device also paginated on the fly if, for example, a user selected a different font size. (EX-1027 ¶¶[0042]-[0044]; EX-1002 ¶35.)

III. THE ’111 PATENT

A. Overview

The ’111 patent describes well-known steps to download, assemble and render images and corresponding audio. (EX-1002 ¶36.) It describes accessing a library to download a portion of an audio stream and one or more images, the images associated with time information relative to the audio timeline. (EX-1001, 2:54-3:2; 6:53-7:2.) It describes assembling a page from the images, assigning time information to the page, and simultaneously rendering the page and corresponding audio. (*Id.*, 29:15-30:56, 39:34-43.)

B. Claims

Claims 1-16 are challenged in this petition. Claims 1 and 16 are independent. Claim 1 is representative and recites:

A method comprising:

- [a] accessing, by a client device, a network accessible library having stored thereon an image stream of static graphic images and a corresponding audio stream;
- [b] downloading to the client device from the network accessible library one or more static graphic images from the image stream, wherein the one or more static graphic images are each associated with time information including at least one of a start time, an end time, and a duration relative to a timeline of the audio stream;
- [c] assembling, by the client device, a first page from the one or more static graphic images;
- [d] assigning to the first page, by the client device, time information including at least one of a start time, an end time, and a duration relative to the timeline of the audio stream, wherein the time information for the first page is determined on the basis of the time information for the one or more static graphic images;
- [e] downloading to the client device from the network accessible library a portion of the audio stream including a first time offset, wherein the first time offset corresponds to a first position on the first page; and
- [f] simultaneously rendering the first page and the portion of the audio stream on the client device by using the time information for the one or more static graphic images or for the first page, wherein the portion of the audio stream is rendered in dependence upon the first time offset.

The dependent claims recite trivial variations, such as the library including a server, the number of images rendered being dependent on the size of the image or device capabilities, assembling a second page, and automatically turning the page.

C. Prosecution

The Examiner allowed the claims without a single Office Action. (EX-1095, 23-30.) None of the references herein were meaningfully considered by the Examiner. Hay (relied on herein for dependent claims 6 and 7) was cited in a lengthy IDS but none of the other references were before the Examiner.

D. Priority

The '111 patent's earliest possible priority date is December 13, 2005. (EX-1001, 1-2.) Petitioner does not concede that 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). Here, a POSITA would have had at least a bachelor's degree in electrical engineering, computer engineering, or computer science, and at least three years of industry or academic experience in the design, development, and/or implementation of content rendering and/or distribution systems. (EX-1002 ¶¶27-31); *see In re GPAC Inc.*, 57 F.3d 1573, 1579 (Fed. Cir. 1995). Work experience could substitute for formal education and additional formal education could substitute for work experience. (EX-1002 ¶29.)

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, Petitioners assume the claims are not invalid under §112.

VI. STATEMENT OF PRECISE RELIEF REQUESTED

A. Grounds

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

Ground	Challenged Claims	References
1A	1-2, 4-5, 16	Prabhakaran and Duncan
1B	3	Prabhakaran, Duncan, and Templeman
1C	6	Prabhakaran, Duncan, and optionally Hay
1D	7	Prabhakaran, Duncan, and optionally Hay and/or Heckerman
1E	8-11	Prabhakaran, Duncan, and DTB
1F	12-15	Prabhakaran, Duncan, and Dr. Seuss
2A	1-2, 4-5, 8-11, 16	DTB, Steele, and Duncan
2B	3	DTB, Steele, Duncan, and Templeman

Ground	Challenged Claims	References
2C	6	DTB, Steele, Duncan, and Hay
2D	7	DTB, Steele, Duncan, Hay, and Prabhakaran
2E	12-15	DTB, Steele, Duncan, and Dr. Seuss

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 35 U.S.C. §102(b) because it published more than one year before the '111 patent's earliest possible priority date of December 13, 2005: Dr. Seuss (1957); Prabhakaran (1996); Templeman (1998); Heckerman (2001); DTB (2002); Hay (2002); Duncan (2003); and Steele (April 2004). (EX-1053, 1; EX-1030, 1; EX-1028, 1; EX-1027, 1; EX-1097 ¶¶51-78.)

These references are analogous art because they are from the same field of endeavor as the '111 patent, e.g., content distribution and/or rendering. (EX-1002 ¶22); *Unwired Planet, LLC v. Google Inc.*, 841 F.3d 995, 1000 (Fed. Cir. 2016). 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 1-2, 4-5, AND 16 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN AND DUNCAN.

A. Claim 1

1. Preamble

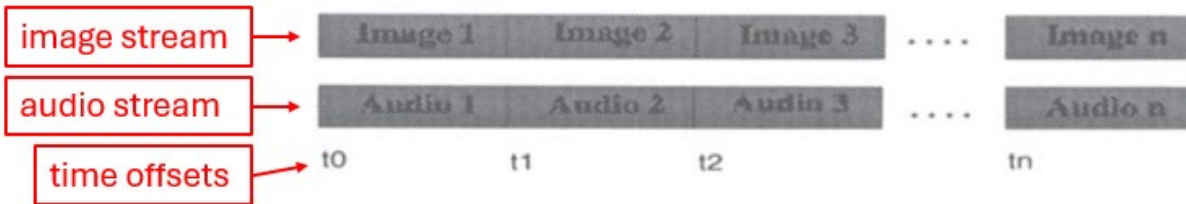
The preamble recites “[a] method comprising.” Prabhakaran discloses a method including the steps of (1) acquiring multimedia objects from servers, (2) synchronizing multimedia objects in the time domain, and (3) concurrently presenting synchronized multimedia objects. (EX-1026, 177-81, 206-10; EX-1002 ¶40.) Accordingly, to the extent the preamble is limiting, it is disclosed or rendered obvious by Prabhakaran. (EX-1002 ¶40.)

2. 1[a]: Network Accessible Library

Element 1[a] recites “accessing, by a client device, a network accessible library having stored thereon an image stream of static graphic images and a corresponding audio stream.”

Prabhakaran discloses that client devices retrieve (via a network) “multimedia objects” including text, images, audio, and video from databases stored on servers. (EX-1026, 177-78, 181 (network/database), 206 (network/servers), 210 (multimedia objects retrieved from server); EX-1002 ¶42.) Synchronization points (also referred to as “temporal relations” or “points of synchronization”) between the multimedia objects are “pre-defined” and “stored along with the multimedia objects.”

(EX-1026, 177, 213.) Figure 1 shows a presentation of images corresponding to audio clips at time offsets (e.g., t_0 , t_1 , t_2 , t_n):



(*Id.*, 178, Fig. 1 (excerpt); EX-1002 ¶42.)

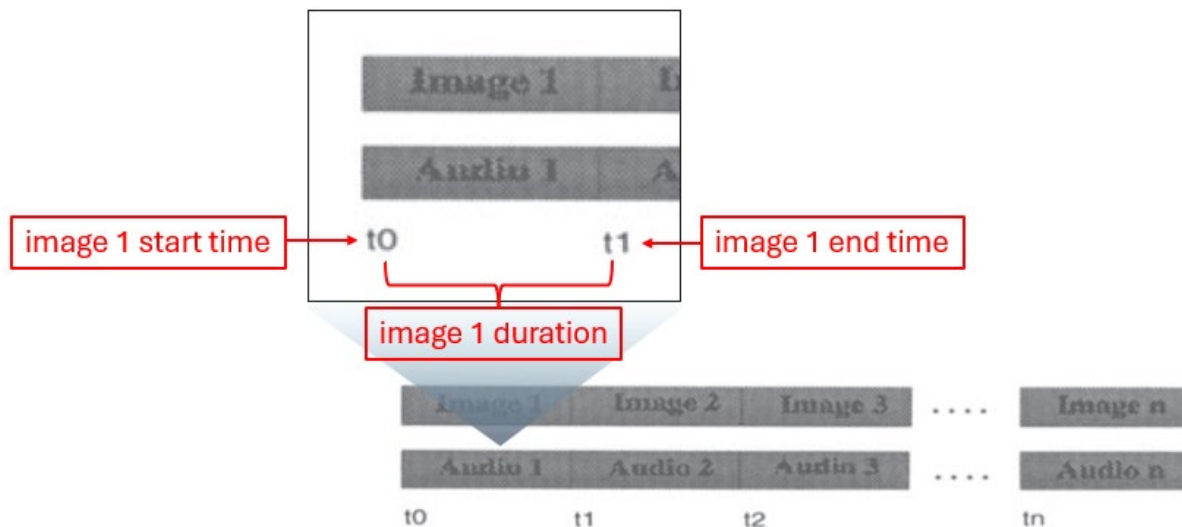
Prabhakaran therefore discloses or renders obvious a client device accessing (e.g., retrieving multimedia objects from) a network accessible library (e.g., multimedia database on server) having stored thereon an image stream of static graphic images (e.g., Image 1, Image 2, Image 3, ... Image n) and corresponding audio stream (e.g., Audio 1, Audio 2, Audio 3, ... Audio n). (EX-1002 ¶¶41-43.)

3. 1[b]: Downloading Image Associated with Time Information

Element 1[b] recites “downloading to the client device from the network accessible library one or more static graphic images from the image stream, wherein the one or more static graphic images are each associated with time information including at least one of a start time, an end time, and a duration relative to a timeline of the audio stream.”

Prabhakaran’s client retrieves, from the server via a network, images from an image stream. (*Supra* §VII.A.2.) Pre-defined temporal relations between the image

and audio streams are stored in the multimedia database. (*Id.*) Figure 1 shows each image corresponds to an audio clip (e.g., Image 1 corresponds to Audio 1) such that the corresponding image and audio clip share the same start time, end time, and duration:



(EX-1026, 178, Fig. 1 (excerpt); EX-1002 ¶45.) Each image is therefore displayed relative to the same timeline as the corresponding audio file. (*Id.*)

Prabhakaran therefore discloses or renders obvious downloading (e.g., retrieving) to the client device from the network accessible library (e.g., multimedia database on server) one or more static graphic images from the image stream (e.g., Image 1), each associated with time information including at least one of a start time, an end time, and a duration (e.g., t_0-t_1) relative to a timeline of the audio stream. (EX-1002 ¶¶44-46.)

4. 1[c]: Assembling a First Page

Element 1[c] recites “assembling, by the client device, a first page from the one or more static graphic images.” Prabhakaran discloses assembling a first page because it renders a page comprising, e.g., Image 1. (EX-1002 ¶48.)

Prabhakaran’s process for assembling a page is the same as the ’111 patent’s. (*Id.*)

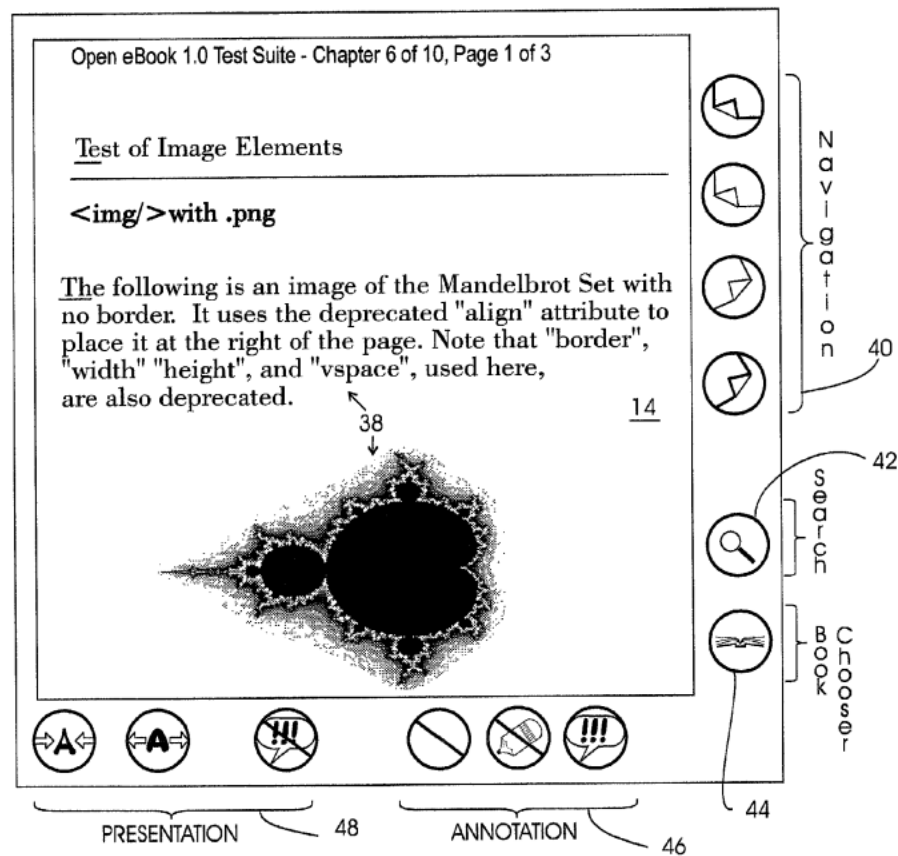
The ’111 patent describes a “dynamic approach” where pages are created at run time. (EX-1001, 29:34-63.) “[P]ages are defined on the basis of screen size, font size, font type, presentation style, included graphics, other ancillary content, writing style ... and text availability.” (*Id.*, 29:35-39.) The available space on a page is “filled with appropriately formatted text and ancillary content” including images. (*Id.*, 29:47-49.) A page may contain only one image or more than one image. (*Id.*, 30:30-37; EX-1002 ¶48.)

Prabhakaran similarly explains that the user can modify the presentation by, for example, “scal[ing] the spatial requirements” and “handl[ing] spatial clash[.]” (EX-1026, 178 (“The points of synchronization in a multimedia presentation can be modified by the user going through the presentation.... for example, a user may interact by giving inputs such as ... scale the spatial requirements, handle spatial clash”); EX-1002 ¶49.) Users can modify the size of objects displayed on the screen and therefore assemble a page in the same way disclosed by the ’111 patent (e.g.,

modifying content size at run time). (*Id.*) A POSITA would have understood that on client devices with large displays, the client device may arrange Images 1-3 on a single page, modifying the synchronization points so that Images 1-3 were all displayed on a page from t0 to t3. (EX-1002 ¶49.) On client devices with smaller displays, the client device may fit only Images 1-2 on a page, modifying the synchronization points so that Images 1-2 were displayed on a page from t0 to t2. (*Id.*)

Prabhakaran therefore discloses or renders obvious assembling (e.g., in response to user input scaling the object size), by the client device, a first page (e.g., display) from one or more static graphic images. (*Id.* ¶¶47-50.)

Even if Prabhakaran did not disclose or render obvious this limitation, it would have been obvious in view of Duncan. (*Id.* ¶51.) Duncan discloses an electronic book device that assembles pages including images. (EX-1027.) An exemplary page is shown in Figure 3:



(*Id.*, Fig. 3.) Just as the '111 patent describes “a dynamic approach to pagination” (EX-1001, 29:34-47), Duncan describes a “paginator” that can “function dynamically at any time” (EX-1027 ¶[0042]). Duncan, like the '111 patent, discloses re-paginating in response to a change in font size. (*Compare* EX-1001, 29:34-45 with EX-1027 ¶[0042]-[0044].) The client’s microprocessor performs this pagination. (EX-1027 ¶[0021]-[0025], Fig. 1 (client’s microprocessor executes pagination logic).) Duncan therefore discloses a client device performing this limitation. (EX-1002 ¶51.)

A POSITA would have been motivated to use Duncan’s pagination method in Prabhakaran’s system for several reasons. (*Id.* ¶¶52-57.)

First, a POSITA would have understood the benefits of allowing users to adjust the size of images and text being rendered. (*Id.* ¶53.) Indeed, Duncan teaches that display scalability would beneficially accommodate people with visual impairments. (EX-1027 ¶[0005]; EX-1002 ¶53.) Moreover, a POSITA understood that adjusting component sizing improved user comfort as different users may prefer different text or image sizes depending on their vision, screen size, or reading preference. (EX-1002 ¶53.) Allowing the client device to adjust the sizing and/or positioning of displayed components also ensures optimal viewing on a device regardless of screen size or device capabilities. (*Id.*)

Second, Duncan explains that the preferred output modes for its system align with Prabhakaran’s presentation (e.g., including “visual graphics and sound”). (EX-1027 ¶¶[0008]-[0009]; EX-1002 ¶54.)

Third, Prabhakaran states that “presentation of multimedia information to the user involves spatial organization” and that the user can “scale the spatial requirements[.]” (EX-1026, 177-78.) This would have incentivized a POSITA to look for methods of scaling image size. (EX-1002 ¶55.) A POSITA would have understood Duncan’s pagination method to be a viable and beneficial method of doing so. (*Id.*)

Fourth, the combination represents nothing more than the simple addition of one known element (e.g., Duncan’s pagination method) to another known element (e.g., Prabhakaran’s displaying images from an image stream) to obtain predictable results (e.g., arranging a page of image(s)). (EX-1002 ¶56); *KSR*, 550 U.S. at 417.

Fifth, the combination represents the use of a known technique (e.g., Duncan’s pagination method) to improve a similar method (e.g., Prabhakaran’s displaying images) in the same way. (*Id.*)

Sixth, the combination applies a known technique (e.g., Duncan’s pagination method) to a known method (e.g., Prabhakaran’s rendering) that is ready for improvement and yields predictable results (e.g., a page of image(s)). (*Id.*)

Finally, assembling an electronic page was commonplace, well-understood, and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (*Id.* ¶57 (citing EX-1053, 11:32-47; EX-1054, 2:7-30; EX-1055, 12:15-22; EX-1056, 1:37-40).)

5. 1[d]: Assigning Time Information to First Page

Element 1[d] recites “assigning to the first page, by the client device, time information including at least one of a start time, an end time, and a duration relative to the timeline of the audio stream, wherein the time information for the first page is determined on the basis of the time information for the one or more static graphic images.” The ’111 patent describes the start and end time of a page as the time

information associated with “the first and last words on each page.” (EX-1001, 30:8-12.) While much of the ’111 patent’s disclosure relates to text, it could also apply to images. (EX-1002 ¶60.) The duration of a page can be determined by direct lookup of “the start time of the first word on the page” and calculating “[t]he start time of the last word plus the duration of the last word.” (*Id.*, 27:6-10, 30:12-17, 30:30-37.) Where “there is only one piece of [] content on the page, the start time and end time will be that associated with that one piece of [] content.” (*Id.*, 30:35-37.) Because the timing information (e.g., start, end, duration) of the page is based on the time assigned to the contents (e.g., text, images, etc.) of the page, a client device “assigns” time information to a page by determining what is displayed on a page (e.g., as a result of font size changes). (EX-1002 ¶60.)

Prabhakaran’s client computer assigns time information to a page in at least three ways.

a. Assembling a Page

Prabhakaran assigns time information to a first page in the same manner described by the ’111 patent. (EX-1002 ¶62.)

Prabhakaran, alone or in combination with Duncan, discloses assembling a page by determining what content is displayed on each page (e.g., through font and image size). (*Supra* §VII.A.4; EX-1002 ¶63.) Where Prabhakaran’s client assembles a page comprising only Image 1, the client device assigns time to that page

associated with the audio rendered when Image 1 is displayed (i.e., start at t_0 , end at t_1 , and a duration of t_0-t_1). (EX-1026, 178, Fig. 1; EX-1002 ¶63.) However, where Prabhakaran’s client scales the spatial aspects of the visual components of the presentation to include, for example, Images 1-3 on a page (*supra* §VII.A.4), the client device assigns time to that page associated with audio rendered for Images 1-3 (i.e., start at t_0 , end at t_3 , and a duration of t_0-t_3). (EX-1002 ¶63.) Prabhakaran expressly refers to the relationship “among the objects to be presented” as “*relative temporal relationships*.” (EX-1026, 179; *see also id.*, 179-80 (the specification to “present object A with ... object B” is a “relative temporal specification”).)

Even if Prabhakaran did not disclose or render obvious this limitation, it would have been obvious in view of Duncan. Duncan, like the ’111 patent, discloses identifying “the start and end position of a page [] being displayed” when rendering a digital talking book. (*Compare* EX-1027 ¶[0028] *with* EX-1001, 30:8-12 (determining time offsets for “the first and last words on each page” when producing synchronized audiovisual output); EX-1002 ¶64.)

Accordingly, Prabhakaran alone or in combination with Duncan discloses or renders obvious assigning to the first page (e.g., Image 1 or Images 1-3), by the client device (e.g., based on scaling the spatial aspects of the images), time information including at least one of a start time (e.g., t_0), an end time (e.g., t_1 or t_3), and a duration (e.g., t_0-t_1 or t_0-t_3) relative to the timeline of the audio stream, wherein the

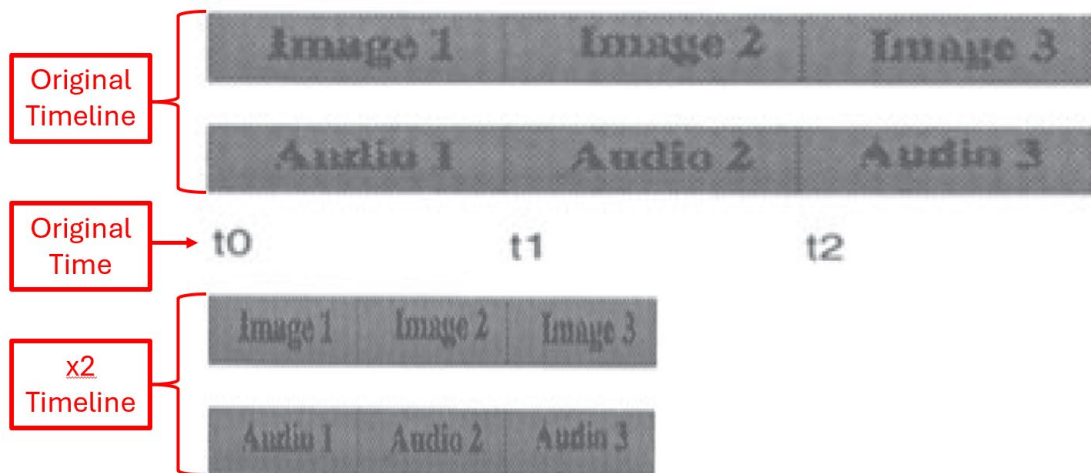
time information for the first page is determined on the basis of the time information for the one or more static graphic images. (EX-1002 ¶¶62-65.) A POSITA would have been motivated to combine Duncan’s paginator with Prabhakaran’s synchronized rendering of images and audio. (*Supra* §VII.A.4; EX-1002 ¶65.)

Moreover, a POSITA would have reasonably expected success in using a paginator (e.g., Duncan’s) with synchronized pre-recorded audio in systems like Prabhakaran’s (e.g., where visual content like images is pre-linked with pre-recorded audio) at least because doing so would have been “straightforward to implement,” according to Duncan’s inventors. (EX-1068, 284; EX-1002 ¶66.)

b. Changing the Presentation Speed

Prabhakaran discloses that the user can “scale the speed of presentation.” (EX-1026, 178.) Scaling the speed alters the amount of time each image is displayed. (EX-1002 ¶67.) When the user scales the presentation speed, the client assigns new time information for each image, and consequently for the page on which it is displayed, to maintain synchronization with the accelerated audio. (*Id.*)

For example, if a user accelerated the speed of the presentation depicted in Figure 1 by a factor of 2, the client would assign new time information to each page (e.g., image):



(EX-1026, 178, Fig. 1 (excerpt); EX-1002 ¶68.)

The newly assigned time information for the page (start, end, duration) is relative to the timeline of the audio stream because the time information assigned to the page (t_0 , $t_{0.5}$, etc.) is based on the time information of the images, which Prabhakaran discloses is “relative” to the timeline of the audio stream. (EX-1002 ¶69.)

The newly assigned time information is also relative to the original audio timeline because each new transition/synchronization point (e.g., Image 1/Audio 1 to Image 2/Audio 2) is dependent on (e.g., calculated as a ratio of) the original audio transition/synchronization points. (EX-1064, 1916 (defining “relative” as “expressed as the ratio of the specified quantity to the total magnitude or to the mean of all the quantities involved”); EX-1065, 1040 (defining “relative” as “something dependent upon external conditions for its specific nature, size, etc. (opposed to absolute)”); EX-1002 ¶70.)

The time information for the first page is determined on the basis of the time information for the images because the accelerated display time of the pages (e.g., images) are calculated based on (e.g., as a ratio of) the time information for the images. (EX-1002 ¶71.)

Again, this is precisely why Prabhakaran teaches using “relative temporal relationships” between the images (pages) and the audio. (*Id.* ¶72.) When a user “scale[s] the speed of presentation,” the time information assigned to the new image/page is “relative to” the original timeline (e.g., duration is half the original time when the speed is doubled). (*Id.*)

Accordingly, Prabhakaran discloses or renders obvious assigning to the first page (e.g., Image 1), by the client device (e.g., by scaling the speed), time information including at least one of a start time (e.g., t_0), an end time (e.g., $t_{0.5}$), and a duration (e.g., $t_0 - t_{0.5}$) relative to the timeline of the audio stream, wherein the time information for the first page is determined on the basis of the time information for the one or more static graphic images. (*Id.* ¶¶67-73.)

c. Translating Synchronization Times into Presentation Times

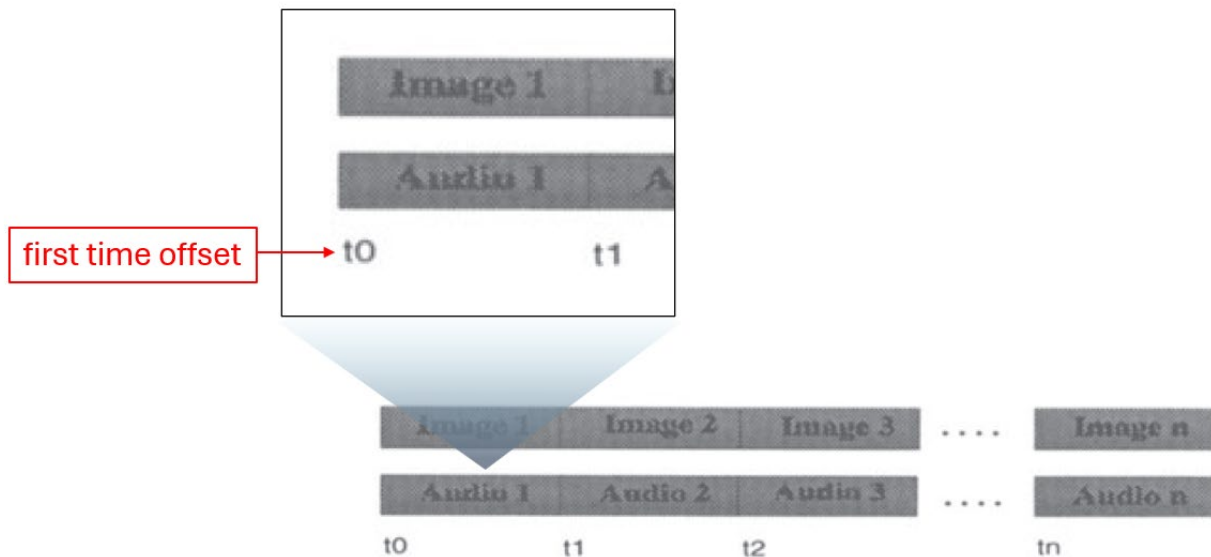
A POSITA would have understood that Prabhakaran’s client assigns time by translating the proposed synchronization points (e.g., t_0 , t_1 , etc.) to actual presentation times. (*Id.* ¶74.) In a contemporaneous textbook, Prabhakaran’s author explained that the pre-defined temporal relationships between media objects (e.g., the

synchronization points between image and audio streams) “have to be *translated* into a presentation schedule that specifies the time instants and durations of the object presentations.” (EX-1066, 157 (emphasis in original); *see also* EX-1067 ¶¶[0044]-[0045]; EX-1002 ¶74.) The client assigns presentation times to the pages by converting the relative timeline assigned to the objects to “the real-world time at which the event happens.” (EX-1070, 13; EX-1002 ¶74.) Future objects (e.g., objects occurring at t1, t2, tn) are then assigned presentation (e.g., real-world) times based on the presentation time of prior objects using the same relative, pre-assigned temporal spacing. (*Id.*)

6. 1[e]: Downloading Audio

Element 1[e] recites “downloading to the client device from the network accessible library a portion of the audio stream including a first time offset, wherein the first time offset corresponds to a first position on the first page.” The ’111 patent discloses a default time offset of zero. (EX-1001, 17:18-19.) The ’111 patent also acknowledges that downloading a portion of an audio stream was not new. (*Id.*, 2:16-21; *see also* EX-1004; EX-1002 ¶77.)

As shown in Figure 1, Prabhakaran’s client downloads a portion of the audio stream (e.g., Audio 1), which includes a first time offset (e.g., t0), and that time offsets corresponds to a first position (starting time or starting spatial position) on the first page (e.g., comprising Image 1 or Images 1-3):



(EX-1026, 178, Fig. 1 (excerpt); EX-1027, Fig. 3; EX-1002 ¶¶78.) The segments are retrieved from a network accessible library (e.g., multimedia database on server). (*Supra* §VII.A.3; *see also* EX-1026, 206-09 (disclosing an “optimal retrieval schedule” of the segments of Prabhakaran’s audio stream); EX-1002 ¶¶78.)

Prabhakaran therefore discloses or renders obvious downloading (e.g., retrieving) to the client device from the network accessible library (e.g., multimedia database on server) a portion of the audio stream (e.g., Audio 1 or Audio 1-3) including a first time offset (e.g., t_0) that corresponds to a first position (e.g., starting time or starting spatial position) on the first page (e.g., page comprising Image 1 or Images 1-3). (EX-1002 ¶¶76-79.)

Indeed, even if Prabhakaran disclosed downloading the entire audio stream, it would still disclose this claim limitation because “a portion” includes the entirety of

the audio stream. (EX-1002 ¶80); *see Kom Software, Inc. v. NetApp, Inc.*, No. 2021-1075, 2021 WL 5985360, at *1 (Fed. Cir. Dec. 17, 2021) (rejecting the argument that “portion” must mean “less than the entirety of each file” where a dependent claim included the phrase “only a portion”); (*cf.* EX-1001, claim 14 (referencing “only a portion”).)

7. 1[f]: Simultaneously Rendering Images and Audio

Element 1[f] recites “simultaneously rendering the first page and the portion of the audio stream on the client device by using the time information for the one or more static graphic images or for the first page, wherein the portion of the audio stream is rendered in dependence upon the first time offset.”

Prabhakaran discloses rendering the page and corresponding audio using the relevant time information. (*Supra* §§VII.A.4-VII.A.6.)

Because the multimedia presentation is rendered according to—and in dependence on—the synchronization points, the audio and image streams are rendered in dependence on the time offset. (EX-1002 ¶84.) For example, Audio 1 and Image 1 are rendered in dependence upon the time offset, t_0 . (*Id.*) To the extent PO argues this limitation requires the rendering to occur in response to, e.g., a bookmark, play command, etc., it still would have been obvious to a POSITA in view of Prabhakaran

alone or in combination with Duncan. (*Supra* §VII.A.4; EX-1027 ¶¶0032] (navigation buttons allow users to navigate to specific page or “go to the end or start of the book”); EX-1002 ¶84.)

Prabhakaran therefore discloses or renders obvious simultaneously rendering (e.g., presenting concurrently) the first page (e.g., Image 1 or Images 1-3) and the portion of the audio stream (e.g., Audio 1 or Audio 1-3) on the client device by using the time information (e.g., t0, t1, t2) for the one or more static graphic images (e.g., Image 1, Image 2, Image 3) or for the first page, wherein the portion of the audio stream is rendered in dependence upon the first time offset (e.g., t0). (*Supra* §§VII.A.3 (image time information), VII.A.4 (first page), VII.A.5 (first page time information), VII.A.6 (audio portion with first time offset); EX-1002 ¶¶82-87.)

B. Claim 2

Claim 2 depends from claim 1 and recites that “the network accessible library includes one or more server devices.”

Prabhakaran discloses a network accessible library stored on “server(s),” plural. (EX-1026, 206-13; *supra* §VII.A.2.) Prabhakaran therefore discloses or renders obvious this limitation. (EX-1002 ¶¶88-90.)

C. Claim 4

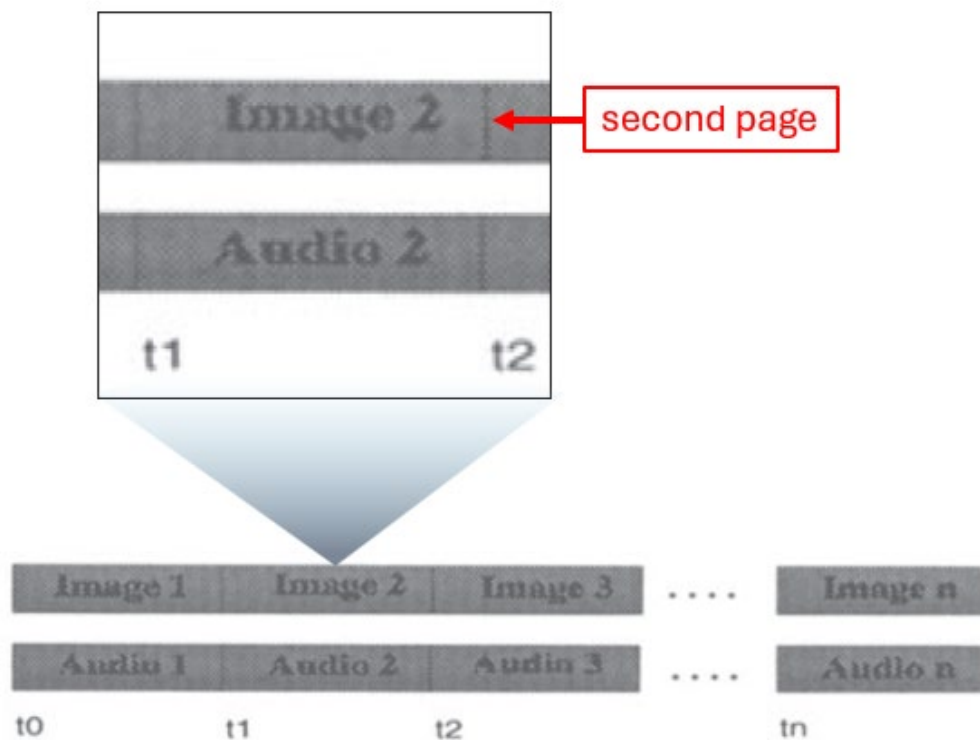
Claim 4 depends from claim 1 and recites “wherein the first time offset is a start time of the first page, a start time of one of the one or more static graphic images, or a bookmarked position.”

Prabhakaran discloses or renders obvious a first time offset (e.g., t_0) that is the start time (e.g., t_0) of the first page (*supra* §VII.A.4) and a start time (e.g., t_0) of one of the one or more static graphic images (Image 1). (EX-1002 ¶¶91-93.)

D. Claim 5

Claim 5 depends from claim 1 and recites steps for “a second page” identical to claim 1’s steps for assembling, and assigning time information to, a first page. (EX-1002 ¶94.) Prabhakaran discloses or renders obvious assembling and assigning time information to a second page for the same reasons it discloses assembling and assigning time information to a first page. (*Supra* §§VII.A.4, VII.A.5; EX-1002 ¶95.)

For example, a POSITA would have understood that if Prabhakaran assembled a first page comprising Image 1 only, then Prabhakaran would assemble a second page comprising Image 2, and would assign time information to that second page (e.g., start t_1 , end t_2 , and duration t_2-t_1):



(EX-1026, 178, Fig. 1 (excerpt); EX-1002 ¶¶96.) The assigned time information would be relative to the timeline of the audio stream (just like the first page) and the time information would be determined on the basis of the time information (t1 to t2) for the image (Image 2) displayed on the page. (EX-1002 ¶¶96.)

Even if Prabhakaran did not disclose assembling a second page and assigning time information, it would have been obvious in view of Duncan for the same reasons as for limitations 1[c]-1[d]. (*Supra* §§VII.A.4-VII.A.5; EX-1027 ¶¶[0033] (disclosing advancing to a “next page”); EX-1002 ¶¶97.) A POSITA would have been motivated to combine Prabhakaran and Duncan. (*Supra* §§VII.A.4-VII.A.5; EX-1002 ¶¶97.)

E. Claim 16

Claim 16 recites “[a] non-transitory computer-readable storage medium having stored thereon instructions which, when executed by a processor of a computing device, cause the computing device to” perform the steps of claim 1.

Prabhakaran discloses a “client” in a “computer network.” (EX-1026, 206.) A POSITA would have understood the client computer to comprise a “non-transitory computer readable storage medium” that includes instructions for performing the steps described. (EX-1002 ¶101.) Thus, Prabhakaran discloses the preamble. (*Id.*) Duncan also discloses this. (*Id.*; EX-1027 ¶¶[0010], [0023].)

After the preamble, claim 16 merely recites steps identical to those in claim 1. (EX-1002 ¶102.) Prabhakaran and Duncan together disclose or render obvious each of those steps and thus render obvious claim 16. (*Supra* §VII.A; EX-1002 ¶¶100-03.)

VIII. GROUND 1B: CLAIM 3 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND TEMPLEMAN.

Claim 3 depends from claim 1 and recites that “the number of the one or more static graphic images included on the first page is determined on the basis of the size of the one or more static graphic images and the graphic rendering capability of the client device.”

Claim 3’s additional limitation would have been obvious in view of Prabhakaran and Duncan. (*Supra* §VII.A.4; EX-1002 ¶105.)

Even if Prabhakaran and Duncan did not render this limitation obvious, it would have been obvious in view of Templeman. (EX-1002 ¶105.)

Templeman discloses a system for manipulating and displaying information where the system determines constraints on the display. (EX-1053, Abstract; EX-1002 ¶106.) Templeman’s devices may automatically select a display format (referred to as a “form” or “metaform”) depending on the capabilities of the device. (EX-1053, 3:26-48; EX-1002 ¶106.) Specific formats “may be developed for particular screen display types to effectively utilize available space and size.” (EX-1053, 11:50-53, 12:21-29.) For example, if the display device “is a six inch diagonal screen for a PDA, individual frame size will be more severely constrained than if display device [] were a fifteen inch diagonal monitor on a desktop computer system.” (*Id.*, 9:3-14.) Templeman’s device then fills the page with content until the page is filled (e.g., “the page bottom constraints are violated”) before filling the next page with remaining content. (*Id.*, 11:17-41, 10:46-65 (disclosing importing images in addition to text).)

A POSITA would have been motivated to determine the number of images displayed on a page based on the size of the images and the display constraints (e.g., fill a page with images where the page size is determined by the screen size as disclosed by Templeman) for multiple reasons. (EX-1002 ¶¶107-12.)

First, a POSITA would have been motivated to look to a reference like Templeman to understand how to “scale the spatial requirements” as suggested by Prabhakaran. (*Supra* §VII.A.4; EX-1002 ¶108.)

Second, a POSITA would have been motivated, in view of Templeman, to select the number of images Prabhakaran’s client displays on a page based on the image size and device capabilities. A POSITA would have been motivated to display, for example, a single image (e.g., Prabhakaran’s Image 1) per page on a 6-inch screen, but display multiple images (e.g., Prabhakaran’s Images 1-3) per page on a 15-inch screen. (*Supra* VII.A.4; EX-1053, 9:44-46; EX-1002 ¶109.)

Third, a POSITA would have been motivated to select the images to be displayed on the page based on image and screen size because Templeman warns that “shrink[ing] the size of the [content] to correspond to the smaller screen” would be “unsatisfactory as it may make the [content] illegible [and/or] result in a document having an excess of whitespace[.]” (EX-1053, 2:49-50; EX-1002 ¶110.)

Fourth, the combination represents nothing more than the simple addition of one known element (e.g., modifying the amount of content on a page depending on content size and screen size) to another known element (e.g., Prabhakaran’s pages comprising images) to obtain predictable results (e.g., pages where the number of images is based on image and screen size). (EX-1002 ¶111); *KSR*, 550 U.S. at 417.

Fifth, the combination represents the use of a known technique (e.g., modifying the amount of content on a page depending on content and screen size) to improve a similar method (e.g., Prabhakaran’s display) in the same way. (*Id.*)

Sixth, the combination applies a known technique (e.g., modifying the amount of content on a page depending on the content and screen size) to a known method (e.g., Prabhakaran’s) that is ready for improvement and yields predictable results (e.g., arranging images based on the size of the image and the size of the screen). (*Id.*) The combination would improve the user experience by allowing users to better see the content on their device. (EX-1002 ¶111.)

Finally, a POSITA would have reasonably expected success because pagination based on the size of the content and the size of the display could be implemented through routine code that was commonly known and well understood by a POSITA. (*Id.* ¶112.) Moreover, Duncan describes providing audiobooks via HTML and Templeman teaches that HTML-formatted data can be “readily imported or easily translated” for use with its metaforms. (EX-1027 ¶¶[0025], [0041]; EX-1053, 7:62-67.)

IX. GROUND 1C: CLAIM 6 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND OPTIONALLY HAY.

Claim 6 depends from claim 5 and recites “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 as the portion of the audio stream is rendered.”

Prabhakaran discloses multimedia presentations that a user can “freeze and restart.” (EX-1026, 178.) A POSITA would therefore have understood that Prabhakaran’s display of an image stream progresses without user action as the audio stream is rendered. (EX-1002 ¶117.) Accordingly, Prabhakaran discloses or renders obvious automatically turning (e.g., progressing without user action) from the first page (e.g., Image 1) to the second page (e.g., Image 2) in dependence upon the time information for the first page or for the second page (t_1) as the portion of the audio stream (Audio 1 to Audio 2) is rendered. (*Id.*)

Even if Prabhakaran did not disclose or render obvious this limitation, it would have been obvious over Duncan. Duncan teaches that, “[w]hen the system reaches the end of the text of the current page, it automatically advances the viewer to the next page.” (EX-1027 ¶[0033].) In view of this teaching, a POSITA would have been motivated to implement Prabhakaran’s multimedia presentation such that it automatically advances to the next “page” in dependence upon the timing information for the first page (e.g., end time (t_1)) or the second page (e.g., begin time (t_1)). (EX-1002 ¶118.)

Even if Prabhakaran and Duncan did not disclose and render obvious this limitation, it would have been obvious in view of Hay.

Hay discloses that “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 ¶119.) 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.*)

A POSITA would have been motivated to automatically turn Prabhakaran’s pages (e.g., as disclosed by Duncan or Hay) for multiple reasons. (EX-1002 ¶¶120-24.)

First, a POSITA would have understood that automatically turning the page would improve the reader’s experience and be more convenient. (*Id.* ¶121.) For example, automatic page turning would allow a smooth flow of audio and avoid unnecessary interruptions in the reading experience at the end of every page. (*Id.*) It would also eliminate the need to press a button or provide another input at the end of each page. (*Id.*)

Second, Prabhakaran discloses that a user can “freeze” a presentation synchronized with time tags. (EX-1026, 178; EX-1002 ¶122.) A POSITA would have been incentivized to look for methods to automatically progress Prabhakaran’s presentation using time tags, which Hay discloses. (EX-1026, 178; EX-1028 ¶¶[0100], [0217]; EX-1002 ¶122.)

Third, the combination represents nothing more than the simple addition of one known element (e.g., automatic page turning) to another known element (e.g., multimedia presentations) to obtain predictable results (e.g., continuous rendering of content). (EX-1002 ¶123); *KSR*, 550 U.S. at 417.

Fourth, the combination uses a known technique (e.g., automatic page turning) to improve a similar method (e.g., Prabhakaran’s content rendering) in the same way. (*Id.*)

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

Finally, methods to automate page turning were commonplace, well-understood, and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (*Id.* ¶124 (citing EX-1059, 49:6-7; EX-1057, Abstract; EX-1058, 4:32-40).)

X. GROUND 1D: CLAIM 7 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND OPTIONALLY HAY AND/OR HECKERMAN.

Claim 7 depends from claim 6 and recites “wherein the portion of the audio stream is rendered without any perceptible interruption.” Prabhakaran renders this obvious because a POSITA would have understood that Prabhakaran’s multimedia presentation is presented without any interruption. (EX-1002 ¶129.) The ’111

patent discusses two means for rendering an audio stream without interruption: (1) “segmenting an audio stream into a plurality of small digital audio files” to avoid download delays and (2) doing so “using gaps in the natural language of the audio” such as periods of silence between words. (EX-1001, 1:49-64; 2:54-3:2, 35:24-34; EX-1002 ¶129.) Both of these would have been obvious. (EX-1002 ¶129.)

Prabhakaran explains that “multimedia objects can be very large and hence can consist of many packets” retrieved on a schedule to avoid transfer delays. (EX-1026, 206-09; EX-1002 ¶130.) Thus, it would have been obvious that Prabhakaran downloads an audio stream in portions to avoid interruptions caused by downloading a larger file. (EX-1002 ¶130 (citing EX-1073, 2-3 (“quicker download” is the “predictable result” of “breaking each audio files into smaller ones”).)

Prabhakaran therefore discloses or renders obvious a method by which portions of an audio stream would be rendered without perceivable interruption. (EX-1002 ¶131.)

This limitation also would have been obvious in view of Heckerman. (EX-1002 ¶132.) Like the ’111 patent, Heckerman discloses segmenting audio streams at natural language gaps. (*Id.*) Heckerman uses speech recognition, where silence is modeled as a word that can be recognized. (EX-1030, Abstract, 4:1-13,

11:49-67, Figs. 5-7.) Heckerman then uses time stamps associated with the confirmed silences in the audio stream to generate synchronized text/audio files. (*Id.*; EX-1002 ¶132.)

A POSITA would have been motivated to combine Heckerman’s method for segmenting an audio stream with Prabhakaran’s segmented audio stream for several reasons. (EX-1002 ¶¶133-38.)

First, a POSITA would have understood that segmenting the audio at a period of silence would have led to smoother audio rendering. (*Id.* ¶134.)

Second, a POSITA would have been “motivated to do so in order to exploit acoustic cues to segment recorded speech into semantically meaningful chunks.” (EX-1074, 5; EX-1002 ¶135.)

Third, the combination represents the simple addition of one known element (e.g., segmenting audio based on natural language gaps) to another known element (e.g., Prabhakaran’s segmented audio) to obtain predictable results (e.g., smoother audio rendering). (EX-1002 ¶136); *KSR*, 550 U.S. at 417.

Fourth, the combination uses a known technique for segmenting audio (e.g., based on natural language gaps) to improve a similar method (e.g., Prabhakaran’s) in the same way. (*Id.*)

Fifth, the combination applies a known technique (e.g., Heckerman’s segmenting audio based on natural language gaps) to a known method (e.g., Prabhakaran’s) that is ready for improvement and yields predictable results (e.g., smoother audio rendering). (*Id.*)

Finally, according to a European patent Examiner evaluating a similar claim in a foreign equivalent of the ’111 patent, segmenting audio based on natural language gaps was “merely one of several alternative straightforward possibilities which the skilled person would select[,]” “notoriously well-known” (EX-1072, 5), and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶¶137-38 (citing EX-1069, 2:4-6; EX-1063, 113-15; EX-1075, 9-10).)

XI. GROUND 1E: CLAIMS 8-11 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND DTB.

A. Claim 8

Claim 8 depends from claim 1 and recites “setting a second time offset as a bookmarked position, wherein the second time offset corresponds to a second position on the first page or a position in the timeline of the audio stream.”

DTB explains how to “set bookmarks [] at any point in a [digital talking book], whether based on the audio file or the textual content file.” (EX-1003, 45-50.) The bookmark positions are identified using a time offset that corresponds to “the exact position ...in the audio file” and therefore correspond to a position in the audio

stream timeline. (*Id.*; EX-1002 ¶143.) Because the audio file and text/image files are linked, the time offsets in the bookmark file also correspond to the text or images being rendered on the page. (EX-1002 ¶143.)

For example, if the first page of Prabhakaran’s presentation comprised Image 1 and Audio 1, and its duration (from t_0 to t_1) was 10 minutes, a user may set a bookmark (as taught by DTB) at 5 minutes. (*Id.* ¶144.) This combination would disclose or renders obvious setting a second time offset ($t_{0.5}$) as a bookmarked position, wherein the second time offset ($t_{0.5}$) corresponds to a second position (halfway through page duration or second spatial position) on the first page or a position in the timeline of the audio stream ($t_{0.5}$ or 5 minutes). (*Id.*)

A POSITA would have been motivated to combine DTB’s bookmark functionality with Prabhakaran’s multimedia presentation method for several reasons. (*Id.* ¶¶145-54.)

First, Prabhakaran discloses that a “a user may interact by giving inputs such as ... freeze and restart[.]” (EX-1026, 178.) A POSITA would have understood that a user freezing a presentation would be motivated to set a bookmark because a bookmark would allow the user to “restart the presentation” at the same point where the user left off. (EX-1002 ¶146.)

Second, a POSITA would have looked to an industry standard like DTB when evaluating what features to implement in its system. (*Id.* ¶147.)

Third, DTB teaches other benefits of bookmarking, including the ability to export bookmarks to other devices (so the user can resume at the same point later on a different device) and other users (for sharing content). (EX-1003, 45; EX-1002 ¶148.) A POSITA would have understood that bookmarking saves time and frustration by starting from a bookmarked position instead of having to redownload and replay content. (EX-1002 ¶148.) A POSITA also would have understood that bookmarking allows users to identify the specific portion of media they wish to share with other users which would reduce confusion and save time on behalf of the user receiving the bookmarked content. (*Id.*)

Fourth, the combination represents the simple addition of one known element (e.g., bookmark functionality) to another known element (e.g., Prabhakaran's presentations) to obtain predictable results (e.g., restart presentation where the user left off). (*Id.* ¶149); *KSR*, 550 U.S. at 417.

Fifth, the combination represents the use of a known technique (e.g., DTB's bookmark functionality) to improve a similar method (e.g., Prabhakaran's presentation) in the same way. (*Id.*)

Sixth, the combination applies a known technique (e.g., DTB's bookmark functionality) to a known method (e.g., Prabhakaran's) that is ready for improvement and yields predictable results (e.g., restart presentation where user left off). (*Id.*)

A POSITA would have reasonably expected success in making the modification because bookmarking was well-known, disclosed in numerous prior art references, and routine to a POSITA. (EX-1002 ¶¶150-53 (citing EX-1060; EX-1069; EX-1062; EX-1033).) Moreover, DTB reflects an industry standard, and a POSITA would have expected success in implementing Prabhakaran’s multimedia presentation using the industry standard. (EX-1002 ¶154.)

B. Claim 9

Claim 9 depends from claim 8 and recites “transmitting the bookmarked position to a network accessible server device or a second other client device.”

DTB discloses this limitation because it discloses that playback systems can “export the [bookmarks] to other compliant playback devices.” (EX-1003, 45.)

A POSITA would have been motivated to combine DTB’s transmittable bookmark functionality with Prabhakaran’s multimedia presentation method for the same reasons discussed above. (*Supra* §XI.A.) It would also allow users to share passages from books with others and it would allow the user to restart the presentation from the bookmarked position on another device. (EX-1002 ¶159.) Moreover, it would reflect the simple addition of one known element (e.g., transmittable bookmark functionality) to another known element (e.g., Prabhakaran’s distributed multimedia presentations) to obtain predictable results (e.g., ability for the same or a different user to restart from bookmarked position on different device); it represents

the use of a known technique (transmittable bookmark) to improve a similar method (Prabhakaran’s presentations) in the same way; and it applies a known technique (transmittable bookmark) to a known method (Prabhakaran’s presentations) that is ready for improvement and yields predictable results (e.g., allowing users to resume presentation on a different device). (*Id.*); *KSR*, 550 U.S. at 417.

A POSITA would have reasonably expected success in making the modification because transmitting bookmarks was well-known, disclosed in numerous prior art references, and routine to a POSITA. (EX-1002 ¶¶160-64 (citing EX-1062; EX-1069; EX-1060; EX-1004).)

C. Claims 10 and 11

Claim 10 depends from claim 1 and recites “wherein the time information for the one or more static graphic images is external to the audio stream.” Claim 11 depends from claim 10 and recites “wherein the time information for the one or more static graphic images is stored in a descriptor file.”

Prabhakaran discloses that the order of objects in the time domain “is explicitly formulated and stored along with the multimedia objects.” (EX-1026, 177, 213.) Prabhakaran therefore discloses or renders obvious these limitations. (EX-1002 ¶167.)

This limitation also would have been obvious in view of DTB. (EX-1002 ¶168.) DTB discloses a synchronization file, referred to as a SMIL file, that is used

to synchronize audio, text, and image files to provide a multimedia presentation (e.g., an audiobook). (EX-1003, 3; *see also id.*, 21, 61, 23-34.) DTB’s SMIL file includes “parallels” that describe the sequence of content and what content should be rendered simultaneously. (*Id.*, 21-27.) The parallels include “time-based positions” represented as clipBegin and clipEnd attributes for any audio rendered in the parallel. (*Id.*) Images referenced in the parallel are rendered at the same time as the audio. (*Id.*)

As shown below, DTB’s SMIL file (highlighted in yellow) is separate from (e.g., external to) to the [audio stream](#) (in blue):

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

(*Id.*, 14-15; EX-1002 ¶169.) DTB therefore discloses or renders obvious time information (e.g., parallels and audio clip time information) stored in a descriptor file (SMIL file) that is external to the audio stream. (EX-1002 ¶169.)

A POSITA would have been motivated to combine DTB’s external descriptor file with Prabhakaran’s stored temporal relations for several reasons. (*Id.* ¶¶170-75.)

First, a POSITA would have been motivated to adhere to a standard for synchronizing distributed multimedia. (*See* EX-1003, 21.) A POSITA would also have been motivated to use SMIL’s external descriptor file because, as a standard, it allowed for consistent rendering across more devices. (*Id.*; EX-1006 (Yoshimura), 1779 (POSITA could “easily create SMIL files for describing streaming content because SMIL is a standard and popular scene description language”); EX-1002 ¶171.)

Second, Prabhakaran discloses “pre-defined temporal relations” that are “stored along with the multimedia database.” (EX-1026, 213.) This would have incentivized a POSITA to look for other references, such as DTB, that describe how to store such temporal relations. (EX-1002 ¶172.)

Third, a POSITA would have understood the benefits of using DTB’s SMIL file, including the ability to index, navigate, synchronize, and retrieve files in large multimedia presentations. (*Id.* ¶173.)

Fourth, the combination represents nothing more than the simple addition of one known element (e.g., DTB’s descriptor file) to another known element (e.g., Prabhakaran’s stored, pre-defined temporal relations) to obtain predictable results (e.g., rendering consistent with a standard). (*Id.* ¶174); *KSR*, 550 U.S. at 417.

Fifth, the combination represents the use of a known technique (e.g., DTB’s descriptor file) to improve a similar method (e.g., Prabhakaran’s stored, pre-defined temporal relations) in the same way. (*Id.*)

Sixth, the combination applies a known technique (e.g., DTB’s descriptor file) to a known method (e.g., Prabhakaran’s stored, pre-defined temporal relations) that is ready for improvement and yields predictable results (e.g., allowing for rendering according to a standard). (*Id.*)

A POSITA would have reasonably expected success in making the modification because, as a standard, DTB’s descriptor file was well-known and commonly used by a POSITA. (EX-1006, 1779 (POSITA could “easily create SMIL files for describing streaming content because SMIL is a standard and popular scene description language”); EX-1002 ¶175 (citing EX-1003, 21).)

XII. GROUND 1F: CLAIMS 12-15 WOULD HAVE BEEN OBVIOUS IN VIEW OF PRABHAKARAN, DUNCAN, AND DR. SEUSS.

A. Claim 12

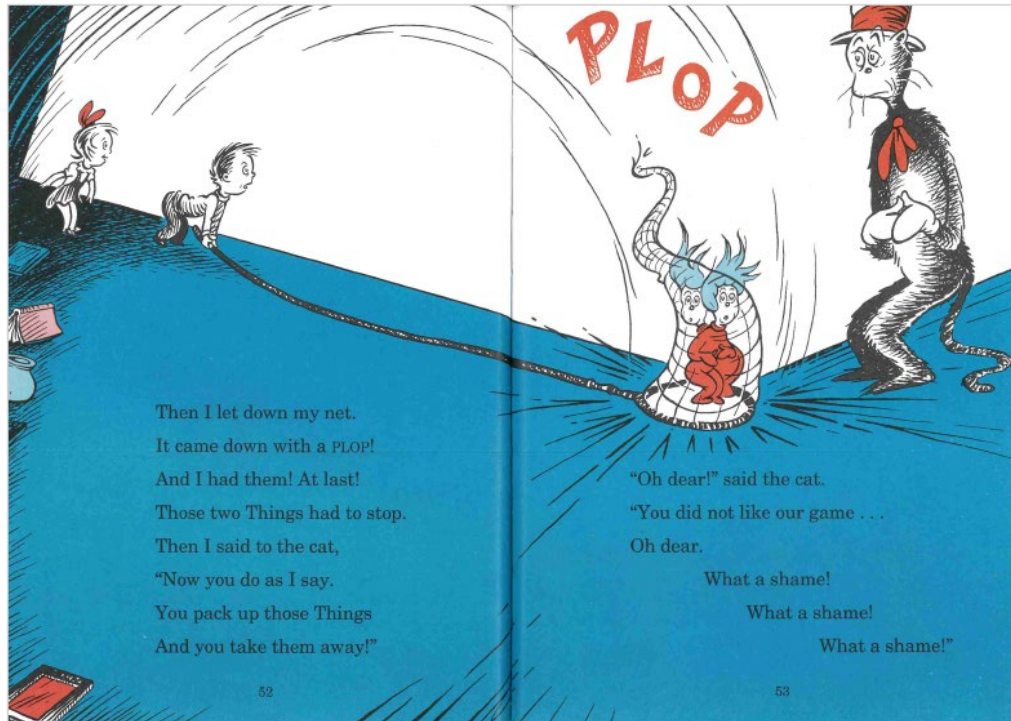
Claim 12 depends from claim 1 and recites “wherein the one or more static graphic images contain elements that form text characters when rendered.” This limitation recites the content of the information displayed (e.g., text characters) and is therefore not entitled to patentable weight. *SAP Am., Inc. v. InvestPic, LLC*, 898 F.3d 1161, 1169 (Fed. Cir. 2018); *see also In re Marco Guldenaar Holding B.V.*, 911 F.3d 1157, 1162 (Fed. Cir. 2018) (printed matter not inventive).

Regardless, the claim is obvious. Images containing text were ubiquitous (e.g., labels, advertisements, logos, infographics, menus, and picture books).

(EX-1002 ¶180.) It would have been obvious to a POSITA that Prabhakaran's images (e.g., Image 1) would contain elements that form text characters when rendered. (*Id.*) Indeed, this would occur each time that Prabhakaran's system is used to render an image (e.g., presentation title page, book cover, album cover, etc.) that includes text. (*Id.*)

Even if Prabhakaran did not render this limitation obvious, it would have been obvious in view of *Dr. Seuss*. (EX-1031, 1); *Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc.*, 109 F.3d 1394, 1396 (9th Cir. 1997). As illustrated below, Dr. Seuss includes a variety of images that contain text characters:





(EX-1031, 4-5, 52-53; EX-1002 ¶181.)

Dr. Seuss therefore discloses or renders obvious one or more static graphic images containing elements that form text characters (e.g., “BUMP!” and/or “PLOP”) when rendered. (EX-1002 ¶182.)

A POSITA would have been motivated for the following reasons to use the text and images along with a recorded audio narration of *The Cat in the Hat*, such as the 2003 recording narrated by actor Kelsey Grammer, to create a digital talking book that renders *The Cat in the Hat* through Prabhakaran’s system. (See EX-1079; EX-1002 ¶¶183-86.)

First, a POSITA would have been motivated to use Prabhakaran’s system to deliver multimedia content for books that contained images comprising text and corresponding text passages, as in the Dr. Seuss reference. Indeed, Dr. Seuss would have been a natural choice for content because “The Cat in the Hat” and other Dr. Seuss books were widely popular, selling more than 35 million copies between 1931 and 1991. (EX-1002 ¶184); *Dr. Seuss*, 109 F.3d at 1396, 1403 (“[t]he good will and reputation associated with Dr. Seuss’ work is substantial”).

Second, the combination represents nothing more than the simple addition of one known element (e.g., Dr. Seuss’s images containing text) to another known element (e.g., Prabhakaran’s image stream) to obtain predictable results (e.g., allowing text in images to be synchronized with other forms of media like plain text). (EX-1002 ¶185); *KSR*, 550 U.S. at 417.

Third, the combination represents the use of a known technique (e.g., rendering images that include text) to improve a similar method (e.g., Prabhakaran’s image stream) in the same way. (*Id.*)

Fourth, the combination applies a known technique (e.g., including text in images) to a known method (e.g., Prabhakaran’s image stream) that is ready for improvement and yields predictable results (e.g., synchronizing images with text with other forms of media). (*Id.*)

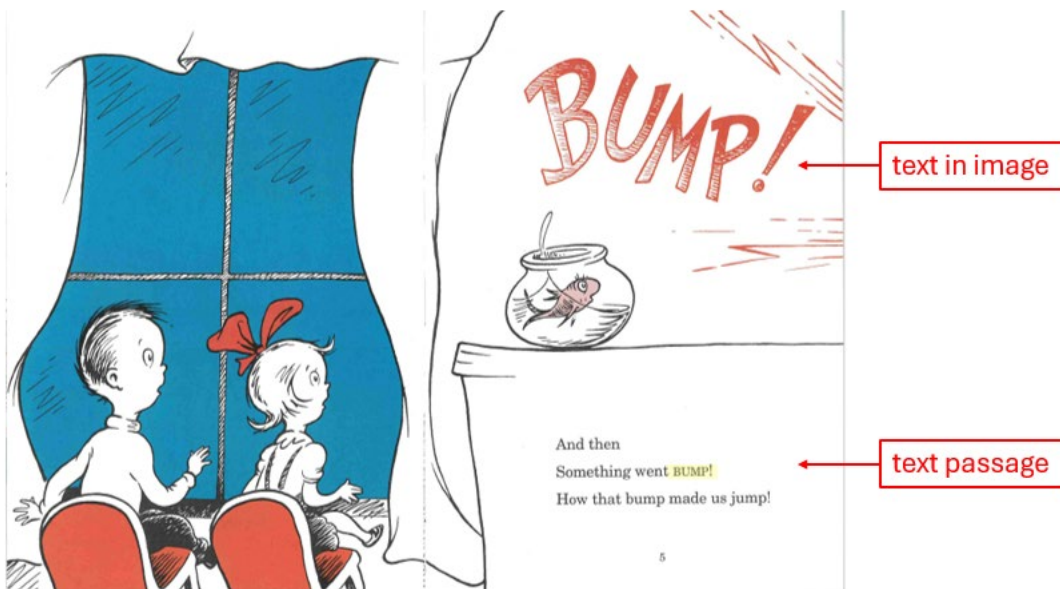
Finally, a POSITA would have reasonably expected success because Dr. Seuss merely provides the content and does not alter the system in any way. (EX-1002 ¶186.)

B. Claim 13

1. 13[a]: Rendering Text Passages

Claim 13 depends from claim 12. Element 13[a] recites “rendering on the first page one or more text passages representing the text characters formed by the elements in the one or more static graphic images.” This limitation is directed to the content being displayed and is therefore not entitled to patentable weight. (*Supra* §XII.A.)

As illustrated below, Dr. Seuss comprises text passages representing the text in the image:



(EX-1031, 4-5.)

Prabhakaran alone or in combination with Dr. Seuss renders obvious rendering on the first page one or more text passages representing the text characters formed by the elements in the one or more static graphic images. (EX-1002 ¶¶190-91.) A POSITA would have been motivated to combine the references, and reasonably expected success. (*Supra* §XII.A; EX-1002 ¶191.)

2. 13[b]: Assigning Time Information

Element 13[b] recites “assigning to each of the one or more text passages time information including at least one of a start time, an end time, and a duration relative to the timeline of the audio stream, wherein the time information for each text passage is determined on the basis of the time information for the one or more static graphic images.”

Prabhakaran discloses that text objects have pre-defined temporal information. (*Supra* §VII.A.3; EX-1002 ¶194.) A POSITA would have understood that a text passage representing text in an image would be assigned the same temporal information as the image with which it is associated. (EX-1002 ¶194.)

Prabhakaran, alone or in combination with Dr. Seuss, therefore renders obvious assigning to each of the one or more text passages (e.g., text), time information (e.g., temporal information) including at least one of a start time, an end time, and a duration (e.g., t_0 to t_1) relative to the timeline of the audio stream, wherein the time information for each text passage (e.g., text) is determined on the basis of the time

information (e.g., t0 to t1) for the one or more static graphic images (e.g., start and end time of Image 1). (*Id.* ¶195.)

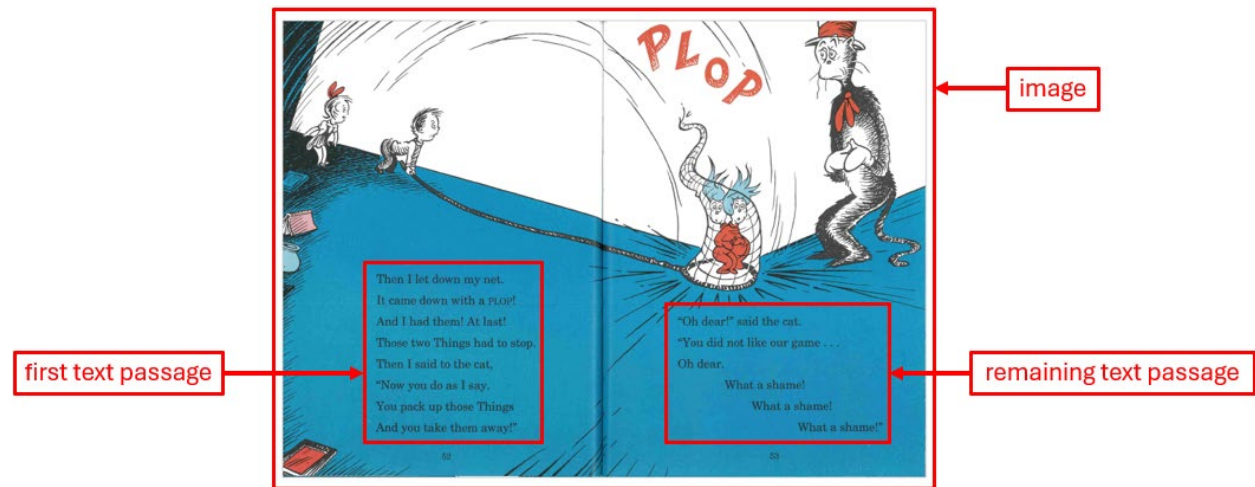
C. Claim 14

Claim 14 depends from claim 13 and recites “wherein only a portion of a first text passage associated with a first static graphic image is rendered on the first page.”

Duncan discloses that increasing the font size of text necessitates repagination. (*Supra* §VII.A.4; EX-1002 ¶198.) A POSITA would have understood that a text passage would spread across two pages if the font size was sufficiently increased, leaving only a portion of the text passage rendered on the first page. (EX-1001, 40:54-67; EX-1002 ¶198.)

Prabhakaran in view of Duncan therefore renders obvious that only a portion of a first text passage associated with a first static graphic image is rendered on the first page. (EX-1002 ¶199.) A POSITA would have been motivated to combine Duncan’s pagination method with Prabhakaran’s multimedia presentation. (*Supra* §VII.A.4; EX-1002 ¶199.)

Dr. Seuss also illustrates an example wherein only a portion of a first text passage associated with a first static graphic image is rendered on the first page:



(EX-1031, 52-53; EX-1002 ¶200.)

D. Claim 15

Claim 15 depends from claim 14 and recites “wherein a remaining portion of the first text passage is rendered on a second page, wherein the first static graphic image is replicated on the second page in association with the remaining portion of the first text passage, and wherein the second page is repaginated.”

Duncan discloses repagination when a font size increase pushes text onto a second page. (EX-1027 ¶[0042].) A POSITA would have understood that the image associated with the text passage would appear on all pages the text passage spans to maintain synchronization between the text passage and image. (EX-1001, 40:54-67; EX-1002 ¶204.) Furthermore, a POSITA would have desired this result because the viewer would want to keep viewing the image along with its associated text. (*Id.*)

Prabhakaran in view of Duncan therefore renders obvious that a remaining portion of the first text passage is rendered on a second page, wherein the first static graphic image is replicated on the second page in association with the remaining portion of the first text passage, and wherein the second page is repaginated. (*Supra* §VII.A.4; EX-1002 ¶205.) A POSITA would have been motivated to combine Duncan's pagination method with Prabhakaran's multimedia presentation. (*Id.*)

Moreover, Dr. Seuss renders this obvious because, as shown above for claim 14, a POSITA would have understood that the remaining portion of the first text passage would be rendered on a second page. It would have been obvious to repeat the first static graphic image on both pages in an e-reader format. (EX-1002 ¶206.) This page would be repaginated as described by Duncan. (*Id.*)

XIII. GROUND 2A: CLAIMS 1-2, 4-5, 8-11, AND 16 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, AND DUNCAN.

A. Claim 1

1. Preamble

DTB discloses a method of creating digital talking books that can be accessed over the Internet and rendered on a playback device. (EX-1003, vii, 21-22.) Accordingly, to the extent the preamble is limiting, it is disclosed or rendered obvious by DTB. (EX-1002 ¶210.)

2. 1[a]: Network Accessible Library

DTB specifies a standard for multimedia presentations accessed online for playback on a playback device. (EX-1003, vii (“The DTB [] can include not just the audio rendition of the work, but the full textual content and images as well.”), 3 (SMIL is “a standard for [] playback of multimedia presentations over the Internet”), 21, 60, 97-98.)

DTB therefore discloses or renders obvious a client device (e.g., playback device) accessing (e.g., over the Internet) a network accessible library (e.g., content accessible over the Internet) having stored thereon an image stream of static graphic images and corresponding audio (e.g., sequence of audio and image files corresponding as specified in parallels). (EX-1002 ¶¶212-13.)

Even if DTB did not disclose or render obvious this limitation, it would have been obvious in view of Steele. (*Id.* ¶¶214-20.) Steele discloses networked multimedia databases containing SMIL files and associated content:

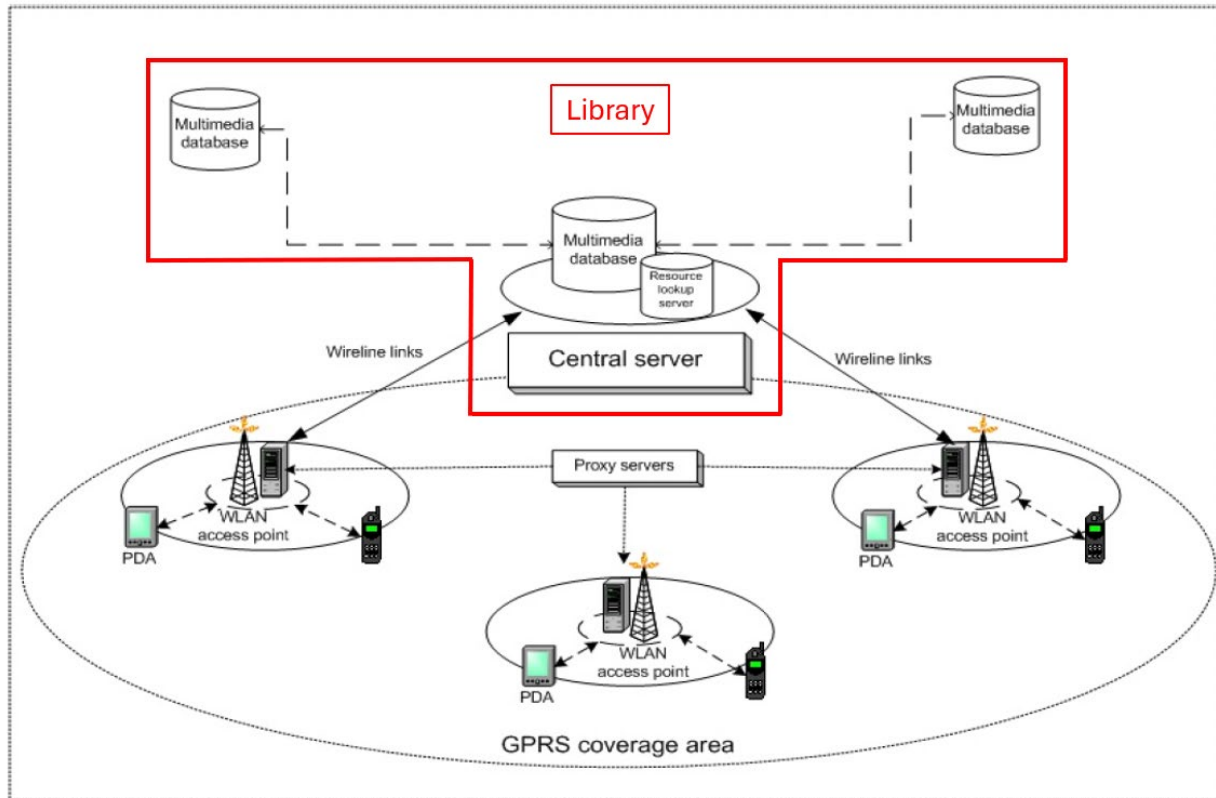


Figure 2: System architecture

(EX-1029, Fig. 2; EX-1002 ¶214.) Steele further discloses that “SMIL allows for the sending of [] still images, and sound” comprising a presentation. (EX-1029, 1.) The multimedia files are stored on servers (*id.*, 2), which are a network accessible library.

A POSITA would have been motivated to combine Steele’s network accessible multimedia library (database/servers) with DTB for several reasons. (EX-1002 ¶¶215-19.)

First, DTB discloses “playback of multimedia presentations over the Internet” using SMIL files (EX-1003, 3), which would have incentivized a POSITA to look

to other references, such as Steele, for methods to store SMIL files and the multimedia content referenced by SMIL files in Internet-accessible locations. (EX-1002 ¶216.)

Second, DTB is a SMIL-based system and Steele discloses storing and distributing SMIL files and their associated content. (EX-1002 ¶217.)

Third, the combination represents the simple addition of one known element (e.g., Steele’s Internet-accessible database) to another known element (e.g., DTB’s internet-based playback system) to obtain predictable results (e.g., accessing multimedia content online). (EX-1002 ¶218); *KSR*, 550 U.S. at 417.

Fourth, the combination represents the use of a known technique (e.g., Steele’s networked multimedia database) to improve a similar method (e.g., DTB’s multimedia playback over the Internet) in the same way. (*Id.*)

Fifth, the combination applies a known technique (e.g., Steele’s networked multimedia database) to a known method (e.g., DTB’s multimedia playback over the Internet) that is ready for improvement and yields predictable results (e.g., accessing multimedia content online). (*Id.*)

Finally, downloading content from network accessible multimedia libraries was commonplace, routine, and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶219 (citing

EX-1004 ¶[0010], Fig. 1; EX-1027 ¶[0023]; EX-1028 ¶¶[0004], [0185]; EX-1032, 12:53-67; EX-1061, 3:44-57.)

3. 1[b]: Downloading Images Associated with Time Information

DTB synchronizes images and audio elements using “time containers” called parallels (“<par>”). (EX-1003, 21, 26-27.) Parallels define a “time grouping in which multiple media objects play back at the same time.” (*Id.*, 21; *see also id.*, 26.) The objects may include the “audio rendition of the work” as well as “full textual content and images.” (*Id.*, vii; *see also id.*, 26-27.) The parallels include pointers to the audio and image files referenced in the parallel. (*Id.*; EX-1002 ¶222.)

Parallels with audio clips include “clipBegin” and “clipEnd” attributes specifying the start and end “of a continuous audio file as a time offset from the start of the audio file” to be rendered. (EX-1003, 27.) An image in a parallel with an audio clip is displayed while the audio clip is rendered; therefore, the image is associated with the clipBegin and clipEnd of the audio clip. (*Id.*; *see also id.*, 6, 30 (“[d]uration of image display will be equal to that of the longest media object or time container contained within the same par”), 32; EX-1002 ¶223.)

DTB’s Example 7.2 below shows a sample implementation of SMIL for an image and its associated caption and producer’s note. (EX-1003, 32-33.) “The image file is presented in parallel with text and audio versions of the figure caption and

a producer’s note describing the figure.” (*Id.*) “The entire group is wrapped in a par, with the image file rendered simultaneously with a sequence of two pars.” (*Id.*)

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

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

(*Id.*, 32; EX-1002 ¶224.)

It would have been obvious to a POSITA to download image and audio clips referenced by a SMIL file to the DTB player. (EX-1003, vii, 21 (“playback of multimedia presentations over the Internet”); EX-1002 ¶225.)

DTB therefore discloses or renders obvious downloading to the client device (e.g., playback system) from the network accessible library (e.g., server) one or more static graphic images from the image stream (e.g., images identified in image element), each associated with time information including at least one of a start time (e.g., clipBegin), an end time (e.g., clipEnd), and a duration (e.g., clipBegin to clipEnd) relative to a timeline of the audio stream (e.g., an audio file). (EX-1002 ¶226.)

Even if DTB did not render obvious this limitation, it would have been obvious in view of Steele. (EX-1002 ¶227.) Steele discloses that, rather than embedding all the multimedia files, a SMIL file may “just provide[] references (URLs) for the multimedia files and these can be *downloaded* separately from these specified URLs.” (EX-1029, 2 (emphasis added); *see also id.*, 1 (multimedia includes images and sound), 1 (multimedia presentations need to be downloaded and viewed incrementally).)

A POSITA would have been motivated to combine Steele’s network accessible multimedia library storing content available for download with DTB’s method. (*Supra* §XIII.A.2; EX-1002 ¶228.)

4. 1[c]: Assembling a First Page

DTB discloses assembling a page on a client device in at least two ways. (EX-1003, 6-7, 23-25, 32, 59-60; EX-1002 ¶231.)

First, DTB discloses that the player assembles a page because the player renders the elements in the parallel, e.g., an image and text. For example, when implementing DTB’s Example 7.2, the player would assemble a first page comprising at least an image (fig1.png) and text (rs.xml#caption_1):

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

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

(EX-1003, 32; EX-1002 ¶232.)

Second, like the '111 patent, DTB's players use style sheets to control presentation styles and assemble pages. (*Compare* EX-1001, 29:7-14 *with* EX-1003, 59-60 ("Developers of playback systems may implement user interface features that support local control of style sheets, thereby allowing the user to define styles that supersede default player- or producer-defined styles."), 7, 4 ("[i]t is in the best interest of users that ... players make information available in as many presentation modes as practical, including ... large print with user-specifiable size and text re-wrapping for players with visual display"); EX-1002 ¶233.) A POSITA would have understood that a DTB player would use the style sheets (which control, e.g., image size and number) to assemble a page. (EX-1002 ¶233.)

Third, playback devices must support SMIL elements and attributes such as <layout>, <region>, and <customTest>, all of which impact page assembly. (EX-1003, 23-25.) Layout controls "where on a visual, audio, or tactile rendering

space various producer-defined elements, e.g., figures, text, footnotes, etc. are displayed.” (*Id.*, 24.) Region controls “the position, size, and scaling of media objects (e.g., text, img).” (*Id.*) CustomTest defines “the kinds of structures (e.g., page numbers, notes, line numbers, etc.) which the user can choose to have presented or skipped during normal playback of a DTB.” (*Id.*, 25.) Playback devices parse the SMIL file and assemble pages based on these instructions. (EX-1002 ¶234.) Thus, DTB discloses that the client device assembles a first page from at least one image. (*Id.*)

Even if DTB did not disclose or render obvious this limitation, it would have been obvious in view of Duncan. (*Supra* §VII.A.4.) A POSITA would have been motivated to combine Duncan’s paginator with DTB for several reasons. (EX-1002 ¶¶235-39.)

First, both references seek to “accommodate sight-impaired readers.” (EX-1027 ¶[0004]; EX-1003, vii.) It would have been obvious to a POSITA that Duncan’s pagination method would have made DTB’s system more accessible to print-disabled persons, which was a primary motivation of the DTB standard. (EX-1003, Abstract; EX-1027 ¶¶[0004], [0042]; EX-1002 ¶236.)

Second, DTB discloses the use of style sheets for rendering an electronic book. (EX-1003, 60.) This would have incentivized a POSITA to look for methods, like Duncan’s, using style sheets in electronic book devices. (EX-1002 ¶237.)

Third, the combination represents nothing more than the addition of one known element (e.g., Duncan’s pagination method) to another known element (e.g., DTB’s style sheets) to obtain predictable results (e.g., displaying one or more images on a page with proper spatial organization). (*Id.* ¶238); *KSR*, 550 U.S. at 417.

Fourth, the combination represents the use of a known technique (e.g., Duncan’s pagination method) to improve a similar method (e.g., DTB’s style sheets) in the same way. (*Id.*)

Fifth, the combination applies a known technique (e.g., Duncan’s pagination method) to a known method (e.g., DTB’s style sheets) that is ready for improvement and yields predictable results (e.g., displaying image(s) on a page with proper spatial organization). (*Id.*) The combination would improve the user experience because users could customize the viewing experience. (EX-1002 ¶238.)

Finally, methods involving pagination were commonplace and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (*Supra* §VII.A.4; EX-1002 ¶239.)

5. 1[d]: Assigning Time Information to First Page

DTB players assign time information to a page in at least two ways. (EX-1002 ¶¶242-46.)

**a. Assigning Time Information by
Assembling a Page**

DTB (alone or in combination with Duncan) discloses assembling a page. (*Supra* §XIII.A.4; EX-1002 ¶243.) A POSITA would have understood that, when a DTB player applies parses the SMIL file to assemble a page, it assigns time information including a start time, end time, and duration to the page. (*Id.*)

Even if DTB did not disclose or render obvious this limitation, it would have been obvious in view of Duncan. (*Supra* §VII.A.5.) A POSITA would have understood that where a page includes one image, the client device determines the duration of that display to be the difference between the start and end times (clipBegin; clipEnd) of that image. (EX-1002 ¶244.) Likewise, a POSITA would have understood that where the page includes multiple images (*supra* §XIII.A.4), the client device determines the duration of that display to be the difference between the start time of the first image and the end time of the last image. (EX-1027 ¶[0028]; EX-1002 ¶244.)

A POSITA would have been motivated to use a paginator (e.g., Duncan's) with synchronized pre-recorded audio in a system like DTB (e.g., where visual content like images is pre-linked with pre-recorded audio) with a reasonable expectation of success. (*Supra* §§XIII.A.4, VII.A.5.a; EX-1002 ¶245.)

**b. Assigning Time Information by Changing
the Presentation Speed**

DTB strongly recommends that “playback systems implement Time-Scale Modification (TSM) to enable user control of playback speed.” (EX-1003, 3; *see also id.*, 63, 22.) When the user adjusts playback speed, the client device assigns new time information for each page to maintain synchronization with the accelerated audio. (*Supra* §VII.A.5.b; EX-1002 ¶246.)

6. 1[e]: Downloading Audio

Audio content in a parallel would be downloaded in the same manner as the image content. (*Supra* §XIII.A.3; EX-1002 ¶249.)

DTB (alone or in combination with Steele) therefore discloses or renders obvious downloading to the client device (e.g., playback system) from the network accessible library a portion of the audio stream (e.g., audio clip) including a first time offset (e.g., clipBegin), wherein the first time offset corresponds to a first position on the first page. (*Supra* §§XIII.A.3 (downloading disclosed or obvious in view of DTB and/or Steele), VII.A.6 ; EX-1002 ¶250.)

Even if DTB did not render a first position on the first page obvious, it would have been obvious in view of Duncan. (*Supra* §XIII.A.5.) A POSITA would have been motivated to combine Duncan’s first position on a first page with DTB’s page of one or more images. (*Id.*; EX-1002 ¶251.)

7. 1[f]: Simultaneously Rendering Images and Audio

DTB players simultaneously present images with corresponding audio clips using synchronization information in the SMIL file. (EX-1003, vii, 6, 7; EX-1002 ¶254.)

Because the audio stream and associated image are associated by the parallel, the audio is rendered in dependence on the clipBegin/clipEnd information in the parallel. (EX-1002 ¶255.)

DTB, in combination with Duncan and Steele, therefore discloses or renders obvious simultaneously rendering (e.g., simultaneously presenting images with corresponding audio clips) the first page and the portion of the audio stream on the client device (e.g., playback device) by using the time information for the one or more static graphic images or for the first page, wherein the portion of the audio stream is rendered in dependence upon the first time offset. (*Supra* §§XIII.A.3 (image time information), XIII.A.4 (first page), XIII.A.5 (first page time information), XIII.A.6 (audio portion with first time offset); EX-1002 ¶256.)

B. Claim 2

Steele discloses “[m]ultiple servers storing multimedia files” where “[c]ertain multimedia files may be duplicated at many servers.” (EX-1029, 2.) Steele therefore discloses or renders obvious that the network accessible library includes server devices. (EX-1002 ¶260.) A POSITA would have been motivated to combine

Steele’s servers with DTB’s multimedia playback over the Internet. (*Id.*; *supra* §XIII.A.2.)

C. Claim 4

This limitation would have been obvious in view of DTB. For example, DTB shows a first time offset (clipBegin=“00:00:00”) that is the start time of a first page (p1):

```
<par id="p1" customTest="pagenum">  
  <text region="text" src="rs.xml#pg_1" />  
  <audio src="rs_fwdx.mp3" clipBegin="00:00:00"  
    clipEnd="00:00:00.91" />  
</par>
```

(EX-1003, 32; EX-1002 ¶264.) It would have been obvious to a POSITA to include in the parallel above an image element instead of a text element. (EX-1002 ¶264.)

Alternatively, DTB’s bookmark points to the ID of a parallel that contains a time offset. (EX-1003, 48.)

DTB therefore discloses or renders obvious a first time offset that is the start time (e.g., clipBegin=“00:00:00”) of the first page (e.g., p1), a start time of the one or more static graphic images, or a bookmarked position (e.g., bookmark pointing to a time offset). (EX-1002 ¶¶264-66.)

D. Claim 5

DTB (alone or in combination with Duncan) discloses or renders obvious assembling a second page (*supra* §XIII.A.4; EX-1003, cover, 2; EX-1002 ¶¶269-70) and assigning time information to a second page (*supra* §XIII.A.5; EX-1002 ¶270).

E. Claims 8-11

DTB discloses or renders obvious the limitations of claims 8-11. (*Supra* §XI; EX-1002 ¶¶272-73.)

F. Claim 16

DTB discloses a playback device that can parse a SMIL file and render a digital talking book. (EX-1003, Abstract; EX-1002 ¶274.) DTB therefore discloses the preamble. (EX-1002 ¶274.)

DTB, Steele, and Duncan together disclose or render obvious claim 16. (*Supra* §§XIII.A; EX-1002 ¶¶274-76.)

XIV. GROUND 2B: CLAIM 3 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, DUNCAN, AND TEMPLEMAN.

Templeman discloses claim 3's limitation. (*Supra* §VIII; EX-1002 ¶277.) A POSITA would have been motivated to combine DTB's method for displaying images with Templeman's modification of the amount of content on a page depending on the size of the content and the size of the screen. (*Id.*)

XV. GROUND 2C: CLAIM 6 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, DUNCAN, AND HAY.

It would have been obvious to a POSITA in view of DTB to automatically advance a digital talking book's page to the next page while audio playback occurs using the timing information in a SMIL file. (EX-1002 ¶279.) This also would have been obvious in view of Duncan or Hay. (*Supra* §IX.)

A POSITA would have been motivated to include automatic page turning (as disclosed by Duncan and Hay) in DTB’s player for several reasons. (EX-1002 ¶¶280-84.)

First, a POSITA would have understood that automatically turning the page would improve the reader’s experience and be more convenient. (*Supra* §IX; EX-1002 ¶281.)

Second, DTB seeks to “accommodate sight-impaired readers.” (EX-1003, vii.) It would have been obvious to a POSITA that automatic page turning would have made DTB’s system more accessible to sight-disabled persons who rely primarily on the audio narration of a book and prefer not to be burdened by pressing a button to advance to the next page. (EX-1002 ¶282.)

Third, the combination represents nothing more than the simple addition of one known element (e.g., automatic page turning) to another known element (e.g., DTB’s rendering a digital talking book) to obtain predictable results (e.g., a continuous audio playback). (*Id.* ¶283); *KSR*, 550 U.S. at 417.

Fourth, the combination represents the use of a known technique (e.g., automatic page turning) to improve a similar method (e.g., DTB’s rendering a digital talking book) in the same way. (*Id.*)

Fifth, the combination applies a known technique (e.g., automatic page turning) to a known method (e.g., DTB’s rendering digital talking book) that is ready for

improvement and yields predictable results (e.g., continuous playback). (*Id.*) The combination would improve the user experience because it would eliminate any need to initiate a page turn. (EX-1002 ¶283.)

Finally, methods to automate page turning were commonplace and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (*Supra* §IX; EX-1002 ¶284.)

XVI. GROUND 2D: CLAIM 7 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, DUNCAN, HAY, AND PRABHAKARAN.

DTB discloses synchronization down to the “paragraph, sentence, or even word level.” (EX-1003, 21.) A POSITA would have understood that the audio is segmented based on natural language gaps (e.g., gaps between words). (EX-1002 ¶286.) DTB therefore renders obvious rendering the portion of the audio stream without any perceptible interruption. (*Id.*)

It also would have been obvious in view of Prabhakaran. (*Supra* §X; EX-1002 ¶287.)

A POSITA would have been motivated to combine Prabhakaran’s retrieval schedule with DTB’s method for several reasons. (EX-1002 ¶¶288-91.)

First, a POSITA would have understood that obtaining multimedia from the Internet would involve delivery delays. (EX-1003, 21; EX-1002 ¶289.) A POSITA would have understood Prabhakaran’s downloading packets of audio on a retrieval schedule to be a viable method for eliminating such delays. (*Id.*)

Second, the combination represents nothing more than the simple addition of one known element (e.g., Prabhakaran’s retrieval schedule) to another known element (e.g., DTB’s playback over the Internet) to obtain predictable results (e.g., an audio presentation free from interruptions). (EX-1002 ¶290); *KSR*, 550 U.S. at 417.

Third, the combination represents the use of a known technique (e.g., Prabhakaran’s retrieval schedule) to improve a similar method (e.g., DTB’s playback over the Internet) in the same way. (*Id.*)

Fourth, the combination applies a known technique (e.g., Prabhakaran’s retrieval schedule) to a known method (e.g., DTB’s playback over the Internet) that is ready for improvement and yields predictable results (e.g., an audio presentation free from interruptions). (*Id.*)

Finally, downloading media streams segmented into small portions was commonplace and disclosed by numerous references, giving a POSITA a reasonable expectation of success in the modification. (EX-1002 ¶291 (citing EX-1004 ¶[0010]; EX-1071 ¶[0045]-[0047]); EX-1073, 2-3.)

XVII.GROUND 2E: CLAIMS 12-15 WOULD HAVE BEEN OBVIOUS IN VIEW OF DTB, STEELE, DUNCAN, AND DR. SEUSS.

DTB alone or in combination with Dr. Seuss renders obvious claims 12-15. (*See supra* §XII; EX-1002 ¶292.) A POSITA would have been motivated to combine Dr. Seuss’s images containing text with DTB’s images for the same reasons a POSITA would have been motivated to combine Prabhakaran with Dr. Seuss. (*Id.*)

XVIII. SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

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

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

Efficiency, fairness, and the merits support institution. *Apple v. Fintiv*, IPR2020-00019, Paper 11 (P.T.A.B. Mar. 20, 2020) (“*Fintiv*”).

A. Factor 1: Potential Stay

On March 20, 2024, PO sued Petitioners for infringement of the ’111 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*” or “the Litigation”). Petitioners will move to stay the Litigation pending resolution of this and related IPRs challenging the patents asserted in the Litigation. The EDVA routinely grants motions to stay cases pending IPR proceedings, including pre-institution, when asserted patents are challenged. *See, e.g., 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 ’111 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 ’111 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 a trial date in the Litigation. The median time to trial in civil cases in the EDVA for 2024 was 14.6 months², but it is clear a longer schedule will be needed here as the case has been pending for nearly 12 months with virtually no activity. (*See infra* §XIX.C.) The Audible DJ Action was filed recently, has no trial date, validity is not currently at issue, 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. 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*,

² 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.

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 complaint in the Litigation on March 20, 2024. Petitioners filed a motion to dismiss or transfer on May 31, 2024 and no hearing date has been set for that motion. Otherwise, the parties have invested very little in the Litigation. Petitioners have not answered the complaint. The parties have not had a Rule 26(f) conference and fact discovery has not yet opened. The parties have not exchanged proposed claim constructions or submitted claim construction briefs. No expert reports have been served. No case schedule has been set. Thus, much work still remains, including fact discovery, claim construction, expert reports, expert discovery, dispositive motions, pretrial motions, and trial. Likewise, there has been no investment in the Audible DJ Action.

Because the remaining investment in the Litigation 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 Litigation and/or Audible DJ Action are stayed, Petitioners could not pursue in those proceedings 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 Litigation 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. 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 Litigation or Audible DJ Action will go to trial before a final written decision 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, as demonstrated above, 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).

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

The Office has not considered Prabhakaran, Duncan, Templeman, Hecker-man, DTB, Steele, and/or Dr. Seuss. 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”). As shown, the references here clearly disclose each claim element. 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).

XXI. CONCLUSION

Amazon requests that the Board institute trial and cancel all challenged claims.

XXII. 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 '111 patent against Petitioners in a patent infringement lawsuit captioned *Audio Pod IP, LLC v. Amazon.com, Inc. et al.*, No. 2:24-cv-00185 (E.D. Va., filed March 20, 2024). Audible also filed a complaint for declaratory judgment of noninfringement of the '111 patent, captioned *Audible, Inc. v. Audio Pod IP, LLC*, No. 1:25-cv-02158 (S.D.N.Y., filed March 14, 2025).

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
<p>Colin B. Heideman (Reg. No. 61,513) 2cbh@knobbe.com BoxSEAZNL2185LP5@knobbe.com</p> <p><u>Postal and Hand-Delivery Address:</u> Knobbe, Martens, Olson, & Bear, LLP 925 4th Ave., Ste. 2500 Seattle, WA 98104 Telephone: (206) 405-2000 Facsimile: (206) 405-2001</p>	<p>Joseph R. Re (Reg. No. 31,291) 2jrr@knobbe.com</p> <p><u>Postal and Hand-Delivery Address:</u> Knobbe, Martens, Olson, & Bear, LLP 2040 Main Street, 14th Floor Irvine, CA 92614 Telephone: (949) 760-0404 Facsimile: (949) 760-9502</p> <p>Christie R.W. Matthaei (Reg. No. 62,933) 2crw@knobbe.com Nathan D. Reeves (Reg. No. 77,806) 2ndr@knobbe.com</p> <p><u>Postal and Hand-Delivery Address:</u> Knobbe, Martens, Olson & Bear, LLP 925 4th Ave., Ste. 2500 Seattle, WA 98104 Telephone: (206) 405-2000 Facsimile: (206) 405-2001</p> <p>Daniel Hughes (Reg. No. 76,592) 2dph@knobbe.com</p> <p><u>Postal and Hand-Delivery Address:</u> Knobbe, Martens, Olson & Bear, LLP 3579 Valley Centre Dr., Ste. 300 San Diego, CA 92130 Telephone: (858) 707-4000 Facsimile: (858) 707-4001</p>

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 BoxSEAZNL2185LP5@knobbe.com.

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

Petitioners certify that the '111 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 sixteen claims is requested. Payment for any additional fees due may be charged to the above-referenced Deposit Account.

Amazon.com, Inc. v. Audio Pod IP, LLC
IPR Petition – U.S. Pat. No. 10,805,111

Respectfully submitted,

KNOBBE MARTENS OLSON & BEAR, LLP

Dated: March 21, 2025

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

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

APPENDIX

Listing of Claims from U.S. 10,805,111	
Claim 1	
1[pre]	A method comprising:
1[a]	accessing, by a client device, a network accessible library having stored thereon an image stream of static graphic images and a corresponding audio stream;
1[b]	downloading to the client device from the network accessible library one or more static graphic images from the image stream, wherein the one or more static graphic images are each associated with time information including at least one of a start time, an end time, and a duration relative to a timeline of the audio stream;
1[c]	assembling, by the client device, a first page from the one or more static graphic images;
1[d]	assigning to the first page, by the client device, time information including at least one of a start time, an end time, and a duration relative to the timeline of the audio stream, wherein the time information for the first page is determined on the basis of the time information for the one or more static graphic images;
1[e]	downloading to the client device from the network accessible library a portion of the audio stream including a first time offset, wherein the first time offset corresponds to a first position on the first page; and

Listing of Claims from U.S. 10,805,111	
1[f]	simultaneously rendering the first page and the portion of the audio stream on the client device by using the time information for the one or more static graphic images or for the first page, wherein the portion of the audio stream is rendered in dependence upon the first time offset.
Claim 2	
--	The method of claim 1, wherein the network accessible library includes one or more server devices.
Claim 3	
--	The method of claim 1, wherein the number of the one or more static graphic images included on the first page is determined on the basis of the size of the one or more static graphic images and the graphic rendering capability of the client device.
Claim 4	
--	The method of claim 1, wherein the first time offset is a start time of the first page, a start time of one of the one or more static graphic images, or a bookmarked position.
Claim 5	
5[a]	The method of claim 1, further comprising:
5[b]	assembling a second page from the one or more static graphic images; and

Listing of Claims from U.S. 10,805,111	
5[c]	assigning to the second page time information including at least one of a start time, an end time, and a duration relative to the timeline of the audio stream, wherein the time information for the second page is determined on the basis of the time information for the one or more static graphic images.
Claim 6	
6[a]	The method of claim 5, further comprising:
6[b]	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 as the portion of the audio stream is rendered.
Claim 7	
--	The method of claim 6, wherein the portion of the audio stream is rendered without any perceptible interruption.
Claim 8	
8[a]	The method of claim 1, further comprising:
8[b]	setting a second time offset as a bookmarked position, wherein the second time offset corresponds to a second position on the first page or a position in the timeline of the audio stream.
Claim 9	
9[a]	The method of claim 8, further comprising:
9[b]	transmitting the bookmarked position to a network accessible server device or a second other client device.

Listing of Claims from U.S. 10,805,111	
Claim 10	
--	The method of claim 1, wherein the time information for the one or more static graphic images is external to the audio stream.
Claim 11	
--	The method of claim 10, wherein the time information for the one or more static graphic images is stored in a descriptor file.
Claim 12	
--	The method of claim 1, wherein the one or more static graphic images contain elements that form text characters when rendered.
Claim 13	
13[a]	The method of claim 12, further comprising:
13[b]	rendering on the first page one or more text passages representing the text characters formed by the elements in the one or more static graphic images; and
13[c]	assigning to each of the one or more text passages time information including at least one of a start time, an end time, and a duration relative to the timeline of the audio stream, wherein the time information for each text passage is determined on the basis of the time information for the one or more static graphic images.

Listing of Claims from U.S. 10,805,111	
Claim 14	
--	The method of claim 13, wherein only a portion of a first text passage associated with a first static graphic image is rendered on the first page.
Claim 15	
--	The method of claim 14, wherein a remaining portion of the first text passage is rendered on a second page, wherein the first static graphic image is replicated on the second page in association with the remaining portion of the first text passage, and wherein the second page is repaginated.
Claim 16	
16[pre]	A non-transitory computer-readable storage medium having stored thereon instructions which, when executed by a processor of a computing device, cause the computing device to:
16[a]	access a network accessible library having stored thereon an image stream of static graphic images and a corresponding audio stream;
16[b]	download from the network accessible library one or more static graphic images from the image stream, wherein the one or more static graphic images are each associated with time information including at least one of a start time, an end time, and a duration relative to a timeline of the audio stream;
16[c]	assemble a first page from the one or more static graphic images;

Listing of Claims from U.S. 10,805,111	
16[d]	assign to the first page time information including at least one of a start time, an end time, and a duration relative to the timeline of the audio stream, wherein the time information for the first page is determined on the basis of the time information for the one or more static graphic images;
16[e]	download from the network accessible library a portion of the audio stream including a first time offset, wherein the first time offset corresponds to a first position on the first page; and
16[f]	simultaneously render the first page and the portion of the audio stream by using the time information for the one or more static graphic images or for the first page, wherein the portion of the audio stream is rendered in dependence upon the first time offset.

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. 10,805,111** contains 13,961 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 21, 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. 10,805,111** and **ACCOMPANYING EXHIBITS** are being served on March 21, 2025 via Federal Express overnight mail on counsel of record for U.S. Patent No. 10,805,111 at the Correspondence Address of record below:

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

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

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

Dated: March 21, 2025

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