UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

CELLULAR SOUTH, INC.,
Patent Owner.

Case IPR2025-00875
Patent 9,940,972 B2
Issue Date: April 10, 2018

Title: VIDEO TO DATA

DECLARATION OF DR. HENRY HOUH

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

Page

TABLE OF CONTENTS
(continued)

Page

TABLE OF CONTENTS
(continued)

Page

TABLE OF CONTENTS
(continued)

Page

TABLE OF CONTENTS
(continued)

Page

I, Henry Houh, Ph.D., declare as follows:

I. INTRODUCTION AND QUALIFICATIONS

A. Qualifications and Experience

1. My complete qualifications and professional experience are described in my *Curriculum Vitae*, a copy of which can be found attached to this declaration as **Appendix A**. The following is a brief summary of my relevant qualifications and professional experience.

2. I received a Ph.D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology ("MIT") in 1998. Beforehand, I received a Master of Science degree in Electrical Engineering and Computer Science in 1991, a Bachelor of Science degree in Electrical Engineering and Computer Science in 1989, and a Bachelor of Science degree in Physics in 1990, all from MIT.

3. I am currently self-employed as an independent technical consultant. From its founding in 2012 until October 2022, I was also president of a company that provides supplemental science, technology, engineering, and mathematics ("STEM") education to children of all ages.

4. I first entered telecommunications in 1987 when I worked as a summer intern at AT&T Bell Laboratories as part of a five-year dual degree program at MIT. I continued to work at AT&T Bell Laboratories as part of this MIT program. While

at MIT, I was a teaching assistant ("TA") in the Electrical Engineering and Computer Science Department's core Computer Architectures course. I first was a TA as a senior for a role typically reserved for graduate students. I later became head TA. The course covered various topics in computer architectures. As a TA, I helped write homework assignments, lab assignments, and exams. I also taught in some of the recitation sections.

5. Later, as part of my doctoral research at MIT from 1991-1998, I was a research assistant in the Telemedia Network Systems ("TNS") group at the Laboratory for Computer Science. The TNS group built a high-speed gigabit network and created applications that ran over the network. Example applications included ones for remote video capture, processing, and display of video on computer terminals. In addition to working on the design of core network components, designing and building the high-speed links, and designing and writing the device drivers for the interface, I also set up the group's web server. Also, I helped to maintain, install, and upgrade the networking devices used within the group, along with other graduate students.

6. I also helped to build the web pages that initiated the above-mentioned video sessions via a web interface. Vice President Al Gore visited our group in 1996 and received a demonstration of—and remotely drove—a radio-controlled toy car

with a wireless video camera mounted on it that was built by our group. This toy car device received commands transmitted over a network from a remote computer, and video data from the toy car was transmitted wirelessly then over a computer network back to the user controller. On occasion, we allowed users visiting our web site to drive the toy car from their remote computer while they watched the video on their computer. The video stream was encoded by TNS-designed hardware, streamed over the TNS-designed network, and displayed using TNS-designed software.

7. Our video environment included an extensive amount of video processing. The closed captions were extracted from the analog video broadcasts, and our group built an extensive toolkit to perform video processing functions on the live video as it was being captured and optionally recorded. Our system included features such as motion detection, template matching, model matching, live green screening, face detection, histogram generation and matching, feature matching, edge detection, scene change detection, image transformation, filtering, image segmentation, object motion tracking, and other features.

8. I defended and submitted my Ph.D. thesis, titled "Designing Networks for Tomorrow's Traffic," in January 1998. As part of my thesis research, I analyzed local area and wide area flows to show a more efficient method for routing packets in a network, based on traffic patterns at the time. The traffic flow data included

Ethernet, IP, TCP or UDP, and RTP header information, which I analyzed to come to the conclusions in my thesis.

9. While I was in graduate school, I started a company that provided web site design services and targeted advertising opportunities. We also sold targeted advertising placements. As the web became popular first at colleges and universities, we focused initially on building college recruiting web sites and sold advertisements to companies wishing to target students to recruit at specific colleges. We developed technology that utilized the web user's college affiliation to customize advertisements targeted to students at that college.

10. From 1997 to 1999, I was a Senior Scientist and Engineer at NBX Corporation, a start-up that made business telephone systems for streaming packetized audio over data networks instead of using traditional telephone lines. NBX was later acquired by 3Com Corporation and the phone system is still used today by numerous businesses.

11. As part of my work at NBX, I designed the core audio reconstruction algorithms for the telephones, as well as the packet transmission algorithms. I also designed and validated the core packet transport protocol used by the phone system. The protocol was used for all signaling in the phone system, including for the setup of conference calls. The NBX system also featured a computer interface for initiating

phone calls, which could also initiate conference calls. The NBX system also supported the Telephony Application Programming Interface ("TAPI") that allowed other computer programs to integrate with our system telephony features. We obtained U.S. Patent No. 6,967,963, entitled "Telecommunication method for ensuring on-time delivery of packets containing time-sensitive data," as part of this work. The first release of the phones had integrated Ethernet repeater hubs embedded in them, and I worked with the documentation and field support people to develop sets of guidelines so that customers would not build their networks to violate the Ethernet specifications, typically by chaining too many hub repeaters and NBX phones together. The first release of the phones also worked using Ethernet-layer packets (without the use of IP). I also programmed the first prototype of our phone which communicated using IP, and I demonstrated our IP phones working over the Internet when we attended a trade show in California. The phone connected over the Internet to our headquarters in Andover, MA.

12. From 1999-2004, I was employed by Empirix or its predecessor company, Teradyne. Teradyne had traditionally made products to test chips and hardware, but the founder and board chairman, Alex d'Arbeloff wanted to expand the company's products to include software systems testing. Teradyne made a number of acquisitions to achieve this goal, and I was hired to work for Mr.

d'Arbeloff to work with all the acquired software test divisions. These divisions were later spun-out of Teradyne into a separate company, Empirix. Empirix was a leader in test tools for software systems including telecommunications protocols and systems, providing functional testing tools as well as load testing tools. From 2000-2001, I conceived and built a test platform for testing Voice-over-IP (VoIP). The first application on this new test platform was a cloud emulator for simulating the effects of transmitting VoIP over a busy network. The test platform was based on a network processor chip, which could be programmed to cut-through packets while processing packet data such as various protocol later headers including addresses included therein and even packet data contents. I also designed a protocol analyzer built on the same platform. The application captured and performed protocol decoding at various layers in the protocol stacks of captured packets, including detailed Ethernet header decoding, IP header decoding, TCP header decoding, UDP header decoding, RTP header decoding, and many other specified protocols. The application was also designed to reconstruct entire conversations that spanned multiple packets.

13. In 2006, as part of my role at BBN Technologies, I helped found PodZinger Inc., now known as RAMP Inc. PodZinger utilized BBN's speech recognition algorithms to search through the spoken words in audio and video

segments. Our speech recognition algorithms were also capable of extracting the emotional prosody of the speech input. While I was Vice President of Operations and Technology, PodZinger followed its initial prototype with a full streaming audio and video search solution. The metadata generated from speech recognition process was used in the search result user interface so that the user could see highlights of their search term in the audio transcript, the context around their search term, and the user could select a word in the transcript to jump the video to play from exactly the point where the selected word was uttered.

14. I also created a social networking web site, which BBN sold to a venture-funded startup company. In the process of creating the web site, I designed and specified the authentication and authorization protocols.

15. I applied for and received a number of patents in the above areas. In particular, U.S. Patent No. 9,697,230 describes "an automated method and apparatus for generating metadata enhanced for audio/video search-driven applications. The apparatus includes a media indexer that obtains an media file/stream (e.g., audio/video podcasts), applies one or more automated media processing techniques to the media file/stream, combines the results of the media processing into metadata enhanced for audio/video search, and stores the enhanced metadata in a searchable index or other data repository." (EX1006, 4:31-39.) A media indexer "applies a

predefined set of audio or video processing routines to derive a portion of the enhanced metadata from the media content." (*Id.*, 5:19-21.) The indexer utilizes speech recognition, natural language processing, and video frame analyzers, that can recognize scenes, watermarks, objects, faces, and overlay text. (*Id.*, Fig. 1B, 5:22-34, 6:43-57.) Related advertising may then be displayed utilizing the results of such an analysis. (*Id.*, 15:52-16:25.) A partial list of my patents and patent applications include:

- U.S. Patent No. 9,697,231, H. Houh, J. N. Stern, R. Spina, M. Meteer [sic], "Methods and apparatus for providing virtual media channels based on media search," July 4, 2017.
- U.S. Patent No. 9,697,230, H. Houh, J. N. Stern, "Methods and apparatus for dynamic presentation of advertising, factual, and informational content using enhanced metadata in search-driven media applications," July 4, 2017. *See also* WO2007056485.
- U.S. Patent No. 7,801,910, H. Houh and J. N. Stern, "Method and apparatus for timed tagging of media content," September 21, 2010. *See also* US20070112837, US20090222442, WO2007056535.
- U.S. Patent Application Publication No. 20070106660, "Method and

apparatus for using confidence scores of enhanced metadata in search-driven media applications"

- U.S. Patent Application Publication No. 20070118873, "Methods and apparatus for merging media content"
- U.S. Patent Application Publication No. 20090222442, "User-directed navigation of multimedia search results"
- U.S. Patent Application 11/395,732, "User-Directed Navigation of Multimedia Search Results"

16. I have been retained by counsel for Petitioner to provide my expert opinion in connection with the above-captioned proceeding as set forth herein.

B. Materials Considered

17. The analysis that I provide in this Declaration is based on my education, research, and experience, as well as the documents I have considered. In forming my opinions, I read and considered U.S. Patent No. 9,940,972 B2 ("'972" or "'972 patent") (**EX1001**) and its prosecution history. I cite to the following documents in my analysis below:

Exhibit No.	Description of Document
1001	U.S. Patent No. 9,940,972 B2 to Naeem Lakhani et al. (filed February 7, 2014, issued April 10, 2018) ("'972" or "'972 patent")

Exhibit No.	Description of Document
1003	U.S. Patent Application Publication No. 2012/0078712 A1 to James A. Fontana et al. (filed September 27, 2010, issued March 29, 2012) (“**Fontana**”)
1004	U.S. Patent Application Publication No. 2007/0112630 A1 to Wai Kit Lau et al. (filed November 7, 2006, issued May 17, 2007) (“**Lau**”)
1005	U.S. Patent Application Publication No. 2012/0239401 A1 to Takayuki Arakawa (filed November 26, 2010, issued September 20, 2012) (“**Arakawa**”)
1006	U.S. Patent No. 9,697,230 B2 to Henry Houh et al. (filed March 31, 2006, issued July 4, 2017)
1007	’972 Patent File History
1008	Adobe Flash Video File Format Specification, Version 10.1, https://rtmp.veriskope.com/pdf/video_file_format_spec_v10_1.pdf
1009	Excerpts from Scot Hacker, MP3 The Definitive Guide (1st ed. 2000)
1010	U.S. Patent Application Publication No. 2011/0305394 A1 to David W. Singer et al. (filed June 15, 2010, issued December 15, 2011) (“**Singer**”)
1011	U.S. Patent No. 9,407,942 B2 to Andrew Brenneman (filed November 7, 2012, issued August 2, 2016) (“**Brenneman**”)
1012	U.S. Patent Application Publication No. 2004/0125877 A1 to Shin-Fu Chang et al. (filed April 9, 2001, published July 1, 2004) (“**Chang**”)

II. PERSON OF ORDINARY SKILL IN THE ART

18. The face page of the ’972 patent shows that the earliest patent application to which the ’972 patent claims priority was filed August 15, 2013

(61/866,175).[1] Therefore, I understand that the claims of the '972 patent are subject to the patent laws in effect under the America Invents Act ("AIA"), including sections 102 and 103 regarding invalidity based on prior art, which apply to patent applications having an effective filing date on or after March 16, 2013. I understand that under the AIA, an assessment of claims of a patent should be undertaken from the perspective of a person of ordinary skill in the art is to be made as of the effective filing date of the patent (i.e., the earliest of the actual filing date of the patent or application, or the filing date of the earliest application with priority). I note that, for patents subject to pre-AIA patent laws, such an assessment should be made as of the earliest claimed priority date (i.e., the "time the invention was made"). For purposes of my analysis, whether the "effective filing date" or "earliest claimed priority date" applies does not affect my analysis. This is because I understand that the alleged "effective filing date" and "earliest claimed priority date" are the same in this case—August 15, 2013.

---

[1] I express no opinion on whether the '972 patent is actually entitled to the benefit of the provisional application to which it claims priority (61/866,175). This issue is not relevant here because all of the prior art references relied upon predate that date. In the event entitlement to the provisional filing date may later become relevant, for example based on arguments later presented by Patent Owner, I reserve my right to provide opinions on that issue at that time.

19. I have also been advised that to determine the appropriate level of a person having ordinary skill in the art, the following factors may be considered: (1) the types of problems encountered by those working in the field and prior art solutions thereto; (2) the sophistication of the technology in question, and the rapidity with which innovations occur in the field; (3) the educational level of active workers in the field; and (4) the educational level of the inventor.

20. The '972 patent states that it "relates to a method and a system for generating various and useful data from videos." ('972, 1:10-11; *see also id.*, 1:27-29 ("The present invention is generally directed to a method to generate data from video content, such as text and/or image-related information.").)

21. In my opinion, a person of ordinary skill in the art as of August 2013 would have possessed a bachelor's degree in electrical engineering, computer science, or similar field, with at least two years of experience in developing and implementing computer software for processing and/or analyzing multimedia content, such as audio, video, or image data. A person could also have qualified as a person of ordinary skill in the art with some combination of (1) more formal education (such as a master's of science degree) and less technical experience, or (2) less formal education and more technical or professional experience.

22. My opinions regarding the level of ordinary skill in the art are based

on, among other things, my experience in the field of computer science, my understanding of the basic qualifications that would be relevant to an engineer or scientist tasked with investigating methods and systems in the relevant area, and my familiarity with the backgrounds of colleagues, co-workers, and employees, both past and present.

23. Although my qualifications and experience exceed those of the hypothetical person having ordinary skill in the art defined above, my analysis and opinions regarding the ’972 patent have been based on the perspective of a person of ordinary skill in the art as of August 2013.

III. STATEMENT OF LEGAL PRINCIPLES

A. Claim Construction

24. I understand that a purpose of claim construction is to determine what a person of ordinary skill in the art would have understood the claim terms to mean. Claim terms are generally given their ordinary and customary meaning, which is the meaning that the term would have to a person of ordinary skill in the art in question as of the effective filing date.

25. I understand that the person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which the disputed term appears, but in the context of the entire patent, including the specification. I

understand that the patent specification, under the legal principles, has been described as the single best guide to the meaning of a claim term, and is thus highly relevant to the interpretation of claim terms. And I understand for terms that do not have a customary meaning within the art, the specification usually supplies the best context of understanding the meaning of those terms.

26. I further understand that other claims of the patent in question, both asserted and unasserted, can be valuable sources of information as to the meaning of a claim term. Because the claim terms are normally used consistently throughout the patent, the usage of a term in one claim can often illuminate the meaning of the same term in other claims. Differences among claims can also be a useful guide in understanding the meaning of particular claim terms.

27. I understand that the prosecution history can further inform the meaning of the claim language by demonstrating how the inventors understood the invention and whether the inventors limited the invention in the course of prosecution, making the claim scope narrower than it otherwise would be. Extrinsic evidence, such as dictionaries, may also be consulted in construing the claim terms.

28. I understand that, in *Inter Partes* Review (IPR) proceedings, a claim of a patent shall be construed using the same claim construction standard that would be used to construe the claim in a civil action filed in a U.S. district court (which I

understand is called the "*Phillips*" claim construction standard), including construing the claim in accordance with the ordinary and customary meaning of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

29. I have been instructed by counsel to apply the "*Phillips*" claim construction standard for purposes of interpreting the claims in this proceeding, to the extent they require an explicit construction. The description of the legal principles set forth above thus provides my understanding of the "*Phillips*" standard as provided to me by counsel.

30. For purposes of my analysis here, I do not believe express claim constructions are necessary because the prior art renders the claims obvious under any reasonable construction.

B. Obviousness (§ 103)

31. I understand that a patent claim is obvious if, as of the effective filing date, it would have been obvious to a person having ordinary skill in the field of the technology (the "art") to which the claimed subject matter belongs.

32. I understand that the following factors should be considered in analyzing obviousness: (1) the scope and content of the prior art; (2) the differences between the prior art and the claims; and (3) the level of ordinary skill in the pertinent

art. I also understand that certain other facts known as "secondary considerations" such as commercial success, unexplained results, long felt but unsolved need, industry acclaim, simultaneous invention, copying by others, skepticism by experts in the field, and failure of others may be utilized as indicia of nonobviousness. I understand, however, that secondary considerations should be connected, or have a "nexus," with the invention claimed in the patent at issue.

33. I understand that a reference qualifies as prior art for obviousness purposes when it is analogous to the claimed invention. The test for determining what art is analogous is: (1) whether the art is from the same field of endeavor, regardless of the problem addressed, and (2) if the reference is not within the field of the inventor's endeavor, whether the reference still is reasonably pertinent to the particular problem with which the inventor is involved.

34. I understand that a person of ordinary skill in the art is assumed to have knowledge of all prior art. I understand that one skilled in the art can combine various prior art references based on the teachings of those prior art references, the general knowledge present in the art, or common sense. I understand that a motivation to combine references may be implicit in the prior art, and there is no requirement that there be an actual or explicit teaching to combine two references. Thus, one may take into account the inferences and creative steps that a person of

ordinary skill in the art would employ to combine the known elements in the prior art in the manner claimed by the patent at issue. I understand that one should avoid "hindsight bias" and *ex post* reasoning in performing an obviousness analysis. But this does not mean that a person of ordinary skill in the art for purposes of the obviousness inquiry does not have recourse to common sense.

35. I understand that when determining whether a patent claim is obvious in light of the prior art, neither the particular motivation for the patent nor the stated purpose of the patentee is controlling. The primary inquiry has to do with the objective reach of the claims, and that if those claims extend to something that is obvious, then the entire patent claim is invalid.

36. I understand one way that a patent can be found obvious is if there existed at the time of the invention a known problem for which there was an obvious solution encompassed by the patent's claims. I understand that a motivation to combine various prior art references to solve a particular problem may come from a variety of sources, including market demand or scientific literature. I understand that a need or problem known in the field at the time of the invention can also provide a reason to combine prior art references and render a patent claim invalid for obviousness. I understand that familiar items may have obvious uses beyond their primary purpose, and that a person of ordinary skill in the art will be able to fit the

teachings of multiple prior art references together like the pieces of a puzzle. I understand that a person of ordinary skill is also a person of at least ordinary creativity. I understand when there is a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her technical grasp. If this finite number of predictable solutions leads to the anticipated success, I understand that the invention is likely the product of ordinary skill and common sense, and not of any sort of innovation. I understand that the fact that a combination was obvious to try might also show that it was obvious, and hence invalid, under the patent laws. I understand that if a patent claims a combination of familiar elements according to known methods, the combination is likely to be obvious when it does no more than yield predictable results. Thus, if a person of ordinary skill in the art can implement a predictable variation, an invention is likely obvious. I understand that combining embodiments disclosed near each other in a prior art reference would not ordinarily require a leap of inventiveness.

37. I understand that obviousness may be shown by demonstrating that it would have been obvious to modify what is taught in a single piece of prior art to create the patented invention. Obviousness may also be shown by demonstrating that it would have been obvious to combine the teachings of more than one item of

prior art. I understand that a claimed invention may be obvious if some teaching, suggestion, or motivation exists that would have led a person of ordinary skill in the art to combine the invalidating references. I also understand that this suggestion or motivation may come from the knowledge of a person having ordinary skill in the art, or from sources such as explicit statements in the prior art. I understand that when there is a design need or market pressure, and there are a finite number of predictable solutions, a person of ordinary skill may be motivated to apply common sense and his skill to combine the known options in order to solve the problem.

38. I understand the following are examples of approaches and rationales that may be considered in determining whether a piece of prior art could have been combined with other prior art or with other information within the knowledge of a person having ordinary skill in the art:

(1) Some teaching, motivation, or suggestion in the prior art that would have led a person of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention;

(2) Known work in one field of endeavor may prompt variations of it for use in the same field or a different field based on design incentives or other market forces if the variations would have been predictable to a person of ordinary skill in the art;

(3) Combining prior art elements according to known methods to yield predictable results;

(4) Applying a known technique to a known device, method, or product ready for improvement to yield predictable results;

(5) Applying a technique or approach that would have been "obvious to try" (choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success);

(6) Simple substitution of one known element for another to obtain predictable results; or

(7) Use of a known technique to improve similar products, devices, or methods in the same way.

39. I understand that, when determining whether a claimed combination is obvious, the correct analysis is not whether one of ordinary skill in the art, writing on a blank slate, would have chosen the particular combination of elements described in the claim. Instead, I understand the correct analysis considers whether one of ordinary skill, facing the wide range of needs created by developments in the field of endeavor, would have seen a benefit to selecting the combination claimed.

40. I understand that the test for obviousness is not whether the features of a secondary reference may be bodily incorporated into the structure of the primary

reference. The test for obviousness, in other words, is not whether the references could be physically combined but whether the claimed inventions are rendered obvious by the teachings of the prior art as a whole.

IV. THE ’972 PATENT

A. Overview of the Specification

41. The ’972 patent purports to offer one solution to the common problem of how to reduce audio and visual information to written form, in order to provide additional content-matched audio-visual information.

42. The ’972 patent describes its purpose at a very high level. It states as its “Technical Field” that it “relates to a method and a system for generating various and useful data from videos.” (’972, 1:10-11.) And the summary of the invention confirms the broad brush the patent purports to take: “The present invention is generally directed to a method to generate data from video content, such as text and/or image-related information.” (’972, 1:27-29.)

43. Audio-visual data such as videos long predates the ’972 patent, and prior artists similarly had long been working on various ways to analyze, search, and generate video content. By the time of the alleged priority date (August 15, 2013), the fields of audio and image analysis for video content were well-developed. Indeed, the Background section of the ’972 patent acknowledges known image

searching techniques to identify matching and similar images, as well as audio-to-text algorithms for transcribing text from audio. ('972, 1:15-23.)

44. The specification of the '972 patent provides primarily a series of high-level, functional explanations for how to implement the alleged invention, with scant information on how to carry out any of its steps. For example, Figure 1 of the '972 patent, shown below, "illustrates an embodiment of present invention," and specifically "[a]n embodiment of video-to-text engine operation":

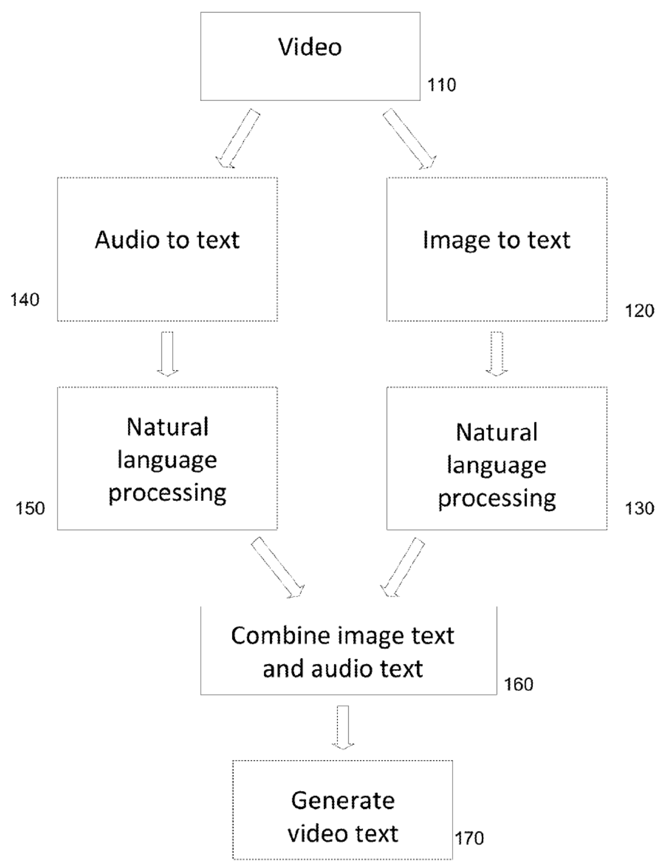


Figure 1

('972, Fig. 1, 2:5, 3:51-52.) At step **110**, "a video stream is presented." ('972, 3:52.) "The engine can extract audio data and image data (e.g. images or frames forming the video) from the video stream." ('972, 3:58-59.) At step **120**, "the video-to-text engine performs an image data processing on the video stream[,]" such as symbol (or object) detection and identification, such as identifying a dog in the video image data. ('972, 3:64-65, 4:3-12.) Notably, the '972 patent explains that this may be performed by an "image processing technique such as color edge detection," but does not explain how to carry out color edge detection. ('972, 4:4-6.) A POSA, however, would have recognized color edge detection as one of many well-known methods for image processing available in 2013. ('972, 1:15-23 (acknowledging image identification was known).) Then, "[b]ased on the identified symbol, a plurality of instances of the symbol can be compared to a topic database to identify a topic (such as an event)[,]" such as identifying the dog as running or jumping. ('972, 4:20-23.) Thus, "text describing a symbol of the video and topic relating to the symbol may be generated." ('972, 4:26-28.) At step **130**, "natural language processing can be applied to the image text." ('972, 4:61-62.) "For example, based on dictionary, grammar, and a knowledge database, the text extracted from video images can be modified as the video-to-text engine selects primary semantics from a plurality of possible semantics." ('972, 4:62-66.) Applying dictionary definitions and grammatical rules to text was nothing new in 2013.

45. "In parallel, at **140**, the video-to-text engine can perform audio-to-text processing on audio data associated with the video." ('972, 5:4-6.) At step **150**, "natural language processing can be applied to the text" and the text "can be given context, an applied sentiment, and topical weightings." ('972, 5:41-45.) Again, the specification is largely silent as to how to implement these natural language processing tools, which would have been well-known to a POSA. ('972, 1:15-23 (acknowledging audio-to-text was known).)

46. At step **160**, "the topics generated from an image or a frame and the topics extracted from audio can be combined. The text can be cross-referenced, and topics common to both texts would be given additional weights." ('972, 5:46-49.) In other words, this is the common step of finding a way to turn an apples-to-oranges comparison (audio and video) into apples-to-apples (give everything a text descriptor or "topic" that can be combined and compared).

47. At step **170**, "the video-to-text engine generates video text, such as text describing the content of the video, using the result of the combined text and cross reference." ('972, 5:49-52.) This is just summarizing the video, such as what a theater critic or museum curator would do for a play or a work of multimedia art. "For example, key words indicating topic and semantic that appear in both texts can

be selected or emphasized." ('972, 5:52-54.) An "advertisement (text, images, or animation)" can then be generated based on the video text and added to the video stream. ('972, 6:24-29.)

48. I discuss additional aspects of the '972 patent in my analysis of the claims below.

B. The Challenged Claims

49. This Declaration addresses claims 1-20 of the '972 patent. Independent claim 1 is representative and recites:

1. A method to generate video data from a video comprising:

[a] generating audio files and image files from the video;

[b] distributing the image files across a plurality of processors and processing the image files in parallel, wherein processing the image files comprises extracting one or more objects and identifying the one or more objects;

[c] processing the audio files;

[d] converting audio files associated with the video to text;

[e] converting the image files associated with the video to video data;

[f] generating a topical meta-data that describes content of the video by deriving semantic information from the identification of the one or more objects and semantic information from the audio files;

[g] adding the topical meta-data to the video; and

[h] cross-referencing the text and the video data based on the generated topical meta-data to determine topics;

[i] generating video text based on the cross-referencing, wherein the video text describes content of the video;

[j] generating a text, image, or animation based on the video text; and

[k] placing the text, image, or animation in the video.

('972, 8:2-25 (Claim 1 (bracketed notation (e.g., "[a]") added).)

50. I address the claims further in my detailed analysis in **Part V** below.

V. APPLICATION OF THE PRIOR ART TO CHALLENGED CLAIMS

51. I have reviewed and analyzed the prior art references and materials listed in **Part I.B** above. In my opinion the claims of the '972 patent would have been obvious to a person of ordinary skill in the art ("POSA") based on the following combinations of the prior art.

Ground	Claims	Basis for Challenge Under § 103
1	1-20	Fontana in view of Lau
2	8-9, 20	Fontana in view of Lau and Arakawa

52. I am informed by counsel that each of the references cited in the

grounds above qualifies as prior art to the challenged claims because each reference was filed and/or published before the earliest filing date to which the ’972 patent can claim priority (August 15, 2013).

A. Brief Summary and Overview of Prior Art

1. References Common to All Grounds

(a) Fontana (EX1003)

53. **Fontana**, U.S. Patent Application Publication No. 2012/0078712 A1 to James A. Fontana et al., is entitled “Systems and Methods for Processing and Delivery of Multimedia Content.” I am informed that Fontana qualifies as prior art because it was filed on September 27, 2010, and was published on March 29, 2012, both before the earliest priority date of the challenged claims. Fontana was not cited during prosecution of the ’972 patent.

54. Fontana is directed to “[m]ethods and systems for providing multimedia content.” (Fontana Abstract.) Like the ’972 patent, Fontana discloses a method that generates images and audio files from a video, performs speech-to-text recognition of the audio, detects objects within the images, and generates and searches within meta-data about the video. Fontana teaches:

> [A] method for providing multimedia content includes receiving multimedia content from a source, and applying a container to the

> multimedia content using one or more computing systems, the container defining an interface through which metadata external to the multimedia content is linked with one or more portions of the multimedia content. The method includes constructing a set of text metadata describing an audio portion of the multimedia content, generating a set of object metadata describing at least a portion of one or more objects appearing in the multimedia content, and returning the container and multimedia content to the source. The method further includes storing the text metadata and object metadata at the one or more computing systems, and upon notification of a request for playback of the multimedia content at the one or more computing systems, providing the text metadata and the object metadata associated with the container for synchronized use during playback of the multimedia content via the container.

(Fontana ¶0010.)

55. The metadata generated by Fontana can then be used to search the multimedia content and synchronously display enhancements to the multimedia data: "The methods and systems described herein provide search and playback enhancements to multimedia content, in part, by processing the content in a back-end server environment to generate metadata describing the content. By integrating enhancements using metadata linked to the content, the enhancements can be synchronously displayed or controlled during playback of the multimedia content,

and can be customized to the particular items within the content as well as the individuals viewing the content." (Fontana ¶0041.)

56. An embodiment of this method is depicted, for example, in Figure 6, which shows "Audio Processing" at 606, "Video Processing" at 608, "Full-Text Search" at 630, and "Video Metadata" at 634.

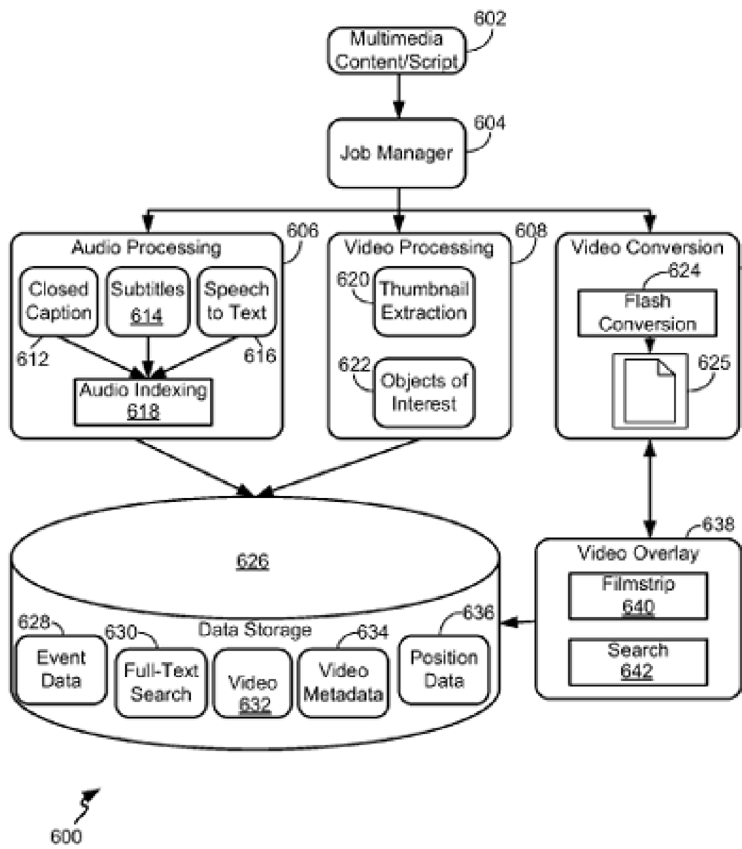


FIG. 6

57. Fontana additionally teaches implementing its method in "a plurality of

distributed computing systems[.]" (Fontana ¶0045.) In one embodiment shown in Figure 3, "the distributed computing network 300 includes a plurality of computing systems, illustrated as servers 302a-c. The servers 302a-c are communicatively interconnected, and each includes a corresponding data storage system 304a-c. The servers 302a-c share a distributed memory cache 306, and are each capable of accessing a shared cache of memory that is not residing in any of data storage systems 304a-c. The servers 302a-c are interfaced to inbound work, such as from a scheduler system (as described in further detail in my analysis below) for coordination and communication of data for processing." (Fontana ¶0052.) "[O]ne or more of the servers 302a-c can include specific graphical processing units for processing lower level video, image or audio algorithms. Other specific capabilities can be included into the servers 302a-c as well. The servers 302a-c are configured to share processing jobs, such that tasks can be performed by one or more of the computing systems, or separated and performed across multiple computing systems in parallel." (Fontana ¶0053.)

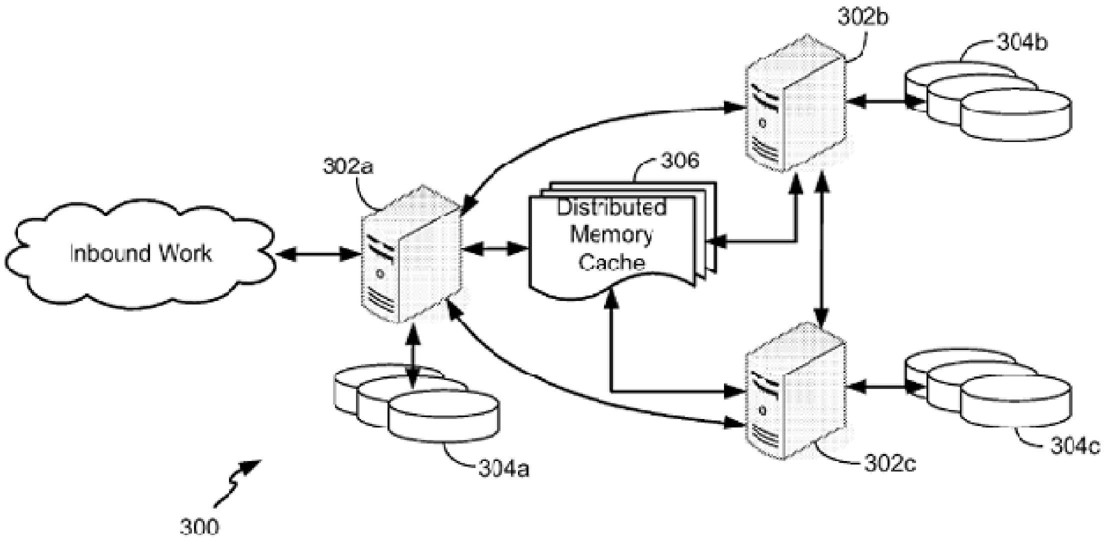


FIG. 3

(Fontana Fig. 3.)

58. I will provide more information about Fontana in my discussion of the claim limitations below.

(b) Lau (EX1004)

59. **Lau**, U.S. Patent Application Publication No. 2007/0112630 A1 to Wai Kit Lau et al., is entitled "Techniques for Rendering Advertisements with Rich Media." I am informed that Lau qualifies as prior art because it was filed on November 7, 2006 and published on May 17, 2007, which are both before the earliest priority date of the challenged claims. Lau was cited during prosecution and mentioned by the examiner as "cited to disclose rendering advertisements with rich

media" but was not relied upon in any ground of rejection. (EX1007, p.0069.)

60. Like Fontana, Lau relates to audio and image processing of multimedia content, such as videos. Lau is directed to a method by which "an advertisement is matched to subject matter in a portion of rich media content, such as a digital video[.]" (Lau Abstract.) "In one embodiment, engine 102 correlates advertisements to subject matter associated with rich media content. Accordingly, an advertisement that correlates to the subject matter associated with the portion of rich media content may be served such that it can be rendered on user device 104 relative to the portion of rich media content. Different methods may be used to correlate or match advertisements to portions of the rich media content." (Lau ¶0020.)

61. One method of matching advertisements to media content disclosed by Lau is based on keyword searching: "For example, an advertisement may be associated with a keyword. When that keyword is used in the rich media, correlation engine 202 correlates the advertisement to a portion of rich media content in which the keyword is used." (Lau ¶0037.) These keywords may be derived from analysis of audio and visual elements of the media:

> Correlation recognition detection techniques may be used to determine that the advertisement is correlated to the portion of rich media content.

> For example, keywords may be detected in the rich media using audio recognition. Audio recognition may include speech recognition, music detection on music portions, sound effect detection on sound effects, etc. Other techniques for keyword detection can include using preset word tags or indicators in the rich media content. Image recognition can be used on visual portions of the rich media content. For example, optical character recognition (OCR), facial recognition, object matching, etc. Other recognition techniques can be employed. For example, any suitable way of determining the content of rich media can be used to correlate a portion of the rich media content to an advertisement.

(Lau ¶0040.)

62. Lau teaches that the advertisement may be generated by identifying "ad units" based on keywords, and then combining those ad units into an advertisement for display with the media content:

> Correlation engine 202, when determining the advertisement, may determine one or more ad units that correlate to the subject matter. For example, based on one or more keywords, ad units from the ad matrix are determined. The ad units are then combined into an advertisement that is correlated to the subject matter. One example of this is BMW may provide a general ad unit for their logo and have a different ad unit for different models, such as the 330 model, 530 model, etc. The logo unit and each of the model units can be combined at runtime based on

> the context of the content. If the content talks about the 330 model then the logo and the 330 ad units may be combined and presented to the user.

(Lau ¶0038.) The matched advertisement can then "be displayed in serial, parallel, or be injected into the rich media content." (Lau ¶0034.)

63. I will provide more information about Lau in my discussion of the claim limitations below.

2. Additional References

(a) Arakawa (EX1005)

64. **Arakawa**, U.S. Patent Application Publication No. 2012/0239401 A1 to Takayuki Arakawa ("Arakawa"), is entitled "Voice Recognition System and Voice Recognition Method." I am informed that Arakawa qualifies as prior art because it is the publication of a PCT filed on November 26, 2010, with a section 371 (c)(1), (2), (4) date of June 8, 2012 and a publication date of September 20, 2012, all of which are before the earliest priority date of the challenged claims. Arakawa subsequently issued as U.S. Patent No. 9,002,709 on April 7, 2015. Arakawa was not cited during prosecution of the '972 patent.

65. I point to Arakawa in Ground 2 for dependent claims 8, 9, and 20, for its recitation of "**segmenting the audio files**" at "**spectrum thresholds**."

66. Like Fontana, Arakawa relates to audio processing. Arakawa teaches "a voice recognition system capable of, while suppressing negative influences from sound not to be recognized, correctly estimating utterance sections that are to be recognized. A voice segmenting means calculates voice feature values, and segments voice sections or non-voice sections by comparing the voice feature values with a threshold value." (Arakawa Abstract.)

67. I will provide more information about Arakawa in my discussion of the claim limitations below.

B. Ground 1: Claims 1-20 Are Obvious Over Fontana in view of Lau

1. Independent Claim 1: "A method to generate video data from a video comprising:" (Claim 1[pre])

68. For reference, claim 1 recites:

1. A method to generate video data from a video comprising:

[a] generating audio files and image files from the video;

[b] distributing the image files across a plurality of processors and processing the image files in parallel, wherein processing the image files comprises extracting one or more objects and identifying the one or more objects;

[c] processing the audio files;

[d] converting audio files associated with the video to text;

[e] converting the image files associated with the video to video data;

[f] generating a topical meta-data that describes content of the video by deriving semantic information from the identification of the one or more objects and semantic information from the audio files;

[g] adding the topical meta-data to the video; and

[h] cross-referencing the text and the video data based on the generated topical meta-data to determine topics;

[i] generating video text based on the cross-referencing, wherein the video text describes content of the video;

[j] generating a text, image, or animation based on the video text; and

[k] placing the text, image, or animation in the video.

('972, 8:2-25 (Claim 1 (bracketed notation (e.g., "[a]") added).)

69. The preamble of claim 1 recites "[**a**] **method to generate video data from a video comprising**." Assuming the preamble provides a claim limitation, Fontana discloses it. Fontana discloses "**generat**[ing] **video data from a video**," such as meta-data describing the video generated from performing audio processing (for example speech-to-text processing), and/or object recognition of visual content of a video.

70. For example, Fontana teaches "<u>constructing a set of text metadata</u>

describing an audio portion of the multimedia content, and generating a set of object metadata describing at least a portion of one or more objects appearing in the multimedia content." (Fontana ¶0008.[2]) Fontana explains that the "multimedia content can include any type of content containing, for example, one or more of images, video, audio, or a combination thereof…. In the context of the present disclosure, a robust example of multimedia content is used in which video and audio information are included[.]" (Fontana ¶0042.) Thus, Fontana's "text metadata" and "object metadata" can be generated from a video, making them "video data from a video."

71. Therefore, Fontana discloses "[**a**] **method to generate video data from a video**." And as explained below, the method as claimed is obvious over Fontana and Lau.

(a) "generating audio files and image files from the video;" (Claim 1[a])

72. Fontana discloses "**generating audio files and image files from the video**" by separately processing the audio and video content of the multimedia content, and further extracting thumbnail images ("**image files**") from the video.

[2] Unless otherwise noted, all emphasis added.

73. For example, Fontana teaches "processing of multimedia content at an audio processing module 606, a video processing module 608, and a video conversion module 610" where "[e]ach of these modules can be executed concurrently (e.g., in parallel), with jobs associated with each module operating on one or more computing systems as defined by a scheduler (e.g., scheduler **406** of Fig. 4)." (Fontana ¶0090.) "The audio processing module 606 is configured to process audio content associated with the multimedia content" while the "video processing module 608 is configured to process the video portion(s) of multimedia content[.]" (Fontana ¶¶0091, 0094.) Fontana further discloses that these two processes are separate: "the various distributed computing systems described in FIGS. 1-5, above, allow for segmenting the processing into discrete portions (e.g., audio, video processing separately, etc.) and parallel pipelined processing of the data to ensure fast content processing[.]" (Fontana ¶0135.) Fontana thus teaches separately processing the audio and video from the multimedia content. As described below, Fontana teaches that these audio and video processing modules are run on audio and video content extracted, that is, "**generated**" from the multimedia content. A POSA would further appreciate that a video, separated from its audio, is a series of images, such that video files are a form of "**image files**."

74. Fontana further teaches that the "audio processing module" can be run

on "**audio files**" that are **generated** from the video. Specifically, Fontana teaches that "[i]f no closed captioning information is present, an audio separation operation 1128 <u>strips, or extracts, the audio from the multimedia content</u>." (Fontana ¶0164.) A POSA would have understood that stripping or extracting the audio from the multimedia content generates an audio file. Fontana additionally teaches that "it is recognized that a number of sources provide speech to text conversion program that approach the conversion differently." (Fontana ¶0169.) One such program disclosed in Fontana is "large vocabulary continuous speech recognition (LVCSR) engines" that "depend on a language model that includes a vocabulary/dictionary <u>for speech-to-text conversion of audio files.</u>" (Fontana ¶0169.) A POSA would thus have appreciated that Fontana's disclosures encompass "**generating audio files … from the video**."

75. Fontana additionally teaches that the "video processing module" **generates** an additional set of "**image files**" in the form of thumbnail images derived from the video. Fontana teaches: "In the embodiment shown the video processing module 608 includes a thumbnail extraction module[,]" that "is arranged to generate thumbnails at possible locations the content provider would like to create an object of interest …. In some embodiments, the thumbnail extraction module 618 generates a series of thumbnails representing scenes throughout the multimedia content."

(Fontana ¶0095.) A POSA would appreciate that a thumbnail is an image. Such images were commonly stored as image files in 2013, such as in gif, tiff, jpg, or tiff formats, for example, which are common formats for graphical image files. A POSA would thus have appreciated that by extracting thumbnail images from the video, Fontana's disclosures satisfy "**generating … image files from the video.**"

(b) "distributing the image files across a plurality of processors and processing the image files in parallel, wherein processing the image files comprises extracting one or more objects and identifying the one or more objects;" (Claim 1[b])

76. Fontana discloses or renders obvious "**distributing the image files across a plurality of processors and processing the image files in parallel, wherein processing the image files comprises extracting one or more objects and identifying the one or more objects**."

77. Fontana discloses implementing its method in a distributed computing environment: "The multimedia processing system 104, although represented by a single computing system, is in preferred embodiments <u>a plurality of distributed computing systems</u>[.]" (Fontana ¶0045.) A POSA would have understood that a distributed computer system is a form of "parallel" processing as required by the '972 patent—by distributing tasks across a plurality of processors, each processor is able to handle a portion of the tasks, resulting in parallel processing of the set of

requests. As Fontana explains, its distributed computing system allows tasks, including "video, image or audio algorithms[,]" to be "separated and performed across multiple computing systems in parallel." (Fontana ¶0053; *see also id.* ¶0090.) As further detailed below, a POSA would have understood that Fontana's distributed computing system includes both parallel processing of image files separate from audio files, and parallel processing for image-processing tasks, resulting in processing multiple image files in parallel. Therefore, Fontana discloses that its distributed computing system distributes image processing requests, such that it "**distribut[es] the image files across a plurality of processors and process[es] the image files in parallel**."

78. For example, Fontana teaches:

> The network 200 can, in certain embodiments, correspond to an architecture underlying the multimedia processing system 104 of FIG. 1, for example in a cloud-based or other distributed computing environment. The network 200 includes, in the embodiment shown, a workflow server 202 interconnected to an integration framework 204 and a storage network 206. The integration framework 204 provides interconnectivity and data sharing among a plurality of computing systems, such that the computing systems can share workloads, messages, and other tasks. The integration framework 204 can be connected to any of a plurality of differing types of computing systems

> 208 capable of sharing workloads; in the embodiment shown, various shared computing systems are illustrated including workstations 208a, grid computing systems 208b, compute clusters 208c, data resources 208d, and one or more high performance computing systems 208e.

(Fontana ¶0048.)

79. Fontana explains that this workflow server "receives inbound data processing requests, for example from a content provider (as further discussed below) and <u>distributes one or more portions of jobs associated with each data processing request to the integration framework 204 and the storage network 206</u>." (Fontana ¶0050.) The resulting system "allows creation of pipelined data processing systems within a distributed computing environment, <u>allowing computationally intensive jobs (e.g., video and audio content processing) to be distributed across a number of computing systems.</u>" (*Id.*) In other words, the workflow server divides computationally intensive jobs, such as video processing (including image processing, as previously discussed), into portions, and distributes those portions across multiple computing systems. A POSA would have appreciated that the result of this system is that these processing tasks, such as image processing, would be divided up and performed in parallel.

80. Fontana further teaches that this distributed and parallel processing of

image files can be carried out by a plurality of servers. Fontana teaches that its method is implemented using "a plurality of computing systems, illustrated as servers 302a-c. The servers 302a-c are communicatively interconnected, … share a distributed memory cache 306, and are each capable of accessing a shared cache of memory[.]" (Fontana ¶0052.) These servers "are interfaced to inbound work … for coordination and communication of data for processing." (*Id.*) Fontana further teaches that "one or more of the servers 302a-c can include specific graphical processing units for processing lower level video, image or audio algorithms…. The servers 302a-c are configured to share processing jobs, such that tasks can be performed by one or more of the computing systems, or separated and performed across multiple computing systems in parallel." (Fontana ¶0053.) As a result, "any of those computing systems can perform all **or a portion** of a processing job as defined by a scheduling algorithm, allowing multimedia content to be processed efficiently when necessary." (Fontana ¶0055.) A POSA would have appreciated that this system of servers, which share processing jobs and workloads by, for example, separating and performing the jobs in parallel, carries out parallel processing, including for processing image files. In the context of image processing, it would have been obvious to a POSA that image processing within Fontana's distributed computing system could **process the image files in parallel** by, for example, distributing the image files for processing across multiple servers.

81. Indeed, Fontana further discloses processing images files to "**extract[] one or more objects and identify[] the one or more objects**" as part of "**processing the image files in parallel**." Fontana explains that its system "includes a scheduler 406" that "in general receives tasks from the frontend 402 as defined by content providers, for example indicating that <u>multimedia content should be processed to generate one or more objects of interest</u>, to create a transcript of the multimedia content, or other typically computationally-intensive functions." (Fontana ¶0061.) "The scheduler 406 <u>receives and routes the content and processing requests</u> to the desired computing systems within the grid 408; the scheduler <u>generally provides the ability to equally distribute resources to all jobs that are running at once</u>[.]" (*Id.*) A POSA would have appreciated these disclosures as teaching that the scheduler is what distributes tasks among servers for parallel processing, and that Fontana's scheduler distributes tasks to process multimedia content to generate one or more objects of interest.

82. As discussed above for claim 1[a], a POSA would have understood that Fontana's image processing algorithm generates image files at least in the form of thumbnail images. Fontana further discloses "**extracting one or more objects and identifying the one or more objects**" as part of processing the image files. The

objects to be identified by Fontana can be either provided to the system or automatically identified by the system.

83. Fontana explains:

> In the embodiment shown the video processing module 608 includes a thumbnail extraction module 618 and an objects of interest module 620. The thumbnail extraction module 618 is arranged to generate thumbnails at possible locations the content provider would like to create an object of interest (for example a first frame, a last frame, and immediately following major scene or sound changes in the content). In some embodiments, the thumbnail extraction module 618 generates a series of thumbnails representing scenes throughout the multimedia content. The objects of interest module 620 generates one or more objects of interest as defined in metadata to be associated with the multimedia content. In various embodiments, the objects of interest module 620 can accommodate input from content providers to identify the objects of interest, or can at least partially automatically identify at least candidate objects of interest for confirmation by a user.

(Fontana ¶0095)

84. Fontana explains that the identification of "objects of interest" can occur automatically within the thumbnail images, specifically, "boundaries of a number of candidate objects of interest could be automatically detected within one or more thumbnails[.]" (Fontana ¶0142.) Fontana further teaches "**extracting one**

or more objects" by "generat[ing] a 'filmstrip' which is a strip of thumbnails containing 'objects of interest' from the video. These objects of interest can be items, people, or conditions in the video that the viewer may be interested in[.]" (Fontana ¶0148.)

85. Fontana additionally teaches "**identifying the one or more objects.**" Fontana discloses, for example, that objects can be identified using OpenCV or other tools, all of which would have been familiar to a POSA:

> After the content is received, a candidate object generation operation 906 generates candidate objects of interest from the multimedia content.... The candidate object generation module can be performed by any of a number of object recognition programs, including computer vision programs. Example computer vision tools include OpenCV, which is a library of motion tracking, facial recognition, gesture recognition, object identification, segmentation, and calibration tools. Other tools, such as MatLab or scale-invariant feature transform (SIFT) algorithms could be included in the object detection process as well.

(Fontana ¶0139.)

86. Fontana additionally teaches that objects could also be identified through "a neural network or other learning model to acquire knowledge of objects typically recognized or identified by users as objects of interest." (Fontana ¶0140.)

Fontana provides as examples of such neural networks those developed by Numenta, Inc.; Vidient Systems, Inc.; and Behavioral Recognition Systems, Inc. (*Id.*)

(c) "processing the audio files;" (Claim 1[c])

87. Fontana discloses "**processing the audio files**," such as processing associated with performing speech-to-text recognition. Fontana discloses an "audio processing module 606" that "is configured to process audio content associated with the multimedia content. In certain embodiments, the audio processing module 606 is configured to generate a full text transcript of the audio included in the multimedia content, to allow content consumers to search and review transcripts for the appearance of desired items." (Fontana ¶0091.) As discussed for claim 1[a] above, Fontana further discloses that this speech-to-text recognition may be performed on **audio files**.

(d) "converting audio files associated with the video to text;" (Claim 1[d])

88. As noted for claim 1[c], Fontana discloses "**converting audio files associated with the video to text**," such as performing speech-to-text recognition to convert audio from the multimedia content (which may be a video, as discussed for the preamble) to text. And as discussed for claim 1[a], this speech-to-text recognition may be performed on **audio files**.

(e) "converting the image files associated with the video to video data;" (Claim 1[e])

89. Fontana discloses "**converting the image files associated with the video to video data**," such as performing facial and/or object recognition to generate data and meta-data about the video content from the extracted thumbnail images.

90. As I discussed for claim 1[b] above, Fontana discloses "a thumbnail extraction module 618 and an objects of interest module 620." (Fontana ¶0095.) The "objects of interest module 620 generates one or more objects of interest as defined in metadata to be associated with the multimedia content." (*Id.*) Fontana also discloses a "candidate object generation operation" that "can generate a number of candidate objects of interest defined by the content provider." (Fontana ¶0141.) Additionally, "boundaries of a number of candidate objects of interest could be automatically detected within one or more thumbnails[.]" (Fontana ¶0142.) These modules are discussed in greater detail for claim 1[b]. A POSA would appreciate that the data generated by the objects of interest module and the candidate object generation operation are "**video data**" as they are data about the video content. (*See, e.g.*, '972, 1:10-11 ("The present invention relates to a method and a system for generating various and useful data from videos."); *id.*, 1:27-29 ("The present invention is generally directed to a method to generate data from video content, such

as text and/or image-related information.”).)

(f) “generating a topical meta-data that describes content of the video by deriving semantic information from the identification of the one or more objects and semantic information from the audio files;” (Claim 1[f])

91. Fontana discloses “**generating a topical meta-data that describes content of the video**” in the form of contextual information, an indexed transcript, keywords, and other meta-data derived from Fontana’s audio and video analyses.

92. The ’972 patent explains topical meta-data as follows:

> Further, the engine applies the topical meta-data to the original full video. The image topics can be stored as topics for the entire video or each image segment. The topic generation process can be repeated for all identifiable symbols in a video in a distributed process. The outcome would be several topical descriptors of the content within a video. An example of the aggregate information that would be derived using the above example would be understanding that the video presented a dog, which was jumping, on the beach, with people, by a resort.

(’972, 4:33-44.) Thus, the topical meta-data can include at least semantic understanding of identified objects (e.g., dog, beach, people, resort) and events (e.g., jumping).

93. Fontana generates metadata describing both the objects recognized in

the images and the audio processing. As to object recognition, Fontana explains that a "plurality of processing operations occur to generate object metadata, text metadata, and format the received multimedia content, for example to generate and store the various types of content-specific metadata described above." (Fontana ¶0120.) Thus, "an object metadata operation 806 generates object metadata corresponding to information about the content overall, as well as objects appearing in or mentioned in the multimedia content. For example, the object metadata can define the overall genre, title, producer, creation date, length or other characteristics of the multimedia content, but can also define people or objects appearing in the content as well." (*Id.*) A POSA would understand the object metadata to correspond to topical meta-data, in the form of identifying people and objects appearing in the video.

94. A POSA would further understand that Fontana discloses identifying topical meta-data in the form of what the '972 patent specification refers to as "events," that is, actions performed by the identified people or objects. Specifically, Fontana teaches that "[t]he candidate object generation module can be performed by any of a number of object recognition programs, including computer vision programs. Example computer vision tools include OpenCV, which is a library of motion tracking, facial recognition, gesture recognition, object identification,

segmentation, and calibration tools." (Fontana ¶0139.) Motions and gestures are events, such as "jumping," that the '972 patent specification identifies as topical meta-data. (*See* '972, 4:20-23, 4:41-44.)

95. For audio processing, "[a] text metadata operation 808 defines text metadata associated with the multimedia content. The text metadata can take any of a number of forms, and can include a transcript of audio data included in the multimedia content, as well as additional textual information that a content presenter would like to display alongside the streamed multimedia content, such as additional contextual information, advertisements, or hyperlinks to other websites or content." (Fontana ¶0121.) Notably, the text metadata can include both a transcript of the audio data, as well as additional text metadata for "additional contextual information." (*Id.*) Furthermore, "[t]he transcript can be indexed … to allow content consumers to search the spoken text transcript, as well as other descriptive information related to the multimedia content." (*Id.*) As a result, the text metadata can include additional text-based descriptive and contextual information about the video content, beyond the transcript of the audio.

96. For example, Fontana discloses identifying keywords in the transcript, which enables searching of the multimedia content: "The text index information 630 can be used to provide a corresponding transcript alongside playback of multimedia

content, or can be used to provide keyword searchability of the multimedia content." (Fontana ¶0107.) Fontana describes example keyword data for facilitating searching the multimedia content as follows:

> FIG. 7I illustrates example keyword data that can be used in association with particular content to facilitate searching of that content. In certain embodiments, the keyword data 706 can be used as a substitute for the text information 630, or can be used to reference a particular location within the text information to allow searching of content or metadata describing the content. In the embodiment shown, the keyword data 706 includes an identifier of the multimedia content as well as the keyword or keywords associated with that content. Other information can be included in the keyword data as well (e.g., links to a particular location within the multimedia content, or other associated keywords, etc.). In certain embodiments, the keyword data 706 can be made available to external search engines, to allow the content or portions of the content to be made available for search access by search engines that are remote from and unaffiliated with the systems and methods described herein.

(Fontana ¶0114.)

97. A POSA would have appreciated that Fontana's above-described object metadata and text metadata (including both transcript and additional descriptive or contextual information, such as keywords) are all "**topical meta-data that describes content of the video**."

98. Furthermore, Fontana's contextual information and meta-data, which correspond to the "**topical meta-data**" of claim 1, are generated "**by deriving semantic information from the identification of the one or more objects and semantic information from the audio files**." A POSA would understand "**semantic information**" to refer to information conveying or associated with the meaning of content. Fontana's topical meta-data is generated by deriving semantic information from the object identification and audio processing because it includes words and concepts ascribing meaning to the audio and video content. For example, a POSA would recognize keywords identified from a text-to-speech transcript as semantic information because they convey meaning about the audio. (*See, e.g.*, Fontana ¶0107 ("The text index information 630 can be used to provide a corresponding transcript alongside playback of multimedia content, or can be used to provide keyword searchability of the multimedia content.").) Likewise, a POSA would recognize identifications of people or objects as semantic information about the meaning of a video or image.

(g) "adding the topical meta-data to the video; and" (Claim 1[g])

99. Fontana discloses or renders obvious "**adding the topical meta-data to the video**." For example, Fontana teaches that "specific start and end times can

be defined, as associated with specific segments of the transcription text. In this way, the transcript could be linked, portion by portion, to the multimedia content based on the time at which the transcribed words are played in the content." (Fontana ¶0107.) A POSA would have appreciated that this specific linking of transcript to video "**add[s] the topical meta-data to the video**." Indeed, the '972 patent specification discloses a similar linking of topical meta-data to the video: "Data generated from an image and/or from audio transcription can be time stamped, for example, according to when it appeared, was heard, and/or according to the video frame from which it was pulled." ('972, 4:28-31.)

100. Similarly, Fontana teaches that "keyword data 706 can be used as a substitute for the text information 630, or can be used to reference a particular location within the text information to allow searching of content or metadata describing the content. In the embodiment shown, the keyword data 706 includes an identifier of the multimedia content as well as the keyword or keywords associated with that content. Other information can be included in the keyword data as well (e.g., links to a particular location within the multimedia content, or other associated keywords, etc.)." (Fontana ¶0114.)

101. Furthermore, Fontana teaches that "text index information 630 can be used to <u>provide a corresponding transcript alongside playback of multimedia</u>

content, or can be used to provide keyword searchability of the multimedia content." (Fontana ¶0107.) A POSA would have appreciated that providing a transcript alongside playback of a video was a form of "**adding the topical meta-data to the video**."

102. Fontana additionally teaches that the topical meta-data can be presented concurrently with the video content, which is yet another form of "**adding the topical meta-data to the video**." As Fontana explains, the "text metadata," which is **topical meta-data**, "can take any of a number of forms, and can include … additional textual information that a content presenter would like to display alongside the streamed multimedia content, such as additional contextual information, advertisements, or hyperlinks to other websites or content." (Fontana ¶0121.) Additionally, "[a]n action definition operation 914 allows a user to define one or more actions associated with each object of interest identified in the multimedia content. Any of a number of different types of actions can be defined. Example actions include display of contextual information identifying the object, as well as including click through actions such as a hyperlink to related content, or other sections of the same piece of content." (Fontana ¶0147.) A POSA would have appreciated that presenting text metadata alongside the video "**add[s] the topical meta-data to the video**."

103. Additionally, Fontana discloses a "container operation 810" that "applies a container to the received multimedia content." (Fontana ¶0122.) This "container operation 1304 can be performed by the multimedia processing systems of the present disclosure, with the container and associated metadata being stored either by the multimedia processing systems or managed by the content provider." (Fontana ¶0176.) Such containers were well-known to a POSA at the time, as they were a common way to store the various components of multimedia content such as videos. Fontana discloses some of these well-known container formats, such as "an Adobe Flash format" as well as "HTML5, Microsoft Silverlight, or other formats[.]" (Fontana ¶0122.)

104. A POSA would have known that container files such as those disclosed in Fontana routinely store metadata as well as multimedia content, and thus would have found it obvious to apply Fontana's container operation to both the multimedia content and its associated metadata. For example, Adobe Flash format, which Fontana discloses, was known to include storage of metadata. (*See, e.g.*, EX1008, Adobe Flash Video File Format Specification, https://rtmp.veriskope.com/pdf/video_file_format_spec_v10_1.pdf (describing Adobe Flash Format metadata options).) Furthermore, MP3 was a well-known container file format for audio, and a POSA would have known that "MP3 files are

capable of storing a certain amount of 'meta-data'—extra information about each file—inside the file itself…. These tags will be inserted automatically by most tools as you rip and encode, or can be added or edited later on, often directly through your MP3 player's interface." (EX1009, MP3: The Definitive Guide, p. 6; *see also id*., pp. 44-45, 105 (ID3 tags are "the extra space in MP3 files that let you store 'meta data' about a file"), 114 ("the ID3v2 specification allows for a huge array of meta-data storage capabilities in MP3 files"), 116 ("Every MP3 file has the ability to store 'meta-data' related to the track in the file itself, in the form of what are known as 'ID3 tags.'"), 363-64, 369.) As another example, U.S. Patent Application Publication No. 2011/0305394 ("Singer") discloses a method that "adds a metadata track to the image data (e.g., video data)" where the "added track includes the generated face detection metadata." (EX1010, Singer ¶0040.)

105. Disclosures in Fontana would have confirmed to a POSA that the container could include both the video and its topical meta-data. As noted above, Fontana explains that the "container operation 1304 can be performed by the multimedia processing systems of the present disclosure, with the container and associated metadata being stored … by the multimedia processing systems[.]" (Fontana ¶0176.) As a further example, Fontana teaches "[a] storage operation 812 stores the content and associated metadata for use." (Fontana ¶0123.) A POSA

would have appreciated that a container operation was one such storage operation.

106. Likewise, Fontana teaches that "[t]he data storage 410 can be configured to store any of a number of different types of data, <u>including the received multimedia content and data associated therewith</u>. In certain embodiments, the data storage 410 <u>includes a set of metadata associated with each piece of multimedia content</u> processed by the computing grid 408, for example as generated by processing the multimedia content." (Fontana ¶0066.)

(h) "cross-referencing the text and the video data based on the generated topical meta-data to determine topics;" (Claim 1[h])

107. A POSA would have found it obvious to combine Fontana with Lau to "**cross-referenc[e] the text and the video data based on the generated topical meta-data to determine topics**."

108. In the combination of Fontana with Lau, the **topical meta-data** generated by Fontana would be searched using the method of Lau to determine a relevant context for the video and ultimately identify advertisements for display at specific times during a video. This search process **cross-references the text and the video data** by matching advertisements using keywords from **the text** (the results of the speech-to-text recognition) and content from the **video data** (metadata on the identified objects of interest) **based on the generated topical meta-data** (i.e.,

by incorporating context and other metadata, *see* claim 1**[f]**).

109. Lau discloses performing a search to determine advertisements whose content and/or topics match the content and/or topics for portions of video content. The search involves counting the number of keyword and/or concept matches for a search term near a particular time in the video content. A POSA would have found it obvious to combine Lau and Fontana by using the **topical meta-data** generated by Fontana (*see* claims 1[f] and 1[g]) as well as **the text and the video data** (*see* claims 1[d] and 1[e]) as keywords and concepts searched by Lau for purposes of identifying ads or ad units to match to the video based on context or subject matter. Each of the ads, ad units, and context or subject matter of the video is a "**topic**." As discussed for claim 1[f], the **topical meta-data** (various contextual information and metadata) is derived from **the text and the video data**, such that searching the topical meta-data **cross-references the text and the video data**, and any such search is **based on the generated topical meta-data**. A POSA would have further found it obvious to use the text and the video data as part of Lau's matching search, which further **cross-references the text and the video data**. As a result, this combination "**cross-referenc[es] the text and the video data based on the generated topical meta-data to determine topics.**"

110. Specifically, Lau teaches matching keywords and concepts in a video

to ads based on context or subject matter: "Correlation engine 202 may relate a unique content ID with a time series of keywords and concepts (that advertisers have purchased), and in turn, relate the keywords and concepts to ads submitted by advertisers." (Lau ¶0051.) This matching aggregates together the different metadata types and uses the resulting aggregated information to match the content to an ad: "for each ad, correlation engine 202 finds candidate content that may be relevant. This is done by searching for content in the index to match the keywords, categories, and concepts associated with the ad to information in the content." (Lau ¶0056.) A POSA would have found it obvious to use the results of Fontana's speech-to-text recognition (**the text**) to generate keywords and other "information in the content" for Lau's search method. Similarly, a POSA would have found it obvious to use the results of Fontana's object recognition module (**the video data**) as concepts and other information in the content for Lau's search method. And because Lau's correlation engine searches "to match the keywords, categories, and concepts," it would further be obvious to perform Lau's search with reference to the contextual information derived from Fontana, which is the **topical meta-data** and is additionally "information in the content." (*See* Lau ¶0056.) Lau teaches that the search to match advertisements to video content may be performed by combining the results of different types of analyses, such as text with images as well as conceptual ideas like context: "Advertisers may buy correlation information, such as keywords, phrases or concepts, either through a bidding process or some other means, and submit their ads and related information to correlation engine 202 through correlation assistant 214. … The phrases may be <u>any combination of words and other information, such as symbols, images, etc.</u> The concepts may be a conceptual idea of something." (Lau ¶0046.) Thus, a POSA would have found it obvious that in the combination of Fontana with Lau, correlating advertisements to the video could **cross-reference the text** (speech-to-text transcript and any derived keywords) **and the video data** (the identified objects of interest) **based on the generated topical meta-data** (additional contextual information).

111. This contextually aware keyword and topic searching aligns with the '972 specification's teaching regarding cross-referencing:

> At 160, <u>the topics generated from an image or a frame and the topics extracted from audio can be combined. The text can be cross-referenced, and topics common to both texts would be given additional weights.</u> At 170, the video-to-text engine generates video text, such as text describing the content of the video, using the result of the combined texts and cross reference. <u>For example, key words indicating topic and semantic that appear in both texts can be selected or emphasized.</u>

('972, 5:46-54.)

112. Returning to Lau, Lau discloses using its method to **determine topics**, that is, subject matter for the ads and ad units. For each ad, Lau identifies a set of candidate locations within the video, then scores each based on the strength of the match:

> For each piece of candidate content associated with an ad, correlation engine 202 determines candidate times where the content may be relevant to the ad. Correlation engine 202 locates the times where the keywords and concepts match. For each candidate time, correlation engine 202 creates an "ad anchor" holding the score for the match. The score may be a linear combination of the following weights:
>
> 1. Probability of the keyword/concept match pulled from the recognition lattice.
>
> 2. Concentration of the match—the more keywords/concepts for the ad matches near the time, the higher the score. One embodiment of this score may be a count of the number of matches within a certain window of the current time….

(Lau ¶¶0057–0059.)

113. As I discuss in greater detail for claim 1[i] below, Lau additionally discloses that "[a]n advertisement may be broken into ad units" that are each "a subset of a larger advertisement. … Thus, advertiser system 106 is not restricted to

just serving an entire advertisement. Rather, the most relevant pieces of the advertisement may be selected from the matrix of ad units." (Lau ¶0023.) As a result, "[c]orrelation engine 202, when determining the advertisement, may <u>determine one or more ad units that correlate to the subject matter</u>," and "the 330 ad units may be combined and presented to the user" as the final advertisement. (Lau ¶0038.)

114. Each of searching for candidate video location matches for ads or ad units based on the subject matter, and assigning weights to those candidate locations, involves **cross-referencing the text and the video data based on the generated topical meta-data** to determine ad or ad unit matches based on subject matter or context ("**topics**").

115. <u>**Rationale and Motivation to Combine (Fontana with Lau)**</u>**:** A POSA would have been motivated to combine Fontana and Lau in order to use Lau's disclosures for how to place advertisements within a video to carry out Fontana's disclosures that advertisements can be linked to video content, and would have had a reasonable expectation of success in the combination. Fontana and Lau are analogous references to the '972 patent. Like the '972 patent, both Fontana and Lau are addressed to the field of processing and generating multimedia content, including video, and both specifically address how to search within such multimedia content.

(’972, 1:27-29 (“The present invention is generally directed to a method to generate data from video content, such as text and/or image-related information.”); ’972, Abstract (“The text and the video content can be cross-referenced with the video.”); Fontana ¶0001 (“[T]he present disclosure relates to systems and methods for processing and delivery of multimedia content.”); Fontana ¶0041 (“In general, the present disclosure relates to methods and systems for receipt, processing, and delivery of multimedia content, as well as enrichment of multimedia content for enhanced search and delivery.”); Lau ¶0002 (“Embodiments of the present invention generally relate to digital media and more specifically to displaying advertisements with rich media content.”); Lau ¶0006 (“In one embodiment, an advertisement is matched to subject matter in a portion of rich media content, such as <u>digital video</u>”); Lau ¶0053 (“The ads may be correlated to content in different ways. In one embodiment, keywords may be associated with each ad. Content may be searched to determine if the content includes the keywords.”).) Both further disclose linking advertisements to the multimedia content. (’972, 1:57-59 (“An advertisement can be placed at a specific time in the video based on the video content and/or section symbol of a video image.”); Fontana ¶0117 (“FIG. 7M illustrates example advertisement data 716 that can be used in association with multimedia content, to link one or more advertisements with multimedia content during playback.”); Lau

¶0002 ("Embodiments of the present invention generally relate to digital media and more specifically to displaying advertisements with rich media content.").)

116. Fontana expressly discloses that advertisement data could be linked to the multimedia content: "FIG. 7M illustrates example advertisement data 716 that can be used in association with multimedia content, to link one or more advertisements with multimedia content during playback." (Fontana ¶0117.) But Fontana does not teach how to identify advertisements to link with multimedia content; rather, "the matching of advertisements and content occurs based on a decision process separate from the content delivery system of the present disclosure." (Fontana ¶0117.) A POSA thus would have been motivated to combine Fontana with an advertisement-matching method, such as that disclosed by Lau, in order to practice Fontana's disclosures of linking advertisement information with the multimedia content.

117. A POSA would have had a reasonable expectation of success in combining Fontana and Lau. It would have been obvious to a POSA that the text and object meta-data generated by Fontana could be used as the keywords/content to search for matching advertisements to video content according to Lau. Fontana makes clear that its keywords are suitable for searching within the video, including with search engines outside of Fontana's disclosures—a POSA would have found it

obvious that Lau's method of searching for ad matches within its search engine is one such application of keyword-based searching within a search engine taught by Fontana. Fontana specifically discloses that its metadata can include an indexed transcript to enable searching, and can be linked to particular start and end times of the video: "The transcript can be indexed, as described below, to allow content consumers to search the spoken text transcript, as well as other descriptive information related to the multimedia content" (Fontana ¶0121) and "specific start and end times can be defined, as associated with specific segments of the transcription text. In this way, the transcript could be linked, portion by portion, to the multimedia content based on the time at which the transcribed words are played in the content" (Fontana ¶0107).

118. Fontana explicitly contemplates that its "text index information 630 … can be used to provide keyword searchability of the multimedia content." (Fontana ¶0107.) Fontana teaches that the content and metadata it generates can be searched through the use of keywords: "The enhanced multimedia content described in the present disclosure generally relates to multimedia content with associated interactive features, for example … associated transcript information linked to the multimedia content for keyword searching[.]" (Fontana ¶0042.) "FIG. 7I illustrates example keyword data that can be used in association with particular content to facilitate

searching of that content. In certain embodiments, the keyword data 706 can be used as a substitute for the text information 630, or can be used to reference a particular location within the text information to allow searching of content or metadata describing the content." (Fontana ¶0114.)

119. Fontana then teaches a "content request operation" that "receives a request related to multimedia content. The specific type of request received in the content request operation 814 can take a number of forms, such as a search query related to keywords appearing in one or more fields of metadata associated with the content (e.g., titles, authors, producers, genre, etc.) or in the transcript or other text associated with one or more pieces of content." (Fontana ¶0126.) Fontana responds to this request through a "provide metadata operation" that "provides metadata (and optionally the multimedia content) in response to the request. The provide metadata operation 816 selects at least a portion of the metadata associated with the content (e.g., including definitions of objects of interest, events, transcript information, position information, etc.) for inclusion with the content during playback." (Fontana ¶0127.)

120. Fontana further expressly contemplates that search engines, other than those expressly disclosed by Fontana, could also be used to search the keyword data generated by Fontana: "In certain embodiments, the keyword data 706 can be made

available to external search engines, to allow the content or portions of the content to be made available for search access by search engines that are remote from and unaffiliated with the systems and methods described herein." (Fontana ¶0114; *see also* Fontana ¶0127 ("In certain embodiments, the preference information can be provided to a remote decision engine that can then indicate a particular type, genre, or other grouping of enhancements to include with the multimedia content.").) Thus, a POSA would have appreciated that the **topical meta-data** generated by Fontana could be used as the keyword/content information for the video in Lau's search method, and would have had a reasonable expectation of success in the combination, as Fontana indeed encourages such combinations.

(i) "generating video text based on the cross-referencing, wherein the video text describes content of the video;" (Claim 1[i])

121. The combination of Fontana with Lau described above—using the topical metadata generated by Fontana to implement Lau's method of matching advertisements to video content—"**generat[es] video text based on the cross-referencing, wherein the video text describes content of the video**." The combination of Fontana and Lau is described for claim 1[h] above, including a motivation to combine and reasonable expectation of success. As relates to claim 1[i], Lau discloses that ads are matched to multimedia content through the **cross-**

referencing by determining a context or subject matter (which I will refer to for simplicity as a "context") ("**video text**") that **"describes content of the video**" and using that context to determine "ad units" from which the ads are generated. Additionally, Fontana discloses storing "advertisement data" including "topics, keywords, or content" that can be matched or related to the video content as additional "**video text**."

122. As discussed for claim 1**[h]**, the result of the cross-referencing are the ads, ad units, and context/subject matter of the video. Lau discloses "ad units" that are pieces of a larger advertisement, where each ad unit is associated with a concept. "An advertisement may be broken into ad units. An ad unit may be a subset of a larger advertisement. … Each ad unit may be associated with a concept. The ad units may be selected individually to form an advertisement." (Lau ¶0023.)

123. Lau explains that there are "[d]ifferent ways of creating an ad unit[.]" (Lau ¶0027.) As examples, an "ad unit may be created by taking a static ad and augmenting the unit with an advertiser-specified message dependent on **context** and keywords." (Lau ¶0027.) Furthermore, A POSA would have appreciated that identifying an "advertiser-specified message dependent on context and keywords" additionally includes a determination of "context," which is a way of describing what a video is about beyond just the keywords. In other words, a POSA would

have appreciated that "context" as used in Lau can be a form of **video text based on the cross-referencing** (derived from the text and video data, which are themselves the result of the audio and image processing) that **describes content of the video**.

124. Lau illustrates the concept of "ad units" with a BMW example that makes clear that Lau's "context" can be **video text based on the cross-referencing** that **describes content of the video**:

> Correlation engine 202, when determining the advertisement, may determine one or more ad units that <u>correlate to the subject matter</u>. For example, <u>based on one or more keywords</u>, ad units from the ad matrix are determined. The ad units are then combined into an advertisement that is correlated to the subject matter. One example of this is BMW may provide a general ad unit for their logo and have a different ad unit for different models, such as the 330 model, 530 model, etc. The logo unit and each of the model units can be combined at runtime based on the context of the content. <u>If the content talks about the 330 model then the logo and the 330 ad units may be combined and presented to the user.</u>

(Lau ¶0038.)

125. In this example, Lau determines that the "subject matter" of the video is the BMW 330 model. This subject matter (or context) is **video text** (BMW 330 model) that is **based on the cross-referencing** (identified "based on one <u>or more</u>

keywords," which can search text and image metadata (*see* Lau ¶0046), as the model appearing in the video content) and that **describes content of the video** ("the content talks about the 330 model"). (*See* Lau ¶0038.) A POSA would therefore have recognized that Lau teaches **generating video text based on the cross-referencing, wherein the video text describes content of the video**.

126. In the combination of Fontana and Lau, information about the advertisements selected through Lau's method are then incorporated into Fontana's "advertisement data" to facilitate display to the viewer:

> FIG. 7M illustrates example advertisement data 716 that can be used in association with multimedia content, to link one or more advertisements with multimedia content during playback. In the embodiment shown, the <u>advertisement data 716 can include an advertiser identifier, a definition of an advertisement, and associated topics, keywords, or content that can be linked to the advertisement</u>. In certain embodiments, <u>the advertisement data 716 is used to link the content to advertisements during playback;</u> in alternative embodiments, the advertisement data 716 is managed to track advertisements appearing with content[.]"

(Fontana ¶0117.)

127. A POSA would have appreciated that Fontana's "advertisement data" in combination with Lau is another form of **video text** (eg, "a definition of an

advertisement, and associated topics, keywords, or content") that is **based on the cross-referencing** (because derived from Lau's method of matching advertisements to video content) and that **describes content of the video** (such as data on "associated topics, keywords, or content" and other data "used to link the content to advertisements during playback").

(j) "generating a text, image, or animation based on the video text; and" (Claim 1[j])

128. The combination of Fontana with Lau described above—using the topical metadata generated by Fontana to implement Lau's method of matching advertisements to video content—"**generat[es] a text, image, or animation based on the video text**." The combination of Fontana and Lau is described for claim 1[h] above, including a motivation to combine and reasonable expectation of success. As relates to claim 1[j], Lau teaches that the "ad units" are selected based on context (which is **video text** as explained for claim 1[i]), and can then be combined together to generate an advertisement. Once the ad units are selected, through the process described for claim 1[i], "[t]he ad units are then combined into an advertisement that is correlated to the subject matter." (Lau ¶0038.) This advertisement may be **a text** (such as "an advertiser-specified message", Lau ¶0027)**, image** (such as the BMW logo, Lau ¶0038)**, or animation** (such as a "video that may serve as pre/mid/post-

roll", Lau ¶0027). The advertisement additionally is **based on the video text** as it is the result of combining the ad units, which are selected by matching to context (**video text**), together into an advertisement. (*See* Lau ¶0038.)

(k) "placing the text, image, or animation in the video." (Claim 1[k])

129. The combination of Fontana and Lau **plac[es] the text, image, or animation in the video**. The combination of Fontana and Lau is described for claim 1[h] above, including a motivation to combine and reasonable expectation of success. As relates to claim 1[k], Lau discloses that the advertisement matched to the video content (the "text, image, or animation") is placed in the video.

130. Lau discloses that the advertisement is placed in the video as pre-roll, mid-roll, post-roll, (that is, video content that plays before, during, or after another video) or even "injected" into the video itself. For example, Lau teaches that the ad units may be turned into "video that may serve as pre/mid/post-roll." (Lau ¶0027.) Lau further explains that "the advertisement may be displayed in serial, parallel, or be injected into the rich media content." (Lau ¶0034; *see also* Lau ¶0084 ("rendering formatter 204 can determine that an advertisement should be rendered serially relative to the portion of rich media content, in parallel to the portion of rich media content, or injected into the rich media content"); Lau claim 5 ("the advertisement

is injected into or laid on top of portion of rich media content").) A POSA would have appreciated that injecting the advertisement into the rich media content (i.e., video), **plac[es] the text, image, or animation in the video**.

2. Claim 2: "The method according to claim 1, further comprising: generating a content-rich video based on the video, the text, and the video data."

131. A POSA would have found claim 1 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, the combination of Fontana with Lau "**generat[es] a content-rich video based on the video, the text, and the video data**."

132. Specifically, a POSA would have understood that the end result of the combination of Fontana with Lau—a video indexed to searchable metadata and into which an advertisement has been added—is a **content-rich video**. As described for claim 1, above, this content-rich video is generated based on **the video** (the initial multimedia content, see claim 1[pre]), **the text** (converted from the audio files, see claim 1[d]), **and the video data** (the object metadata, see claim 1[e]).

3. Claim 3: "The method according to claim 1, further comprising: applying natural language processing to the text to determine context associated with the video."

133. A POSA would have found claim 1 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, Fontana renders obvious

"**applying natural language processing to the text to determine context associated with the video**."

134. As discussed for claim 1[d], Fontana teaches converting the audio files associated with the video to text in the form of a transcript. Fontana then further teaches analyzing this transcript using **natural language processing**: "The search performed within the content can in certain embodiments, be <u>performed based on natural language processing</u> of an existing transcript (closed captioning or subtitles file provided by the content provider) or <u>from a new transcript created using speech to text technology</u> and edited by the content provider." (Fontana ¶0184; *see also* Fontana ¶0100 ("In alternative embodiments, additional search arrangements can be included as well, such as a natural language search[.]").)

135. Fontana then determines the topic of the multimedia content—the **context associated with the video**—based on using natural language processing. Fontana additionally teaches that search queries, including those leveraging the above natural language processing of the text transcript, can be used **to determine context associated with the video**, specifically in the form of keywords or relevant portions of the multimedia content. Fontana explains, for example: "A content request operation 814 receives a request related to multimedia content. The specific type of request received in the content request operation 814 can take a number of

forms, such as a search query related to keywords appearing in one or more fields of metadata associated with the content (e.g., titles, authors, producers, genre, etc.) or in the transcript or other text associated with one or more pieces of content." (Fontana ¶0126.) Additionally, "[a] provide metadata operation 816 provides metadata (and optionally the multimedia content) in response to the request. The provide metadata operation 816 selects at least a portion of the metadata associated with the content (e.g., including definitions of objects of interest, events, transcript information, position information, etc.) for inclusion with the content during playback." (Fontana ¶0127.) A POSA would have appreciated that identifying metadata, portions of the transcript, and portions of the multimedia data related to keywords or other queries is **determin[ing] context associated with the video**.

136. The combination of Fontana with Lau further renders claim 3 obvious, as Lau too discloses using **natural language processing** to determine context associated with a video for purposes of matching it to an advertisement. As discussed previously, Lau teaches that "based on one or more keywords, ad units from the ad matrix are determined." (Lau ¶0038.) Lau additionally teaches that this process can build off of natural language processing: "Recognition engine 212 receives rich media content that may be accessed by a user and uses correlation recognition detection techniques to recognize the content....In another embodiment,

it could be [a] natural language processing engine." (Lau, ¶0041.) A POSA would have appreciated that using recognition detection techniques, such as natural language processing, to recognize the content is **applying natural language processing to the text to determine context associated with the video**.

4. Claim 4: "The method according to claim 2, further comprising: applying natural language processing to the text to extract the topical meta-data."

137. A POSA would have found claim 2 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, for the reasons I explain for claim 3, the combination of Fontana and Lau renders obvious "**applying natural language processing to the text to extract the topical meta-data.**" As discussed for claim 1[f], the **topical meta-data** in the combination of Fontana and Lau includes the text metadata and object metadata generated according to Fontana's method.

138. Specifically, a POSA would have appreciated that Fontana's disclosures of "a search query related to keywords appearing in one or more fields of metadata associated with the content (e.g., titles, authors, producers, genre, etc.) or in the transcript or other text associated with one or more pieces of content" to be a search within **the topical meta-data**. (Fontana ¶0126.) Fontana teaches that searching can **apply natural language processing**, which was also obvious in view of Lau. (*See* claim 3; Fontana ¶¶0100, 0184.) Fontana additionally discloses "[a]

provide metadata operation" that "provides metadata (and optionally the multimedia content) in response to the request." (Fontana ¶0127.) A POSA therefore would have appreciated that, in response to a keyword search query, the method could **apply natural language processing to the text** (transcript, which is part of the text metadata) **to extract the topical meta-data** (i.e., to provide metadata in response to the search request).

5. Claim 5: "The method according to claim 1, further comprising: processing the image files to extract additional text."

139. A POSA would have found claim 1 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, the combination of Fontana with Lau "**process[es] the image files to extract additional text**."

140. For example, Lau discloses "**extract[ing] additional text**" from **images**, such as by performing optical character recognition. Lau explains, "Image recognition can be used on visual portions of the rich media content. For example, <u>optical character recognition (OCR)</u>, facial recognition, object matching, etc." (Lau ¶0040.) A POSA would have appreciated that optical character recognition identifies text appearing in images and extracts that text.

141. It would have been obvious to a POSA that the optical character recognition of Lau could be performed on the **image files** of Fontana—namely, the

thumbnail images within which Fontana performs its "objects of interest module" and "candidate object generation operation." In addition to the motivations to combine explained for claim 1, a POSA would have been motivated to implement OCR, as taught by Lau, within the system of Fontana in order to improve the ability to generate metadata describing the video content and match ads to video content. A POSA implementing Fontana's method would have understood that various algorithms could be used to process the image files, as Fontana expressly teaches that "[o]nce a user has selected one or more objects of interest, a number of optional detection algorithms can be applied to further define those or other objects of interest." (Fontana ¶0144.) A POSA would have been motivated to implement OCR as one of those "optional detection algorithms" in order to capitalize on Lau's teaching that OCR can improve the matching of advertisements to the video content. Lau specifically teaches that OCR can be performed as one of the "[c]orrelation recognition detection techniques" that "may be used to determine that the advertisement is correlated to the portion of rich media content." (Lau ¶0040.) A POSA would have had a reasonable expectation of success because Fontana expressly discloses the use of additional detection algorithms. Additionally, OCR was well-known to a POSA in 2013 as one of many standard detection algorithms for image files.

6. Claim 6: "The method according to claim 5, wherein the additional text is generated by segmenting the image files before processing the image files in parallel."

142. A POSA would have found claim 1 obvious over the combination of Fontana with Lau, as I discuss above. Furthermore, Fontana discloses "**segmenting the image files before processing the image files in parallel**." Thus, in the combination of Fontana and Lau, a POSA would have understood that the image files (i.e., the thumbnail images) could be segmented before image processing, including the OCR processing described for claim 5 above.

143. Fontana teaches segmenting the image files in several ways. As a general matter, as described for claim 1[b], Fontana teaches a distributed computing system for video, audio, and image processing that segments all tasks into discrete portions that can be run in parallel: "As such, the various distributed computing systems described in FIGS. 1-5, above, <u>allow for segmenting the processing into discrete portions (e.g., audio, video processing separately, etc.) and parallel, pipelined processing of the data</u> to ensure fast content processing and resulting usability for content providers." (Fontana ¶0135.) As described for claim 1[b], Fontana's workflow server "received inbound data processing requests, for example from a content provider (as further discussed below) and <u>distributes one or more portions of jobs associated with each data processing request</u> to the integration

framework 204 and the storage network 206." (Fontana ¶0050.) The resulting system "allows creation of pipelined data processing systems within a distributed computing environment, <u>allowing computationally intensive jobs (e.g., video and audio content processing) to be distributed across a number of computing systems.</u>" (*Id.*) A POSA would have appreciated that this use of a distributed computing system **segments** tasks including image processing, prior to processing the images in parallel.

144. Furthermore, Fontana teaches segmenting image files even prior to distributing the image files to the workflow server for processing. Fontana discloses "generat[ing] a series of thumbnails representing scenes throughout the multimedia content" as part of "the thumbnail extraction module," which is run prior to performing the object recognition. (Fontana ¶0095.) This "series of thumbnails representing scenes" **segments the image files** into representative thumbnail images for each scene. Similarly, Fontana explains that "[i]n certain embodiments, the candidate object generation operation 906 splits the multimedia content into a plurality of sections, and generates a thumbnail image associated with each of those sections for preview by the content provider." (Fontana ¶0139.) This embodiment likewise **segments** the images by splitting the multimedia content into sections and generating a thumbnail image for each section.

145. Fontana also discloses that its image processing can use image processing algorithms to segment the images. For example, "[t]he candidate object generation module can be performed by any of a number of object recognition programs, including computer vision programs. Example computer vision tools include OpenCV, which is a library of motion tracking, facial recognition, gesture recognition, object identification, segmentation, and calibration tools." (Fontana ¶0139.) Likewise, Fontana explains that "[t]hese operations associated with each content provider can be, for example, instructions provided to a video or other multimedia-editing web service, for example to define specific elements of multimedia content, such as objects of interest appearing in the content, or to segment, edit, and reprocess the content." (Fontana ¶0058.) Because Fontana implements image processing in a distributed computing system (*see* claim 1[b]), each of these algorithms results in the segmented image files being processed in parallel.

7. Claim 7: "The method according to claim 1, further comprising: determining a motion associated with the one or more objects."

146. A POSA would have found claim 1 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, Fontana discloses or renders obvious "**determining a motion associated with the one or more objects.**"

147. Specifically, Fontana teaches that "[t]he candidate object generation module can be performed by any of a number of object recognition programs, including computer vision programs. Example computer vision tools include OpenCV, which is a library of <u>motion tracking</u>, facial recognition, <u>gesture recognition</u>, object identification, segmentation, and calibration tools." (Fontana ¶0139.) Both "motion tracking" and "gesture recognition" involve **determining a motion associated with the one or more objects**.

148. To the extent Fontana does not expressly teach the limitation of claim 7, it would have been obvious to implement the motion tracking and gesture recognition algorithms of OpenCV—as disclosed in Fontana—as among the "optional detection algorithms" that "can be applied to further define those or other objects of interest." (Fontana ¶0144.)

8. Claim 8: "The method according to claim 1, further comprising segmenting the audio files before processing the audio files in parallel."

149. A POSA would have found claim 1 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, a POSA would have found it obvious in light of Fontana's parallel processing disclosures to "**segment[] the audio files before processing the audio files in parallel**" in order to take advantage of Fontana's teaching that multiple servers can be used to efficiently carry out tasks

like audio processing.

150. Specifically, Fontana teaches that multiple servers can be configured to perform audio processing, and that processing jobs can be separately performed across multiple of these servers in parallel:

> In certain embodiments, the servers 302a-c are specifically designed according to the application the network 300 is intended to support; for example in the case where multimedia data is to be processed using the computing capabilities within network 300, <u>one or more of the servers 302a-c can include specific graphical processing units for processing</u> lower level video, image or <u>audio algorithms</u>. Other specific capabilities can be included into the servers 302a-c as well. <u>The servers 302a-c are configured to share processing jobs, such that tasks can be</u> performed by one or more of the computing systems, or <u>separated and performed across multiple computing systems in parallel</u>.

(Fontana ¶0053.) The "servers 302a-c can correspond to any of the various computing systems 208 of FIG. 2, in that <u>any of those computing systems can perform all or a portion of a processing job as defined by a scheduling algorithm, allowing multimedia content to be processed efficiently when necessary</u>." (Fontana ¶0055.)

151. Fontana additionally teaches that segmentation may occur in order to facilitate parallel processing. For instance, Fontana explains that "operations

associated with each content provider can be, for example, instructions provided to a video or other multimedia-editing web service, for example … to segment, edit, and reprocess the content.” (Fontana ¶0058.) Similarly, Fontana teaches that “the various distributed computing systems described in FIGS. 1-5, above, allow for segmenting the processing into discrete portions (e.g., audio, video processing separately, etc.) and parallel, pipelined processing of the data to ensure fast content processing and resulting usability for content providers.” (Fontana ¶0135.)

152. Fontana thus discloses parallel processing of audio files across multiple servers, and segmenting data to facilitate parallel processing. A POSA would thus have found it obvious that to implement audio processing in parallel, the audio files would first be segmented. In other words, in order for the servers to “separate[]” an audio processing task to “perform[] across multiple computing systems in parallel,” the audio files could first be segmented in order to separate them across the servers. (*See* Fontana ¶0053.) It thus would have been obvious to a POSA based on Fontana to **segment[] the audio files before processing the audio files in parallel**.

9. Claim 9: “The method according to claim 7, wherein the audio files and the image files are segmented at spectrum thresholds.”

153. A POSA would have found claim 7 obvious over the combination of Fontana with Lau, as I discuss above. Likewise, for the reasons discussed for claims

6 and 8, a POSA would have found it obvious over Fontana that "**the audio files and the image files are segmented**."

154. A POSA would further have found it obvious over Fontana to perform the segmentation of audio and image files **at spectrum thresholds** in order to, for example, segment the audio and image files based on scene or content changes. Fontana teaches that "[t]he thumbnail extraction module 618 is arranged to <u>generate thumbnails at possible locations</u> the content provider would like to create an object of interest (for example a first frame, a last frame, and <u>immediately following major scene or sound changes in the content</u>). In some embodiments, the thumbnail extraction module 618 <u>generates a series of thumbnails representing scenes throughout the multimedia content</u>." (Fontana ¶0095.)

155. A POSA would have appreciated both that this division of the multimedia content by "major scene or sound changes in the content" corresponds to a segmentation of the multimedia content, and that identifying "major scene or sound changes in the content" could be performed by identifying **spectrum thresholds**. A POSA would have understood that a spectrum threshold is just a quantifiable valuation of some aspect of the media content at each moment in time – for example, the overall volume at a given time or frame of video, or its overall color saturation value. The '972 patent, for example, gives the example that "the

audio data can be processed and converted into a spectrum. Locations where the spectrum volatility is below a threshold can be detected and segmented. Such locations can represent silence or low audio activities in the audio data." ('972, 5:16-20.) A POSA would have known that one common method of detecting major scene or sound changes is to look for changes above a given **threshold** in these quantifiable valuations of the multimedia content. Such approaches were common in the literature and in practice well before 2013 and would have been well-known to a POSA. U.S. Patent No. 9,407,942 to Andrew Brenneman ("Brenneman"), which claims priority to October 3, 2008, for example, described the use of a "discontinuity profile" that "indicates the degree of discontinuity of each frame of the digital video content" and that "may be indicative of a change of scene … or another point of segmentation of the digital video content." (Brenneman, 11:44-54.) Likewise, U.S. Patent Publication No. 2004/0125877 A1 to Shin-Fu Chang et al. ("Chang"), which was filed on April 9, 2001, teaches "decomposition of video sequences into short shots by detecting a discontinuity in visual and/or audio features," where scene changes may be detected for example "if the frame-to-frame color difference ratio is larger than a given threshold[.]" (Chang ¶¶0008, 0093.) More generally, a POSA would have known that performing a speech-to-text analysis, such as that taught by Fontana, commonly involved performing a Fourier analysis of the audio and segmenting the audio based on frequency information in order to divide the speech into individual phonemes. This process of dividing speech by phonemes is described in Fontana: "Phonetic-based applications separate conversations into phonemes, the smallest components of spoken language; they then find segments within the long file of phonemes that match a phonetic index file representation of target words, phrases and concepts[.]" (Fontana ¶0169; *see also* Fontana ¶0171 (listing example "[p]honetic-based applications useable as one or more of the speech to text conversion programs").) A POSA would have appreciated that these phonetic-based algorithms described in Fontana **segment audio files at spectrum thresholds**.

156. Additionally, a POSA would have known that periods of silence with low energy are commonly identified as gaps between words. A POSA would have known that identifying the gaps between words was used, for example, in the "large vocabulary continuous speech recognition (LVCSR) engines" disclosed by Fontana. (*See* Fontana ¶0169 ("LVCSR engines depends on a language model that includes a vocabulary/dictionary for speech-to-text conversion of audio files. The text file is then searched for target words, phrases and concepts."); Fontana ¶0170 (listing "[e]xample sources of speech to text conversion programs performing LVCSR-based conversions").) This common understanding of the use of silence to segment audio aligns with the '972 patent's disclosure:

> In some embodiments, the segmentation can be performed by a fixed period of time. In another example, quiet periods in the audio can be detected, and the segmentation can be defined by the quiet periods. For example the audio data can be processed and converted into a spectrum. Locations where the spectrum volatility is below a threshold can be detected and segmented. Such locations can represent silence or low audio activities in the audio data. The quiet periods in the audio data can be ignored, and the processing requirements thereof can be reduced.

('972, 5:13-22.) Therefore, a POSA would have found it obvious to use a spectrum threshold for segmenting audio and image files based on scene and/or sound changes, as taught by Fontana.

10. Claim 10: "The method according to claim 1, further comprising: generating an advertisement based on the text and the video data."

157. A POSA would have found claim 1 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, as explained for claim 1, the combination of Fontana and Lau **generat[es] an advertisement based on the text and the video data**. *See* claim 1[j].

158. Specifically, the advertisement generated by Lau is created by compiling ad units, each of which is selected by matching the ad units to the video based on the text (i.e., the audio transcript) and the video data (i.e., the metadata

generated by Fontana). The details of this matching and advertisement generation are explained above for claim 1.

11. Claim 11: "The method according to claim 10, further comprising: placing the advertisement in the video at a preferred time."

159. A POSA would have found claim 10 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, the combination of Fontana with Lau **plac[es] the advertisement in the video at a preferred time**.

160. Specifically, Lau teaches that its "advertisements are <u>time aligned to correlate to the subject matter</u> to maximize the revenue." (Lau ¶0045.) A POSA would have appreciated that time-aligned advertisements must be **placed in the video at a preferred time**. This understanding is consistent with Lau's teachings on how to place advertisements in the video.

161. As discussed for claim 1[k], Lau discloses that "the advertisement may be displayed in serial, parallel, or be injected into the rich media content." (Lau ¶0034; *see also* Lau ¶0084 ("rendering formatter 204 can determine that an advertisement should be rendered serially relative to the portion of rich media content, in parallel to the portion of rich media content, or injected into the rich media content"); Lau claim 5 ("the advertisement is injected into or laid on top of portion of rich media content").) A POSA would have appreciated that injecting the

advertisement into the video **places the advertisement in the video at a preferred time**.

162. Lau likewise discloses that the ad units may be turned into "video that may serve as pre/mid/post-roll." (Lau ¶0027.) Playing an advertisement as pre-roll, mid-roll, or post-roll each likewise **place the advertisement in the video at a preferred time**.

12. Claim 12: "The method according to claim 6 wherein the additional text includes information regarding context associated with the video."

163. A POSA would have found claim 6 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, it would have been obvious that in the combination of Fontana with Lau, **the additional text**—in this combination, the result of the OCR process on the image files—**includes information regarding context associated with the video**. It would have been obvious to a POSA that text generated from the OCR of image files is "information regarding context associated with the video."

164. Lau confirms this understanding, as it explains that its OCR process is performed in order to determine the content of the rich media for purposes of matching the content to an advertisement. A POSA would understand that this corresponds to a "context associated with the video." As Lau teaches:

> Correlation recognition detection techniques may be used to determine that the advertisement is correlated to the portion of rich media content. For example, keywords may be detected in the rich media using audio recognition. Audio recognition may include speech recognition, music detection on music portions, sound effect detection on sound effects, etc. Other techniques for keyword detection can include using preset word tags or indicators in the rich media content. Image recognition can be used on visual portions of the rich media content. For example, optical character recognition (OCR), facial recognition, object matching, etc. Other recognition techniques can be employed. For example, any suitable way of determining the content of rich media can be used to correlate a portion of the rich media content to an advertisement.

(Lau ¶0040.)

13. Claim 13: "The method according to claim 6, wherein the additional text relates to a symbol appearing in the video."

165. A POSA would have found claim 6 obvious over the combination of Fontana with Lau, as I discuss above. For the same reasons as discussed for claims 5 and 6, a POSA would have found claim 13 obvious over the combination of Fontana with Lau. The "additional text" of claims 5, 6, and 13 in the combination of Fontana with Lau is the result of performing OCR on the images, as taught by Lau. A POSA would understand that OCR includes symbols such as letters,

numbers, punctation, and other typographic symbols (e.g., "+"). Thus it would have been obvious to a POSA that **the additional text relates to a symbol appearing in the video**.

14. Claim 14: "The method according to claim 13, wherein the symbol is a brand logo, and wherein the additional text includes information regarding placement and time of appearance of the brand logo."

166. A POSA would have found claim 13 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, a POSA would have found it obvious that the symbol could be a brand logo, and that the additional text includes information regarding placement and time of appearance of the brand logo.

167. A POSA would have understood that the object recognition and OCR processing of the combination of Fontana with Lau could encompass recognizing brands and generating metadata and/or keywords corresponding to those brands. The '972 patent explains that techniques such as object recognition and OCR can be used to identify brands:

> <u>For example, the normalized image frame files can be analyzed for text identification and/or by optical character recognition</u>. The data can be improved through a dictionary verification step. Various maps can be created based on edge detection and/or image segmentation techniques.

> Such techniques can be improved by focusing on regions of interest, for example based on brands, logos, objects, and/or features of interest.

('972, 6:62-7:2.)

168. A POSA would have found such a use of object recognition and/or OCR obvious in view of Fontana and Lau. In fact, Fontana expressly discloses that "[a]dditional objects of interest can be identified by a user," thus permitting a user to specify, for example, a brand logo for the method to recognize. (Fontana ¶0089.) Similarly, Fontana explains that additional objects to recognize can be specified with a script provided with the multimedia content:

> In some embodiments, a content provider can provide a script alongside the multimedia content 602 to the system 600. In such embodiments, the script can contain a number of descriptions of the content, such as dialog occurring in the content, objects and individuals appearing in the content, as well as mood, scene, and other information that can be used at least in part to assist in generating metadata describing the content for use in connection with the systems and methods of the present disclosure.

(Fontana ¶0088.) A POSA would have found it obvious that a brand logo could be one such object specified by a script as appearing within a video.

169. Furthermore, Lau expressly teaches that brands may be a particularly

useful type of content to search for within the video in order to match ad units to specific locations within the video. Lau gives the example of identifying ad units corresponding to the BMW model discussed in the video:

> Correlation engine 202, when determining the advertisement, may determine one or more ad units that correlate to the subject matter. For example, based on one or more keywords, ad units from the ad matrix are determined. The ad units are then combined into an advertisement that is correlated to the subject matter. <u>One example of this is BMW may provide a general ad unit for their logo and have a different ad unit for different models, such as the 330 model, 530 model, etc. The logo unit and each of the model units can be combined at runtime based on the context of the content. If the content talks about the 330 model then the logo and the 330 ad units may be combined and presented to the user.</u>

(Lau ¶0038.) A POSA combining Fontana with Lau therefore would have found it obvious to use brand logos as objects of interest for identification within the image files. A POSA would have been motivated to do so in order to achieve the benefits described by Lau of improved matching of ad units to the video content, and would have had a reasonable expectation of success in light of Fontana's disclosures discussed above that objects of interest may be specified by the user and/or an accompanying script.

170. As explained for claims 1[e], 1[f], and 5, the recognized objects (including the results of OCR and any brand logos) are converted into searchable metadata, i.e., **additional text**. And as I explained for claim 1[h], Fontana teaches that this metadata is linked to where in the video content it occurs, such that **the additional text includes information regarding placement and time of appearance of the brand logo**.

15. Claim 15: "The method according to claim 1, wherein the one or more objects are letters appearing in the video."

171. A POSA would have found claim 1 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, for the same reason as discussed above for claim 5, it would have been obvious to a POSA to implement OCR as taught by Lau as part of the combination of Fontana with Lau. A POSA would have known that the **one or more objects** identified by OCR would include **letters appearing in the video**.

16. Claim 16: "The method according to claim 6, wherein the additional text relates to faces appearing in the video."

172. A POSA would have found claim 6 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, a POSA would have found it obvious that in the combination of Fontana with Lau, the **additional text** could **relate to faces appearing in the video** as both Fontana and Lau teach the use of

facial recognition as part of the image processing.

173. Fontana explains that "[t]he candidate object generation module can be performed by any of a number of object recognition programs, including computer vision programs. Example computer vision tools include OpenCV, which is a library of motion tracking, facial recognition, gesture recognition, object identification, segmentation, and calibration tools." (Fontana ¶0139.) Fontana also teaches the use of "additional detection algorithms" that "can include facial recognition[.]" (Fontana ¶0144.)

174. Similarly, Lau teaches that, "[i]mage recognition can be used on visual portions of the rich media content. For example, optical character recognition (OCR), facial recognition, object matching, etc." (Lau ¶0040.) These image recognition algorithms, including facial recognition, are used to generate meta-data ("**additional text**"): "In the video or visual recognition embodiment, meta-data about the visual content is generated or culled from the content itself." (Lau ¶0042; *see also* Lau ¶0040 ("Other techniques for keyword detection can include using preset word tags or indicators in the rich media content. Image recognition can be used on visual portions of the rich media content. For example, … facial recognition[.]").)

17. Independent Claim 17: "A system for extracting data from a video, comprising:" (Claim 17[pre])

175. Assuming the preamble provides a claim limitation, Fontana discloses it. For the same reasons explained for claim 1, and in particular limitations 1[pre] and 1[a], Fontana discloses a system for extracting data from a video. To the extent the preamble of claim 1 is addressed to a "method" and the preamble of claim 17 is addressed to a "system," Fontana teaches both a method and a system. (*See* Fontana ¶0008 ("In a first aspect, a method for providing multimedia content is disclosed."); Fontana ¶0009 ("In a second aspect, a system for providing multimedia content is disclosed.").)

(a) "a computer processor having a plurality of processors for parallel processing; and" (Claim 17[a])

176. I explained above for claim 1[b] that Fontana discloses a distributed computing system for parallel processing. Fontana further discloses that this system is implemented using a **computer processor having a plurality of processors for parallel processing**:

> In addition, electronic computing device 500 comprises a processing unit 504. As mentioned above, a processing unit is a set of one or more physical electronic integrated circuits that are capable of executing instructions. In a first example, processing unit 504 may execute software instructions that cause electronic computing device 500 to

> provide specific functionality. In this first example, processing unit 504 may be implemented as one or more processing cores and/or as one or more separate microprocessors.

(Fontana ¶0074.)

(b) "a non-transitory computer readable medium containing instructions directing the system to execute the steps of:" (Claim 17[b])

177. Fontana teaches that its system is implemented on **a non-transitory computer readable medium containing instructions directing the system to execute** certain steps. Fontana explains:

> FIG. 5 is a block diagram illustrating example physical components of an electronic computing device 500, which can be used to execute the various operations described above, and provides an illustration of further details regarding any of the computing systems described above in FIGS. 1-4. A computing device, such as electronic computing device 500, typically includes at least some form of computer-readable media. Computer readable media can be any available media that can be accessed by the electronic computing device 500. By way of example, and not limitation, computer-readable media might comprise computer storage media and communication media.

(Fontana ¶0072.) Fontana further explains that this computer readable media can contain computer readable **instructions**, and can be implemented in forms of memory such as RAM, ROM, EEPROM, and flash memory that a POSA would have recognized as **non-transitory computer readable media**:

> In the context of the electronic computing device 500, computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, various memory technologies listed above regarding memory unit 502, non-volatile storage device 510, or external storage device 516, as well as other RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium that can be used to store the desired information and that can be accessed by the electronic computing device 500.

(Fontana ¶0078.)

(c) "converting audio associated with the video to text; " (Claim 17[c])

178. Fontana discloses **converting audio associated with the video to text** for the same reasons discussed above for claim 1[d].

(d) "converting images associated with the video to video data; " (Claim 17[d])

179. Fontana discloses **converting images associated with the video to**

video data for the same reasons discussed above for claim 1[e].

(e) "generating the video data by segmenting image files of the video before processing the image files in parallel; " (Claim 17[e])

180. Fontana discloses **generating video data** for the reasons discussed in claim 1[e], and **processing the image files in parallel**, for the reasons discussed for claim 1[b]. For the same reasons as discussed above for claim 6, Fontana further discloses **segmenting image files of the video before processing the image files in parallel**.

(f) "identifying one or more objects in the image files; " (Claim 17[f])

181. For the reasons discussed above for claim 1[b], Fontana discloses **identifying one or more objects in the image files**.

(g) "generating data topics, from the text and the video data, that describe content of the video by deriving semantic information from the identification of the one or more objects and semantic information from the audio; " (Claim 17[g])

182. For the same reasons discussed above for claim 1[f], Fontana discloses **generating data topics, from the text and the video data, that describe content of the video by deriving semantic information from the identification of the one or more objects and semantic information from the audio**. My analysis for claim

1[f] explains that a POSA would have appreciated that the various metadata generated by Fontana correspond to the "topical meta-data that describes content of the video" within claim 1 of the '972 patent. A POSA would equally recognize that Fontana's generated metadata also correspond to "**data topics … that describe content of the video**." For example, Fontana's object metadata identifying objects are "data topics," as are keywords or other terms derived from the text metadata.

(h) "adding the data topics to the video as meta-data; " (Claim 17[h])

183. For the same reasons discussed above for claim 1[g], Fontana discloses **adding the data topics to the video as meta-data**, including by linking the data topics (i.e., Fontana's text **metadata** and object **metadata**) to specific times in the video, synchronizing the metadata for simultaneous display to the user, and storing the video with its metadata. (*See* claim 1[g].)

(i) "cross-referencing the text, the video data, and the topics with the video based on the generated data topics; " (Claim 17[i])

184. It would have been obvious to combine Fontana with Lau to **cross-referenc[e] the text, the video data, and the topics with the video based on the generated data topics** for substantially the same reasons as I explained above for claim 1[h]. The combination of Fontana and Lau is the same as for claim 1, a POSA

would have been motivated to combine the references for the same reasons as discussed for claim 1, and a POSA would have had a reasonable expectation of success for the same reasons discussed for claim 1. The differences in wording between claim 1[h] and 17[i] do not materially alter the analysis. I address these below.

185. First, claim 1[h] requires cross-referencing "the text and the video data" whereas claim 17[i] requires cross-referencing "the text, the video data, and the topics with the video." As I explained for claim 1[h]:

> Lau discloses performing a search to determine advertisements whose content and/or topics match the content and/or topics for portions of video content. The search involves counting the number of keyword and/or concept matches for a search term near a particular time in the video content. A POSA would have found it obvious to combine Lau and Fontana by using the **topical meta-data** generated by Fontana (*see* claims 1[f] and 1[g]) as well as **the text and the video data** (*see* claims 1[d] and 1[e]) as keywords and concepts searched by Lau for purposes of identifying ads or ad units to match to the video based on context or subject matter. Each of the ads, ad units, and context or subject matter of the video is a "**topic**." As discussed for claim 1[f], the **topical meta-data** (various contextual information and metadata) is derived from **the text and the video data**, such that searching the topical meta-data **cross-references the text and the video data**, and any such search is

> **based on the generated topical meta-data**. A POSA would have further found it obvious to use the text and the video data as part of Lau's matching search, which further **cross-references the text and the video data**. As a result, this combination "**cross-referenc[es] the text and the video data based on the generated topical meta-data to determine topics.**"

186. What claim 1[h] refers to as "topical meta-data" corresponds to the **topics** in claim 17[i], such that for the same reasons discussed for claim 1[h] above, the combination of Fontana with Lau **cross-referenc[e]s the text, the video data, and the topics**.

187. Furthermore, this occurs "**with the video**." Specifically, as I explained for claim 1[h], Lau's search to match ad units to the content identifies a set of candidate locations within the video, then scores each based on the strength of the match:

> For each piece of candidate content associated with an ad, correlation engine 202 determines candidate times where the content may be relevant to the ad. <u>Correlation engine 202 locates the times where the keywords and concepts match. For each candidate time, correlation engine 202 creates an "ad anchor" holding the score for the match.</u> The score may be a linear combination of the following weights:

> 1. Probability of the keyword/concept match pulled from the recognition lattice.
>
> 2. Concentration of the match—<u>the more keywords/concepts for the ad matches near the time, the higher the score.</u> One embodiment of this score may be a count of the number of matches within a certain window of the current time….

(Lau ¶¶0057–0059.) Thus, the same combination described for claim 1[h] **cross-referenc[es] the text, the video data, and the topics with the video.**

188. Finally, claim 1[h]'s cross-referencing is "based on the generated topical meta-data to determine topics" whereas claim 17[i]'s cross-referencing is "**based on the generated data topics**." Here, the topical meta-data for claim 1[h] corresponds to the data topics for claim 17[i], such that this element of claim 17[i] is obvious over Fontana with Lau for the same reasons as claim 1[h].

(j) "generating a text, image, or animation based on the data topics; and" (Claim 17[j])

189. Claim 17[j] essentially combines the end of claim 1[h] ("to determine topics"), claim 1[i], and claim 1[j] into a single limitation – that the system **generat[e] a text, image, or animation based on the data topics**. For the same reasons discussed above for claims 1[h], 1[i], and 1[j], claim 17[j] would have been obvious over the combination of Fontana with Lau.

(k) "placing the text, image, or animation in the video." (Claim 17[k])

190. For the same reasons discussed above for claim 1[k], Fontana discloses **placing the text, image, or animation in the video**.

18. Claim 18: "The system according to claim 17, wherein converting the audio comprises natural language processing."

191. A POSA would have found claim 17 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, for the same reason as discussed above for claim 3, the combination of Fontana with Lau renders obvious that **converting the audio comprises natural language processing**, as both Fontana and Lau disclose natural language processing as part of the audio processing and resulting search. (*See* claim 3, *supra*.)

19. Claim 19: "The system according to claim 17, the computer directs the audio to be converted by at least one node of a cluster and the computer directs the images to be converted by at least one other node of the cluster in parallel."

192. A POSA would have found claim 17 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, for the same reason as discussed above for claim 1[b], Fontana discloses that **the computer directs the audio to be converted by at least one node of a cluster and the computer directs the images to be converted by at least one other node of the cluster in parallel**.

193. As also explained for claim 1[b], Fontana teaches a distributed computer system for parallel processing:

> In the embodiment shown, <u>the distributed computing network 300 includes a plurality of computing systems, illustrated as servers 302a-c.</u> The servers 302a-c are communicatively interconnected, and each includes a corresponding data storage system 304a-c. The servers 302a-c share a distributed memory cache 306, and are each capable of accessing a shared cache of memory that is not residing in any of data storage systems 304a-c. The servers 302a-c are interfaced to inbound work, such as from a scheduler system (as described in further detail in connection with FIG. 4, below) for coordination and communication of data for processing.
>
> In certain embodiments, the servers 302a-c are specifically designed according to the application the network 300 is intended to support; for example in the case where multimedia data is to be processed using the computing capabilities within network 300, <u>one or more of the servers 302a-c can include specific graphical processing units for processing lower level video, image or audio algorithms</u>. Other specific capabilities can be included into the servers 302a-c as well. <u>The servers 302a-c are configured to share processing jobs, such that tasks can be performed by one or more of the computing systems, or separated and performed across multiple computing systems in parallel.</u>

(Fontana ¶¶0052, 0053.) Fontana illustrates this system in Figure 3:

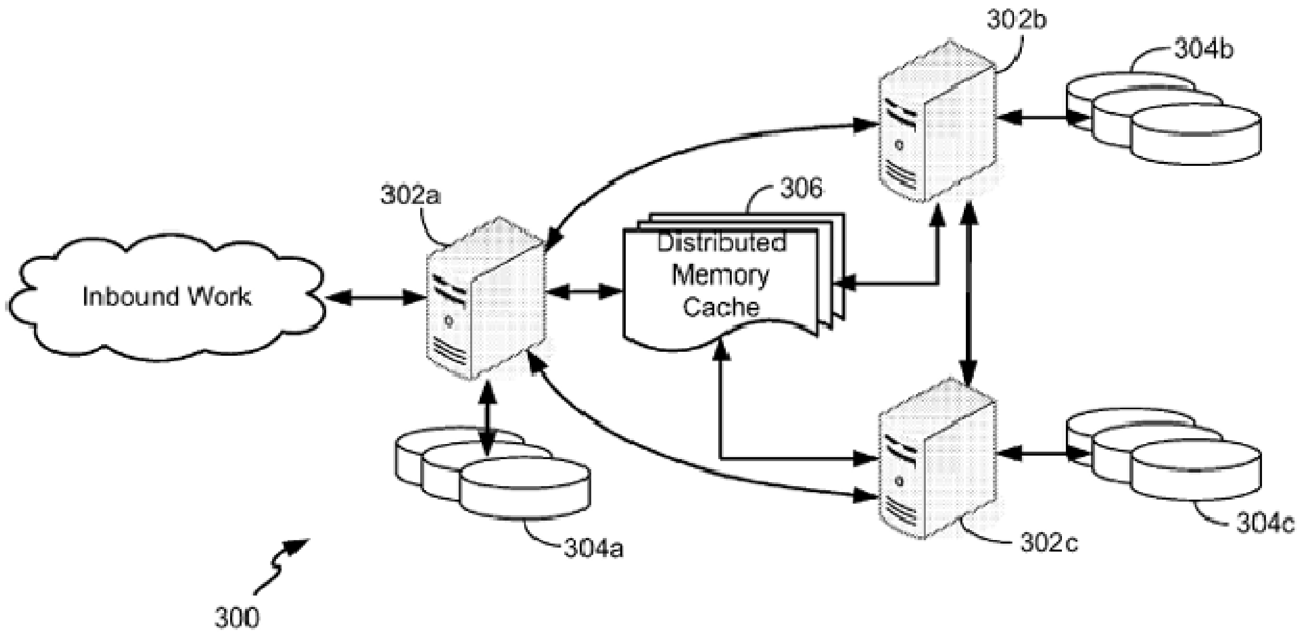


FIG. 3

194. A POSA would have appreciated that each of the servers, such as servers 302a-c in Figure 3, represents **at least one node of a cluster**, such that the distributed computing system of Fontana directs processing tasks to at least one node of a cluster.

195. Fontana further explains that this distributed computing system and network of servers "allow for segmenting the processing into discrete portions (e.g., audio, video processing separately, etc.) and parallel, pipelined processing of the data to ensure fast content processing and resulting usability for content providers."

(Fontana ¶0135.) From this, a POSA would have appreciated that audio and images would be separated and processed in parallel, such that Fontana discloses that **the computer directs the audio to be converted by at least one node of a cluster and the computer directs the images to be converted by at least one other node of the cluster in parallel**.

20. Claim 20: "The server according to claim 17, wherein the audio and the images are segmented at spectrum thresholds."

196. A POSA would have found claim 17 obvious over the combination of Fontana with Lau, as I discuss above. Additionally, for the same reason as discussed above for claim 9, it would have been obvious to a POSA to **segment the audio and the images at spectrum thresholds**.

C. Ground 2: Claims 8-9 and 20 Are Obvious Over Fontana in view of Lau and Arakawa

1. Claim 8: "The method according to claim 1, further comprising segmenting the audio files before processing the audio files in parallel."

197. A POSA would have found claim 1 obvious over the combination of Fontana with Lau, as I discuss above for Ground 1. Additionally, a POSA would have found it obvious to combine Fontana with Arakawa to "**segment[] the audio files before processing the audio files in parallel**" in order to take advantage of Fontana's teaching that multiple servers can be used to efficiently carry out tasks

like audio processing while also increasing efficiency by processing voice clips separately from background noise.

198. Arakawa teaches **segmenting audio** by frame, and further segmenting audio files into voice sections (corresponding to speech) versus non-voice sections corresponding to other noise, prior to performing speech recognition on the voice sections. The result is "a voice recognition system capable of, while suppressing negative influences from sound not to be recognized, correctly estimating utterance sections that are to be recognized." (Arakawa, Abstract.) Arakawa's segmentation into voice versus non-voice sections is performed by comparing each frame to a threshold value, such as Arakawa explains:

> The voice segmentation unit 103 calculates a voice segmentation feature value which indicates possibility of being voice for each frame input sound data. Then, the voice segmentation unit 103 classifies each frame into a voice frame or a non-voice frame by comparing a threshold value (hereinafter, it is referred to as threshold θ) and a voice segmentation feature value for each frame. If a calculated voice feature value for a frame is larger than the threshold value θ, the frame is classified as a voice frame. If a calculated voice feature value is less than the threshold value θ, the frame is classified as a non-voice frame. Then, the voice segmentation unit 103 merges connected voice frames classified above into a voice section (hereinafter, referred to as a first

> voice section). As a voice segmentation feature value, amplitude power for each frame can be used, for example. However, the voice segmentation feature value which indicates possibility of being voice is not limited to amplitude power.

(Arakawa ¶0043.)

199. Arakawa next teaches calculating a feature value for voice recognition for each audio frame:

> The voice recognition feature value calculating unit 105 calculates a feature value used for voice recognition (hereinafter, it is described as a voice recognition feature value) for each frame input sound data. As a voice recognition feature value, cepstrum feature or its derivative feature can be used, value for example.

(Arakawa ¶0044.)

200. Arakawa further teaches using the feature value for each frame to identify words and/or phonemes for speech recognition, as well as to update the threshold for distinguishing voice from non-voice:

> The searching unit 108 calculates, based on the voice recognition feature value, a likelihood of voice and a likelihood of non-voice for each frame, and searches for a word sequence using these likelihoods and the above-mentioned models. The searching unit 108 may search for a maximum-likelihood word sequence among calculated the likelihoods of voice, for example.

> Also, the searching unit 108 segments a section to be the target of the voice recognition (hereinafter, it is referred to as a second voice section) based on the likelihood of voice and the likelihood of non-voice that have been calculated. Specifically, the searching unit 108 segments a section during which the likelihood of voice that has been calculated based on the voice recognition feature value is higher than the likelihood of non-voice that has been calculated based on a voice recognition feature value as the second voice section.
>
> Thus, the searching unit 108 obtains the word sequence corresponding to the input sound (a recognition result) using the feature value for each frame, the vocabulary/phoneme model and the non-voice model, and, in addition to that, obtains the second voice section….
>
> According to a difference between length of the first voice section and length of the second voice section, the parameter updating unit 109 updates the threshold value θ. That is, the parameter updating unit 109 compares the first voice section and the second voice section, and updates the threshold value θ to be used by the voice segmentation unit 103.

(Arakawa ¶¶0047–0050.)

201. Arakawa thus teaches at least three segmentations of the audio files: the first segmentation of voice versus non-voice sections, the second segmentation of voice segments based on phoneme recognition, and the frame-by-frame

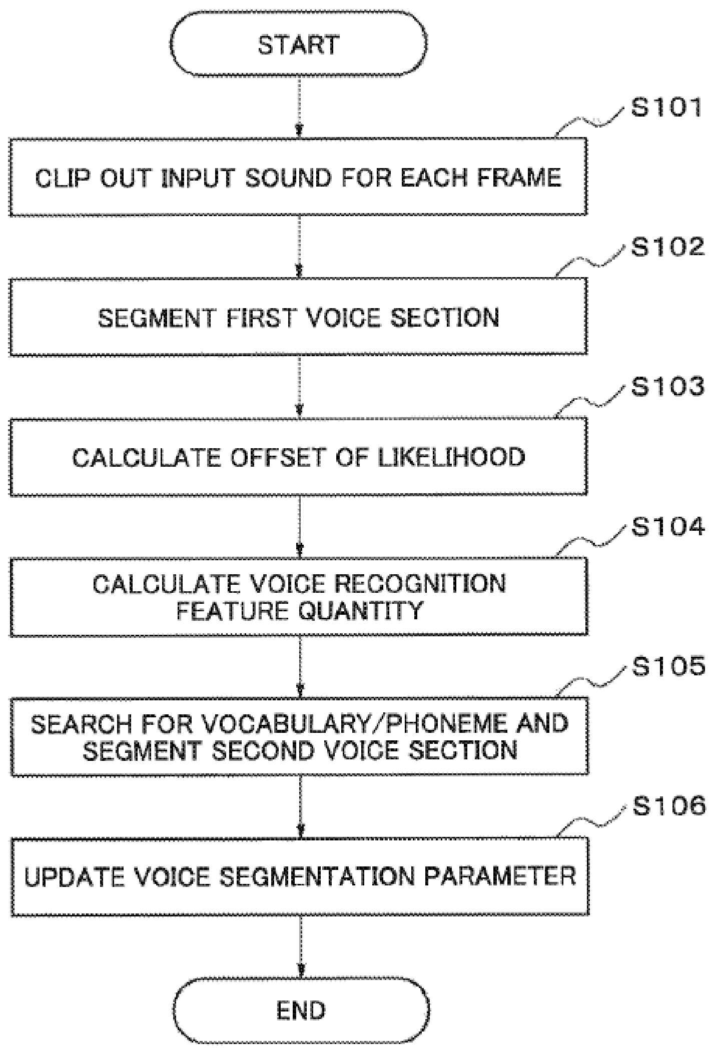
segmentation for purposes of voice segmentation and voice recognition.

202. It would have been obvious to a POSA to combine Arakawa's method of performing voice recognition using segmentation with Fontana in order to take advantage of Arakawa's benefits of improving speech recognition in the presence of background noise. (*See, e.g.*, Arakawa Abstract, ¶0001 ("The present invention relates to a voice recognition system, a voice recognition method and a voice recognition program which recognize voices in an environment where background noise exists."), ¶0012 ("When voice recognition is performed, there is a case where the background noise, line noise and sudden noise such as a sound of hitting a microphone and the like exists. In such case, by using the voice recognition apparatus disclosed in patent document 1 and patent document 2, an error of voice recognition can be suppressed.").)

203. Furthermore, a POSA combining Fontana with Lau would have found it obvious to segment the audio files by frame and into voice sections according to Arakawa's "voice segmentation unit" **before processing the audio files in parallel**. Arakawa explains that dividing the audio into frames, and segmenting voice versus non-voice sections using the voice segmentation unit, each occur prior to performing the speech recognition functions described in Arakawa. This is illustrated in Figure 2, which shows that "clip out input sound for each frame" and "segment first voice

section" each occur before the speech recognition functions (e.g., "search for vocabulary/phoneme"):

Fig.2



(*See* Arakawa Fig. 2.)

204. A POSA would have found it obvious that Fontana's distributed computing system could separate the resulting audio segments across multiple servers to process in parallel for speech recognition—either by having each frame processed in parallel by the "voice recognition feature value calculating unit" and "searching unit" of Arakawa, or by having each voice versus non-voice section of Arakawa processed in parallel. Indeed, Fontana expressly discloses that processes such as audio processing could be separated and performed in parallel in this fashion:

> In certain embodiments, the servers 302a-c are specifically designed according to the application the network 300 is intended to support; for example in the case where multimedia data is to be processed using the computing capabilities within network 300, one or more of the servers 302a-c can include specific graphical processing units for processing lower level video, image or audio algorithms. Other specific capabilities can be included into the servers 302a-c as well. The servers 302a-c are configured to share processing jobs, such that tasks can be performed by one or more of the computing systems, or separated and performed across multiple computing systems in parallel.

(Fontana ¶0053.) I also discuss this for Ground 1 Claim 8.

205. **<u>Rationale and motivation to combine (Fontana and Lau with Arakawa)</u>:** A POSA would have been motivated to combine Fontana and Lau with Arakawa to take advantage of Arakawa's benefits of improving speech recognition

in the presence of background noise, as I discuss above. Additionally, Fontana explains that "a plurality of different speech to text algorithms can be applied." (Fontana ¶0093.) As such, a POSA would have been motivated to identify audio processing algorithms, such as that disclosed by Arakawa, that provided additional benefits. Arakawa is an analogous reference to the '972 patent. Arakawa, like the '972 patent, addresses audio processing using segmentation. ('972, 1:49-53 ("The text from the audio can be generated by first segmenting images of the audio, and then converting the segments of images to text in parallel. The audio can be segmented at spectrum thresholds."); Arakawa, Abstract ("Provided is a voice recognition system capable of, while suppressing negative influences from sound not to be recognized, correctly estimating utterance sections that are to be recognized. A voice segmenting means calculates voice feature values, and segments voice sections or non-voice sections by comparing the voice feature values with a threshold value.").) A POSA would have had a reasonable expectation of success in deploying Arakawa's audio processing algorithms within the context of Fontana's distributed computing for multimedia processing. Indeed, as noted, Fontana expressly teaches that "a plurality of different speech to text algorithms can be applied" (Fontana ¶0093) and that audio processing can be done in parallel using a plurality of servers (Fontana ¶0053).

2. Claim 9: "The method according to claim 7, wherein the audio files and the image files are segmented at spectrum thresholds."

206. A POSA would have found claim 7 obvious over the combination of Fontana with Lau, as I discuss above for Ground 1. Additionally, as discussed above for Ground 2 Claim 8, Arakawa discloses segmenting audio files **at spectrum thresholds** to distinguish voice from non-voice sections of an audio file. It would have been obvious to combine Fontana with Arakawa to **segment[] the audio files and the image files at spectrum thresholds**.

207. The general motivation to combine Fontana with Arakawa is discussed above for Ground 2 Claim 8. A POSA would additionally have been motivated to segment the image files in the same locations as the audio files based on Arakawa's thresholds for the voice segmentation unit. Specifically, Fontana explains that it may be desirable to generate thumbnails (i.e., segment the image files) based on sound changes in the multimedia content: "The thumbnail extraction module 618 is arranged to <u>generate thumbnails at possible locations</u> the content provider would like to create an object of interest (for example a first frame, a last frame, and <u>immediately following major scene or sound changes in the content</u>). In some embodiments, the thumbnail extraction module 618 <u>generates a series of thumbnails representing scenes throughout the multimedia content</u>." (Fontana ¶0095.)

208. A POSA would have found it obvious that Arakawa's voice segmentation unit identifies "sound changes" based on its use of a **spectrum threshold**. (*See* Arakawa ¶0043 ("If a calculated voice feature value for a frame is larger than the threshold value θ, the frame is classified as a voice frame. If a calculated voice feature value is less than the threshold value θ, the frame is classified as a non-voice frame. Then, the voice segmentation unit 103 merges connected voice frames classified above into a voice section (hereinafter, referred to as a first voice section).").) Thus, a POSA would have appreciated that the same segmentation into voice sections versus non-voice sections performed for audio files in Arakawa could be used for generating the thumbnail images to segment the image files in Fontana.

3. Claim 20: "The server according to claim 17, wherein the audio and the images are segmented at spectrum thresholds."

209. A POSA would have found claim 17 obvious over the combination of Fontana with Lau, as I discuss above for Ground 1. Additionally, for the same reason as discussed above for claim 9, it would have been obvious to a POSA to combine with Arakawa to **segment the audio and the images at spectrum thresholds**.

VI. NO SECONDARY CONSIDERATIONS OF NONOBVIOUSNESS

210. As explained in **Part III.B**, I understand that "secondary

considerations" may be used as indicia of nonobviousness, provided that any such consideration has a nexus to the invention claimed. I understand from Petitioner's counsel that Patent Owner in the related district court litigation has not yet identified any evidence related to secondary considerations with a nexus to the claimed invention. Nor am I aware of information, such as commercial success, unexplained results, long felt but unsolved need, industry acclaim, simultaneous invention, copying by others, skepticism by experts in the field, and failure of others, suggesting that the claims addressed in this Declaration are not obvious. To the extent Patent Owner later provides information it claims relates to secondary considerations, I reserve the right to supplement my analysis and opinions to comment on it.

VII. CONCLUSION

211. In signing this Declaration, I recognize that the Declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in this proceeding. If required, I will appear for cross-examination at the appropriate time. I reserve the right to offer opinions relevant to the invalidity of the challenged claims at issue and/or offer testimony in support of this Declaration.

212. I hereby declare that all statements made herein of my own knowledge are true and that all statements are made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001.

Dated: May 9, 2025

Respectfully submitted,

Henry H. Houh.

Henry Houh, Ph.D.
Lexington, MA

APPENDIX A

Henry H. Houh

15 Ballard Ter.
Lexington, MA 02420
henry.houh@gmail.com
cell: 781-325-8273

Education

Massachusetts Institute of Technology, Cambridge, MA

- PhD in Electrical Engineering and Computer Science, February 1998. "Designing Networks for Tomorrow's Traffic," thesis supervised by Professor David Tennenhouse and Professor John Guttag. GPA 4.7/5.0
- Master of Science in Electrical Engineering and Computer Science, February 1991. "Demonstration of a laser repetition rate multiplier," thesis. GPA 4.5/5.0
- Bachelor of Science in Electrical Engineering and Computer Science, June 1989. "Boundary element analysis of arbitrarily shaped dielectric structures," thesis. GPA 4.7/5.0
- Bachelor of Science in Physics, February 1990. GPA 4.7/5.0

MITx, Cambridge, MA

- September 2020 – present: Enrolled in Statistics and Data Science MicroMasters program, a sequence of five MIT-offered graduate-level courses in statistics and machine learning, culminating in a MicroMasters certificate. Have completed four of five required courses.

Experience

Houh Consulting Inc. / Independent Consultant

- 2009 - present: Technical consultant specializing in Social Networking, Web 2.0, Web Site Development, Data Networking, Optical Networking, Telecommunications, Media Streaming and Voice Over IP. Clients include: Akin Gump Strauss Hauer & Feld LLP, Arnold & Porter Kaye Scholer LLP, BBN, Baker Botts LLP, Barnes & Thornburg LLP, BBN Technologies, Beus Gilbert McGroder PLLC, Bracewell LLP, Christensen O'Connor Johnson Kindness PLLC, Cooley LLP, Covington & Burling LLP, Cozen O'Connor P.C., DLA Piper, Davis Polk & Wardwell LLP, Desmarais LLP, Duane Morris LLP, Erise IP, P.A., Faegre Baker Daniels LLP, Finnegan, Henderson, Farabow, Garrett & Dunner, LLP, Fish & Richardson P.C., Goldman Ismail Tomaselli Brennan & Baum LLP, Haynes and Boone, LLP, Kaye Scholer, Kellogg Huber Hanson Todd Evans & Figel PLLC, Kirkland & Ellis LLP, Klarquist Sparkman, LLP, McCarthy Tetrault LLP, McGuireWoods LLP, Morgan, Lewis & Bockius LLP, Munger Tolles & Olson LLP, Orrick, Herrington & Sutfliffe LLP, O'Melveny & Myers LLP, Orrick, Herrington & Sutcliffe LLP, Paul Hastings LLP, Paul, Weiss, Rifkind, Wharton & Garrison LLP, Perkins Coie LLP, Quinn Emanuel Urquhart & Sullivan, LLP, Sidley Austin LLP, Smith Baluch LLP, Sterne, Kessler, Goldstein & Fox PLLC, Troutman Sanders LLP, Venable LLP, Weil, Gotshal &

Manges LLP, Wiley Rein LLP, Wilmer Cutler Pickering Hale and Dorr LLP, and Winston & Strawn LLP.

BlocksCAD Inc. (H3XL spin-out)

- 2014 - present: Founder and member of Board of Directors. BlocksCAD is a visual Computer-Aided Design (CAD) program, designed to teach children mathematics and geometric visualization. BlocksCAD is being used in U.S. schools and worldwide in maker spaces, with 350,000+ users. BlocksCAD was selected to be part of the LearnLaunch 2017 Boost Accelerator, the MIT Play Labs 2018 Summer Accelerator, and the MassChallenge 2018 Boston Cohort. Recipient of DARPA DSO grant in 2014, and SBIR grant from the USDA in 2019.

H3XL Inc. d/b/a Einstein's Workshop (formerly Lexington Robotics)

- 2009 - 2022: Founder and President. Started a science, technology, engineering and math enrichment program and creative/maker space, in 7,000 square feet of space. Serve 2,000+ kids and families annually. As of end of 2022, have delivered an estimated 250,000 student-instruction-hours of STEM courses. Principle Investigator for 2-year DARPA grant to improve 3D Computer Aided Design tool developed in-house for the purposes of teaching 8 year olds and up 3D CAD.

Eons

- 2008 - 2009: Chief Technology Officer. Created product that Eons acquired from BBN Technologies. Integrated BBN product with Eons social networking platform and significantly increased the Eons group creation rate. Helped evaluate advertisement platform offerings and rolled out the "Boom Network" advertising network. Eons raised $32 million from General Catalyst Partners, Charles River Ventures, Sequoia Capital, and Intel Capital.

BBN Technologies

- 2007 - 2008: Delta Division, Vice President of Technology. Grew "Boomerang" counter-sniper project engineering team and significantly de-risked $10 million worth of product deliveries. Identified sales team leading to $10 million dollars of new sales within 6 months and $100+ million in additional orders in the following two years. Boomerang was a significant asset leading to the acquisition of BBN by Raytheon in 2009. Created new business plan and grew team; launched new fully-featured social networking web site in 5 months. Served as lead expert witness in patent infringement lawsuit, resulting in $58 million jury award to client; verdict for patents I testified on were upheld on appeal which resulted in a $120 million settlement.
- 2004 - 2007: Delta Division, Director of Technology, responsible for commercializing IP and creating new businesses. Hired and grew division's initial engineering team. Wrote three business plans, two of which were funded. For call center business plan, acted as general manager, hiring and managing engineering team, inside sales team, and identifying and recruiting a new general manager. Identified and recruited other key employees to Delta Division, including senior members of team leading to successful internal sales growth and spin-outs of projects. Contracted by BBN to BBN spin-out PodZinger as VP of Operations and Technology. Additional areas studied include mesh networking, network security, satellite communications, and semantic web.

PodZinger Inc. (BBN spin-out, also known as EveryZing and now RAMP)

- 2006: Vice President of Operations and Technology. Significantly upgraded capability of consumer-facing search site and redeployed web site from company to co-location facility. Identified key portions of infrastructure for upgrading and cost reduction. Helped write business plan, evaluating advertisement models of revenue. Hired in operations replacement and phased back to BBN.

Commonwealth Capital

- 2004: Entrepreneur-in-residence (informally), performed technical due diligence on business plans, brainstormed ideas for new businesses with venture partner. With venture partner, left for portfolio company BBN to form core of commercialization team.

Empirix, Inc./Teradyne, Inc.

- 2001 - 2004: Chief Technologist, Engineering Manager, Web Application Test Group. Researched potential new product areas; developed product plan and prototype. Responsible for three new and existing products. Managed off-shore development team. Chief architect for all web testing products. Re-architected core testing product, helped write javascript interpreter. Provided technical vision for core product.
- 2000 - 2001: Chief Technologist, Communication Infrastructure Test Group. Responsible for incorporating new technology internally, tracking new technologies, technical evaluation of partnerships and potential acquisitions. Helped develop division strategy. Developed plans which formed core capabilities for successful new products introduced in 2004-5.
- 2000 - 2001: Engineering Manager, Communications Infrastructure Test Group. Execution of new product plan developed in prior role. Grew team from four existing engineers to team of over 30 on immediate team and over 40 on project. Delivered new platform in one year. Platform and derivatives accounted for large portion of booked products for the division within 2 years and is currently (2008) a key portion of new product offerings.
- 2000: Empirix was formed as a spin-out of Teradyne in January 2000. Reported to CEO in carve-out of Empirix from Teradyne.
- 1999: (Teradyne) Director of Business Development, Software Test Units. Reported directly to Chairman of the Board/Founder and then to general manager of software test unit (6 divisions of Teradyne). Evaluated and researched acquisition and partnership candidates. Internally assessed technology position in market and gaps in product lines. Worked with senior division staff to develop new product strategies. Attended internal Teradyne divisional board meetings. Chairman served as my mentor.

3Com Corporation/NBX Corporation

- 1999: Software Engineer 5. Continued work after 3Com acquisition of NBX. Built cross-division relationships for new products and research directions. NBX was acquired by 3Com in March 1999.
- 1997 - 1999: (NBX) Senior Scientist and Engineer. Work in IP Telephony. Architected next-generation product. Protocol design and validation for core protocol now used tens of millions of times daily. Led team in integration of IP protocols into current product. Designed audio reconstruction algorithms. Developed applications for bug analysis and diagnosis of system

problems. Implementation of network simulator. Work on collaborative projects with external partners. Worked to identify gaps in product. Representative at numerous trade shows. Innovated novel methods of using product.

MIT Department of Electrical Engineering and Computer Science, Cambridge, MA

- 1991 - 1998: Research assistant, Telemedia Network Systems Group. Design, development, and implementation of Gigabit ATM network for distributed multimedia system. Studied host interface design issues. ATM network simulation.
- Spring 1989, Fall 1990, Spring 1995: Teaching assistant, Computation Structures digital systems course. (Spring 1995 Head TA)
- 1988 - 1989: Head laboratory teaching assistant for Computation Structures. Responsibilities included writing and revising lab assignments, and maintaining the lab.
- 1987: Laboratory teaching assistant for Computation Structures.
- 1987: Design, construction, and programming of 16-bit computer.

Agora Technology Group, Incorporated

- 1994 - 1996: Founder and CEO. Conceived and oversaw development of targeted advertising-supported Web sites. Responsible for company's vision and direction. Attended the first two WWW Conferences, presenting a workshop and paper at the first, and appearing on the "Commercialization and Economics of the Web" panel and chairing the "Where Commercial Services and the Web Are Headed" panel at the second. Sold company intact; is currently an operating stand-alone company.

AT&T Bell Laboratories, Holmdel, NJ

- 1989 - 1990: Implementation of cascadable all-optical fiber logic gate. Modelocking of all-fiber erbium laser. Construction of modelocked laser repetition rate booster. Strong optics laboratory and fiber optic experience.
- Summer 1988, 1987: Research in integrated optics. Analysis of rectangular waveguides using microwave modeling. Fabrication of integrated optical components.

Honors

- MIT Alumni Association Great Dome Award, 2010, Baker 60th Anniversary Reunion Co-Chair (highest group award given by MIT Alumni Association)
- MIT Alumni Association Presidential Citation Award (now known as Great Dome), 2008, Member of MIT Chairman's Salon committee
- MIT Alumni Association Bronze Beaver Award, 2007 (highest individual award given by MIT Alumni Association)
- MIT Alumni Association Volunteer Honor Roll, February 2004
- MIT Alumni Association Lobdell Award, 1999
- Boston Museum of Science Gold Pin for 1000 hours of Volunteer Service, April 1999
- MIT Alumni Association Presidential Citation Award (now known as Great Dome), 1997, Member of Alumni Online Communications Committee

Patents and Patent Publications

- U.S. Patent #9,697,231, H. Houh, J. Stern, R. Spina, M. Meteer, “Methods and apparatus for providing virtual media channels based on media search,” July 4, 2017.
- U.S. Patent #9,697,230, H. Houh, J. Stern, “Methods and apparatus for dynamic presentation of advertising, factual, and informational content using enhanced metadata in search-driven media applications,” July 4, 2017. See also WO2007056485.
- U.S. Patent #7,975,296, L. Apfelbaum, H. Houh, T. Mayberry and G. Friedman, “Automated security threat testing of web pages,” July 5, 2011. See also US20030159063, WO2003067405.
- U.S. Patent #7,877,736, H. Houh and J. N. Stern, “Computer language interpretation and optimization for server testing,” January 25, 2011. See also US20050138104, WO2005043300.
- U.S. Patent #7,801,910, H. Houh and J. N. Stern, “Method and apparatus for timed tagging of media content,” September 21, 2010. See also US20070112837, US20090222442, WO2007056535.
- U.S. Patent #7,590,542, D. C. Williams, W. C. Hand, H. Houh, A. R. Seeley, “Method of Generating Test Scripts Using a Voice-Capable Markup Language,” September 15, 2009. See also EP1530869, US20030212561, WO2003096663.
- U.S. Patent #6,967,963, H. H. Houh, P. Anderson, C. Gadda, “Telecommunication method for ensuring on-time delivery of packets containing time-sensitive data,” November 22, 2005. See also EP1060400, WO2000033092, CA2318774.
- U.S. Patent #5,144,375, M. C. Gabriel, H. H. Houh, N. A. Whitaker, “Sagnac Optical Logic Gate,” September 1, 1992. Also issued as European Patent # EP0456422, July 23, 1997, German Patent #DE69126913, August 28, 1997
- U.S. Patent Application Publication No. 20020015387, “Voice Traffic Packet Capture and Analysis Tool for a Data Network” (Abandoned in 2007)
- U.S. Patent Application Publication No. 20020016708, “Method and Apparatus for Utilizing a Network Processor as Part of a Test System” (Abandoned in 2007)
- U.S. Patent Application Publication No. 20020016937, “Method and Apparatus for Utilizing a Network Processor as Part of a Test System.” (Abandoned in 2007) See also WO2002011413.
- U.S. Patent Application Publication No. 20070106646, “User-directed navigation of multimedia search results” (Abandoned in 2009)
- U.S. Patent Application Publication No. 20070106660, “Method and apparatus for using confidence scores of enhanced metadata in search-driven media applications” (Abandoned in 2009)
- U.S. Patent Application Publication No. 20070106685, “Method and apparatus for updating speech recognition databases and reindexing audio and video content using the same” (Abandoned in 2015)
- U.S. Patent Application Publication No. 20070118873, “Methods and apparatus for merging media content” (Abandoned in 2015)
- U.S. Patent Application Publication No. 20090222442, “User-directed navigation of multimedia search results” (Abandoned in 2016)
- U.S. Patent Application 11/395,732, “Search snippet creation for audio and video data” (Abandoned in 2009)
- U.S. Patent Application 11/774,931, “Methods and apparatus for managing a social networking web site”
- U.S. Patent Application 11/774,947, “Methods and apparatus for organizing media files”
- U.S. Patent Application 11/774,956, “Methods and apparatus for managing an online event”

- U.S. Provisional Patent Application 61/086,909, "Measuring and ranking relationship activity"
- U.S. Provisional Patent Application 61/086,914, "Detecting media object commonality"
- U.S. Provisional Patent Application 61/086,904, "Message categorization based on message characteristics"
- U.S. Provisional Patent Application 61/086,905, "Photo tagging to request action"

Trials, Testimony and Depositions

- Case No. 1:06CV682 (CMH/BRP), Verizon vs. Vonage, U.S. District Court for the Eastern District of Virginia, filed expert report, was deposed and testified at trial.
- Case No. 1:08CV157 (CMH/TRJ), Verizon vs. Cox, U.S. District Court for the Eastern District of Virginia, was deposed and testified at trial.
- Case No. 5:09-cv-476, Two-Way Media vs. AT&T, U.S. District Court for the Western District of Texas, filed expert report, testified at trial.
- Case No. 2:10-cv-248 (RAJ/FBS), ActiveVideo Networks vs. Verizon, U.S. District Court for the Eastern District of Virginia, filed expert report and was deposed as an expert witness.
- Case No. 1:11-cv-00880-TSE-JFA, Bear Creek Technologies, Inc. vs. Verizon Services Corp., et al, U.S. District Court for the Easter District of Virginia, was deposed as an expert witness.
- Case No. 3:10-CV-298-BBC, AlmondNet, Inc. vs. Microsoft Corp., U.S. District Court for the Western District of Wisconsin, filed expert report.
- Case No. 6:10-cv-00597, Guardian Media Technologies, Ltd. Vs. AT&T Operations, Inc. et al., U.S. District Court for the Eastern District of Texas, Tyler Division, filed expert report.
- Case No. ESCV2010-02282C, The Octopus Solution LLC v. Gary Brown et al., Essex, MA Superior Court, testified at trial.
- Investigation No. 337-TA-882, In the matter of Certain digital media devices, including televisions, Blu-ray disc players, home theater systems, tablets and mobile phones, components thereof and associated software, U.S. International Trade Commission, filed expert reports, was deposed and testified at hearing.
- Investigation No. 337-TA-995, In the matter of Certain communications or computing devices, and components thereof, U.S. International Trade Commission, filed expert reports, and was deposed.
- Case No. 8:12-cv-122-LES-TDT, Prism Technologies LLC v. AT&T Mobility LLC, U.S. District Court for the District of Nebraska, filed expert report, was deposed and testified at trial.
- Case No. 11-2684 (JWL), Sprint Communications Co., L.P., v. Comcast Cable Communications, LLC, et al., U.S. District Court for the District of Kansas, filed expert report and was deposed.
- Case No. 11-2686 (JWL), Sprint Communications Co., L.P., v. Time Warner Cable, Inc., et al., LLC, et al., U.S. District Court for the District of Kansas, filed expert report, was deposed, and testified at trial.
- Case No. 12-487 (SLR), Cox Communications, Inc., et al., v. Sprint Communications Company L.P., et al., U.S. District Court for the District of Delaware, filed expert report and was deposed.
- Case IPR 2014-00039, Microsoft Corporation v. B.E. Technology LLC, U.S. Patent No. 6,628,314. Submitted declaration and cross-examined in deposition.
- Case IPR 2014-00086, Apple, Inc. v. Evolutionary Intelligence, LLC, U.S. Patent No. 7,010,536. Submitted declaration and cross-examined in deposition.
- Case No. 6:11-CV-421, Stragent, LLC v. Intel Corporation, U.S. District Court for the Eastern District of Texas, Tyler Division, deposed and testified at trial as a fact witness.

- IPR2014-01366, IPR2014-01367, IPR2014-01368, Samsung et al. v. Straight Path, submitted IPR declarations and was deposed.
- IPR2015-1006, IPR2015-1007, Cisco v. Straight Path, submitted IPR declarations.
- IPR2014-01457, IPR2014-01459, Microsoft v. Biscotti, submitted IPR declaration and was deposed.
- IPR2016-00302, Apple v. Nonend, submitted IPR declaration.
- IPR2014-00812, Yelp and Twitter v. Evolutionary Intelligence, submitted IPR declaration.
- IPR2015-00307, Cisco v. AIP, submitted IPR declaration.
- Microsoft v. Acacia, submitted IPR declaration.
- IPR2015-00342, AT&T Mobility LLC and Cellco Partnership d/b/a Verizon Wireless v. Solocron, submitted IPR declaration.
- Mediatube v. Bell Canada, Canadian Federal Court, submitted expert report and testified at trial.
- CBM2016-00036, Google v. At Home Bondholder's Liquidating Trust, submitted CBM declaration.
- IPR2016-01198, IPR2016-01201, Apple, Inc. v. VoIP-Pal.com, Inc., submitted IPR declaration and was deposed.
- Case No. 3:14-cv-2824-VC, Trend Micro Incorporated v. Rpost Holdings, Inc., et al, submitted expert report and was deposed
- Akamai v. Limelight Networks, IPR2016-01631, IPR2016-01711, IPR2017-00249, IPR2017-00349, submitted IPR declarations and was deposed
- Hyperlync Technologies v. Verizon Sourcing, No. 650151/2015 New York Supreme Court, submitted declaration and was deposed
- Cascades Streaming Technologies v. Big Ten Networks, submitted expert reports and was deposed
- Klaustech, Inc. v. Google LLC., Case No. 4:10-CV-05899-JSW (N.D. Cal.), submitted expert report and was deposed
- Cisco v. Uniloc, IPR2017-00198, IPR2017-00058, submitted IPR declarations and was deposed.
- Cisco v. Uniloc, IPR2017-00597, submitted IPR declaration and was deposed
- Twitter v. Youtoo Technologies, LLC, IPR2017-01133, IPR2017-01131, submitted IPR declarations and was deposed
- Taser International Inc. (now Axon Enterprise, Inc.) v. Digital Ally Inc., IPR2017-00375, submitted IPR declaration and was deposed
- Taser International Inc. (now Axon Enterprise, Inc.) v. Digital Ally Inc., IPR2017-00376, IPR2017-00515, IPR2017-00775, submitted IPR declarations
- HTC Corporation v. IPA Technologies, IPR2018-00306, IPR2018-00307, submitted IPR declaration
- Hulu v. Soundview, IPR2018-00864, IPR2018-01023, submitted IPR declarations and was deposed
- XpertUniverse v. Cisco, 3:17-cv-3848, U.S. District Court for Northern District of California, submitted declaration in support of claim construction and was deposed
- Improved Search v. Microsoft, C.A. No. 16-cv-650-JFB-SRF (Del.), submitted declaration in support of claim construction
- Apple v. Uniloc, IPR2018-00361, IPR2018-00394, IPR2018-00395, IPR2019-00700, IPR2019-00701, IPR2019-01667, submitted IPR declarations
- LG Electronics v. Uniloc, submitted IPR declaration
- SRC Labs and Saint Regis Mohawk Tribe v. Microsoft, 2:18-cv-00321-JLR, U.S. District Court for Western District of Washington at Seattle, submitted declaration in support of claim construction and was deposed

- Verizon Wireless v. Barkan Wireless IP Holdings LP, submitted IPR declarations
- Verizon Wireless and Samsung Electronics America Inc. v. Barkan Wireless IP Holdings LP, IPR2019-199, IPR2019-200, IPR2019-234, IPR2019-631, IPR2019-632, submitted IPR declarations
- Samsung Electronics America Inc. v. Barkan Wireless IP Holdings LP, IPR2019-100, submitted IPR declaration
- Cisco Systems Inc. v. Meetrix IP LLC, IPR2019-539, IPR2019-540, IPR2019-541, IPR2019-542, IPR2019-543, IPR2019-544, submitted IPR declarations
- GMG Products LLC v. Traeger Pellet Grills LLC, PGR2019-24, PGR2019-34, PGR2019-35, PGR2019-36, submitted PGR declaration and was deposed
- Uniloc 2017 LLC v. Google LLC, 2:18-cv-00502-JRG-RSP, U.S. District Court for the Eastern District of Texas, Marshall Division, submitted declaration in support of claim construction
- Microsoft Corporation v. Uniloc 2017 LLC, IPR2020-00102, IPR2020-00103, submitted IPR declarations
- Microsoft Corporation and HP Inc. v. Synkloud Technologies, LLC, IPR2020-00316, IPR2020-01031, IPR2020-01032, IPR2020-01269, IPR2020-01270, IPR2020-01271, submitted IPR declarations
- Seven Networks, LLC v. Apple Inc, 2:19-cv-115-JRG, U.S. District Court for the Eastern District of Texas, Marshall Division, submitted declaration in support of claim construction
- Unified Patents Inc. v. Ortiz & Associates Consulting, LLC, IPR2019-00743, submitted IPR declaration
- Uniloc 2017 LLC v. Apple Inc., 3:19-cv-01905-JD, U.S. District Court for the Northern District of California, San Francisco Division, submitted declaration in support of claim construction and was deposed
- ClearPlay Inc. v. Dish Network LLC et al.; 2:14-cv-00191, U.S. District Court for the District of Utah, Central Division, submitted expert report and re-examination declarations, was deposed, and testified at trial
- AMC Networks, Inc. et al v. Sound View Innovations, LLC, IPR2020-00482, submitted IPR declaration
- Case No. 1:17-cv-01734-RGA, Sprint Communications Co., L.P., v. Charter Communications, Inc., et al., U.S. District Court for the District of Delaware, filed expert report and was deposed.
- CommScope Technologies LLC et al. v. Barkan Wireless IP Holdings LP et al., IPR2020-00827, IPR2020-00829, IPR2020-00831, IPR2020-00833, IPR2020-00835, IPR2020-00838, submitted IPR declarations
- Cisco Systems Inc. v. Monarch Networking Solutions LLC, IPR2020-01226, submitted IPR declaration
- CoolTV Network.com v. Blackboard, Inc., Facebook, Inc., International Business Machines Corporation, Kaltura, Inc., Limelight Networks, Inc., Microsoft Corporation, Ooyala, Inc., Snap, Inc., and Trapelo Corp., submitted declaration in support of claim construction and was deposed
- Facebook, Inc. v. Onstream Media Corporation, IPR2020-01507, IPR2020-01508, IPR2020-01525, IPR2020-01527, IPR2020-01528, submitted IPR declarations
- Hisense Co LTD. v. Maxell, LTD, IPR2020-01598, IPR2020-01600, submitted IPR declarations
- Microsoft Corp. v. O'Brien, IPR2021-00015, submitted IPR declaration and was deposed
- Sable Networks Inc. v. Cisco Systems, Inc., WA:20-CV-00288-ADA, U.S. District Court for the Western District of Texas, Waco Division, submitted declaration in support of claim construction
- Peloton Interactive, Inc. v. Icon Health and Fitness Inc et al., IPR2021-00342, submitted IPR declaration

- Hulu, LLC v. SITO Mobile R&D IP, LLC et al., CBM2020-00028, IPR2020-00158, IPR2020-00206 , IPR2020-00219 , IPR2020-00265 , IPR2020-00298 , IPR2020-00304 , IPR2020-00308 , submitted CMB and IPR declarations and was deposed
- Twitter, Inc. and Google LLC v. B.E. Technology, LLC, IPR2021-00482, IPR2021-00483, IPR2021-00484, IPR2021-00485, submitted IPR declarations
- MG Freesites LTD v. ScorpCast, LLC, IPR2021-00511, submitted IPR declaration
- Avionte, LLC v. Sandeep Acharya, Samar Basnet, and Aqore LLC, Arbitration, Hon. Jeffrey Keyes (Ret.)
- Microsoft Corporation v. WSOU Investments, LLC d/b/a Brazos Licensing and Development, IPR2021-00751, submitted IPR declaration
- SAP America, Inc. v. Express Mobile, Inc., IPR2021-01144, IPR2021-01145, IPR2021-01146, submitted IPR declarations
- StratosAudio Inc. v. Hyundai Motor America, 6:20-cv-01125-ADA, U.S. District Court for the Western District of Texas, Waco Division, submitted claim construction declaration and was deposedPeloton Interactive, Inc. v. Ifit Health and Fitness Inc, IPR2022-00029, IPR2022-00030, submitted IPR declarations and was deposed
- Investigation No. 337-TA-1263, In the matter of Certain televisions, remote controls, and components thereof, U.S. International Trade Commission, filed expert reports, was deposed, and testified at hearing
- Alexander Stross v. Netease, Inc., et al., U.S. District Court (Central District of California), Case 2:20-cv-00861-AB(JC), filed expert reports, and was deposed
- Ralph Ledergerber v. Netease, Inc., et al., U.S. District Court (Central District of California), Case 2:20-cv-00862-AB(JC), filed expert reports and was deposed
- John Walmsley v. Netease, Inc., et al., U.S. District Court (Central District of California), Case 2:20-cv-00863-AB(JC), filed expert reports and was deposed
- Bernhard Kuhmstedt v. Netease, Inc., et al., U.S. District Court (Central District of California), Case 2:20-cv-02044-AB(JC), filed expert reports and was deposed
- Netflix, Inc. v. CA, Inc., IPR2021-01319, submitted IPR declaration and was deposed
- F45 Training Pty Ltd. v. Body Fit Training USA Inc., C.A. No. 20-1194 (LPS) (D. Del.), filed expert report
- Apple v. Fintiv, IPR2022-00976, IPR2022-01149, IPR2022-01150, IPR2023-00399, submitted IPR declarations and was deposed
- Estech Systems, Inc. v. Howard Midstream Energy Partners d/b/a Howard Energy Partners, C.A. No. 6:20-cv-777-ADA (Western District of Texas, Waco Division), filed expert reports and was deposed
- Echelon Fitness Multimedia, LLC. v. Peloton Interactive, Inc., IPR2020-01187, submitted IPR declaration and was deposed
- VIZIO. V. Maxell Ltd., IPR2022-01458, submitted IPR declaration
- Maxell, Ltd. et el. v. VIZIO, Inc., Case No. 2-21-cv-06758 (C.D. Cal.). filed declaration and was deposed
- Microsoft Corporation v. SurfCast, Inc., submitted IPR declarations
- The Walt Disney Company et al. v. WAG Acquisition LLC, IPR2022-01227, IPR2022-01228, IPR2022-01346, submitted IPR declarations and was deposed
- Rakuten Rewards et al. v. IBM, submitted IPR declarations and was deposed
- Fubotv Media Inc. v. Sito Mobile R&D IP, LLC et al., submitted IPR declarations
- Unified Patents, LLC v. Noblewood IP LLC, IPR2022-01111, submitted IPR declaration
- Unified Patents, LLC. v. Pedersen, Peter Henrik, IPR2023-00029, submitted IPR declaration
- Ericsson Inc v. Dali Wireless Inc., submitted IPR declarations

- Universal Electronics, Inc. et al. v. Roku, Inc., IPR2022-00818, submitted declarations and was deposed
- Google LLC. v. Wildseed Mobile LLC et al., IPR2023-00244, IPR2023-00245, IPR2023-00246, IPR2023-00247, IPR2023-00248, submitted declarations
- Lululemon Athletica Canada Inc. et al. v. Nike, Inc., IPR2023-00180, IPR2023-00424, submitted declarations
- Zoho Corporation, et al. v. Meetrix IP, LLC, IPR2023-00371, IPR2023-00377, IPR2023-00379, IPR2023-00380, IPR2023-00382, submitted declarations
- WSOU Investments, LLC d/b/a/ Brazos Licensing and Development v. Google LLC, 6:20-cv-572, U.S. District Court for the Western District of Texas, Waco Division, submitted reports and was deposed
- Google LLC v. Safecast Limited LLC, IPR2023-00652, submitted declaration
- Investigation No. 337-TA-1338, In the matter of Certain smart televisions, U.S. International Trade Commission, filed reports and was deposed
- Cisco Systems, Inc. v. Orckit IP LLC, IPR2023-00714, submitted declaration
- SupplyPro, Inc. v. Sandvik Machining Solutions AB, et al., AAA Case No. 01-22-000-7897. submitted declaration
- TrackThings LLC v. NETGEAR, Inc., Case No. 22-00981-RGA-JLH, U.S. District Court for the District of Delaware, submitted declaration
- Correct Transmission, LLC v. Nokia of America Corporation, Case No. 2:22-CV-00343-JRG-RSP, U.S. District Court for the Eastern District of Texas, Marshall Division, submitted export reports, was deposed, and testified at trial
- Walt Disney Parks & Resorts US Inc. v. Agile Journeys LLC; IPR2024-00029, submitted declaration
- Aylo Freesites Ltd v. Dish Technologies LLC; IPR2024-00043, IPR2024-00044, IPR2024-00045, IPR2024-00046, IPR2024-00047, IPR2024-00048, IPR2024-00146, IPR2024-00147; submitted declarations and was deposed
- Arm Inc. et al. v. Icpillar LLC; IPR2024-00476, IPR2024-00566; Retained by Arm; submitted declarations
- Microsoft Corp et al. v. Litl LLC, IPR2024-00454, IPR2024-00455, IPR2024-00456, IPR2024-00457, IPR2024-00458; submitted declarations
- Cisco Systems Inc et al. v. InfoExpress Inc, IPR2024-00539, IPR2024-00540, IPR2024-00677, IPR2024-00678; submitted declarations and was deposed
- Vizio Inc v. Multimedia Technologies Pte Ltd; IPR2024-00722, IPR2024-00723; Retained by Vizio; submitted declarations and was deposed
- Sap America Inc v. Isix IP LLC; IPR2024-00615; Retained by SAP; submitted declarations
- Cisco Systems Inc et al. v. Portsmouth Network Corp; IPR2024-00505, IPR2024-00506; submitted declarations and was deposed
- Aylo Freesites Ltd v. Dish Technologies LLC; IPR2024-00512, IPR2024-00513, IPR2024-00514, IPR2024-00515, IPR2024-00516, IPR2024-00517, IPR2024-00518, IPR2024-00519, IPR2024-00940, IPR2024-00941; submitted declarations and was deposed
- Aylo Freesites Ltd v. WellcomeMat LLC; IPR2024-00710, IPR2024-01101; submitted declarations and was deposed
- Verizon Corporate Service Group et al. v. Headwater Partners I LLC; IPR2024-00809, IPR2024-00945; Retained by Verizon; Technology: Mobile device management; Submitted declarations
- Microchip Technology Inc v. Aptiv Technologies Ltd et al.; IPR2024-00228; Retained by Aptiv; Technology: Universal Serial Bus; submitted declaration and was deposed

- Vusion Group SA v. Hanshow Technology Co Ltd; IPR2024-00963; Retained by VusionGroup; Technology: Video analysis; submitted declaration
- Maxell, Ltd. v. TCL Electronic Holdings Ltd. et al.; Case No. 5:23-cv-00108-RWS-JBB; U.S. District Court for the Eastern District of Texas, Texarkana Division; Retained by TCL; submitted declaration and was deposed
- Cerence Operating Company v. Samsung Electronics Co., LTD. et al., Case No. 2:23-cv-00482-JRG-RSP, U.S. District Court for the Eastern District of Texas, Marshall Division, submitted export report on claim construction and was deposed
- MasterClass and FuboTV v. Dish Technologies LLC; IPR2024-00917, IPR2024-00918, IPR2024-00919; submitted declarations and was deposed
- Microsoft Corp v. Dialect LLC; IPR2025-00655, IPR2025-00656, IPR2025-00657, IPR2025-00658, IPR2025-00659; submitted declarations

Publications

- "IP switching: server driven flow classification," H. H. Houh, *Proceedings of the Washington University Workshop on Integration of IP and ATM* , November 1996.
- "Aurora at MIT," D. D. Clark, H. H. Houh, and D. L. Tennenhouse, Editors, *MIT Laboratory for Computer Science Technical Report 673*, December 1995.
- "ViewStation Applications: Implications for Network Traffic," C. J. Lindblad, D. Wetherall, W. Stasior, J. F. Adam, H. H. Houh, M. Ismert, D. Bacher, B. Phillips, and D. L. Tennenhouse, *IEEE Journal of Selected Areas in Communications*, 1995.
- "The VuNet Desk Area Network: Architecture, Implementation, and Experience," H. H. Houh, J. F. Adam, M. Ismert, C. J. Lindblad, and D. L. Tennenhouse, *IEEE Journal of Selected Areas in Communications*, 13 (4), May, 1995.
- "Reducing the Complexity of ATM Host Interfaces," H. H. Houh and D. L. Tennenhouse, Hot Interconnects II Symposium Proceedings, Stanford, August 11-12, 1994.
- "Media-intensive data communications in a `desk-area' network," J. F. Adam, H. H. Houh, M. Ismert, and D. L. Tennenhouse, IEEE Communications, August 1994.
- "ViewStation Applications: Intelligent Video Processing Over A Broadband Local Area Network," C. J. Lindblad, D. J. Wetherall, W. Stasior, B. Phillips, D. Bacher, J. Adam, H. Houh, M. Ismert, and D. L. Tennenhouse, Proceedings of the 1994 USENIX Symposium on High-Speed Networking, Oakland, CA, August 1994.
- "The Media Gateway: Live Video on the World Wide Web," H. H. Houh, C. J. Lindblad, J. Soo, D. L. Tennenhouse, and D. J. Wetherall, Workshop at the 1994 World Wide Web Conference, Geneva, Switzerland, May 1994.
- "Active Pages: Intelligent Nodes on the World Wide Web ," H. H. Houh, C. J. Lindblad, and D. J. Wetherall, Proceedings of the 1994 World Wide Web Conference, Geneva, Switzerland, May 1994.
- "Wavelength Division vs. Code Division Access Methods for Optical Networks," H. H. Houh, Area Exam Paper, May 1993.
- "Experience with the VuNet: A Network Architecture for a Distributed Multimedia System," J. F. Adam, H. H. Houh, D. L. Tennenhouse, Proceedings of the 18th Conference on Local Computer Networks, pp. 70-76, September 1993
- "The VudBoard: A Simple DMA Interface," H. H. Houh, Proceedings of the 4th Gigabit Minijam, January 1994.

- "A Software-Oriented Approach to the Design of Media Processing Environments," D. L. Tennenhouse, J. Adam, D. Carver, H. Houh, M. Ismert, C. Lindblad, W. Stasior, D. Weatherall, D. Bacher, and T. Chang., submitted to the International Conference on Multimedia Computing and Systems, May 1994.
- "A Network Architecture for Distributed Multimedia Systems," J. F. Adam, H. H. Houh, M. Ismert, and D. L. Tennenhouse, submitted to the International Conference on Multimedia Computing and Systems, May 1994.
- "The Viewstation Collected Papers," D. L. Tennenhouse, J. Adam, C. Compton, A. Duda, D. Gifford, H. Houh, M. Ismert, C. Lindblad, W. Stasior, R. Weiss, D. Wetherall, D. Bacher, D. Carver, and T. Chang, MIT Laboratory for Computer Science Technical Report, MIT/LCS/TR-590, November 1993.
- "The Viewstation Collected Papers II," W. F. Stasior, D. L. Tennenhouse, J. F. Adam, D. R. Bacher, H. H. Houh, M. Ismert, C. J. Lindblad, B. M. Phillips, D. J. Wetherall, MIT Laboratory for Computer Science Technical Report, MIT/LCS/TR-696, April 1996.
- "A System's Perspective of the Sagnac Fiber Logic Gates and Their Possible Applications," A. Huang, N. Whitaker, C. Gabriel, H. Avramopoulos, P. M. W. French, H. H. Houh, and I. Chuang, Applied Optics, September 10, 1994
- "Complete Switching in a Three-Terminal Sagnac Switch," H. Avramopoulos, P. M. W. French, M. C. Gabriel, H. H. Houh, N. A. Whitaker, T. Morse, IEEE Phot. Tech. Lett. **3** (3), 235
- "Complete Switching in a Three-Terminal Sagnac Switch," H. Avramopoulos, P. M. W. French, M. C. Gabriel, H. H. Houh, N. A. Whitaker, IEEE/LEOS Annual Meeting, Paper PDP-13, November 1990
- "All-optical phase-locked oscillator," N. A.Whitaker, Jr., H. H. Houh, H. Avramopoulos, T. F. Morse, IEEE/LEOS Annual Meeting, Paper ELT2.4/MOO3, November 1990
- "Passive modelocking of an all-fiber erbium laser," H. Avramopoulos, H. H. Houh, N. A. Whitaker, M. C. Gabriel, T. F. Morse, IEEE/LEOS Conference on Optical Amplifiers and their Applications, Paper PDP-8, August 1990
- "Transverse modes, waveguide dispersion, and 30ps recovery in submicron GaAs/AlAs microresonators," J. L. Jewell, S. L. McCall, A. Scherer, H. H. Houh, N. A. Whitaker, A. C. Gossard, and J. H. English, Appl. Phys. Lett. **55** (1), July 3, 1989

Leadership, Activities and Interests

- Leadership
  - Reading Symphony Orchestra
    - Board of Directors (Treasurer), 2022 – present
  - Discovery Museum (Acton, MA)
    - Board of Directors, 2023 – present
    - Board of Directors, Development Committee, 2023 – present
    - Science and Technology Advisory Council, 2012 – present
    - The Discover Museum received the 2024 National Medal for Museum and Library Service, the nation's highest honor given to museums and libraries that make significant and exceptional contributions to their communities.
  - Tau Beta Pi National Engineering Honor Society
    - Executive Councilor and Secretary, 2024 – 2026
    - Advisor, MA B Chapter at MIT, 2003 – present
    - District Director (National Officer), Tau Beta Pi, New England Area, 1991 – 2003

    - President, MA B Chapter at MIT, Fall 1988 – Spring 1989
    - Laureate award, 1989
- MIT Alumni Association Board of Directors
    - MIT Annual Giving Board, 3 year term, 2024 – 2027
    - National Selection Committee, 3 year term, 2020 – 2023
    - K-12 STEM Initiatives Co-chair, 2013 – 2017
    - Awards Committee Chair, 2012 – 2014
    - Awards Committee, 3 year term, 2011 – 2014
    - Vice President, 2 year term, 2004 – 2006
    - Alumni Ad-Hoc Committee, Chair, 2005-2006
    - Board Member, 2 year term, 1997 – 1999
- MIT Club of Boston
    - Board of Directors, 2006 – 2011
    - K-12 Initiatives Chair, 2009 – 2012
    - VP of Communications, MIT Club of Boston, 2003 – 2006
    - Past-President, MIT Club of Boston, 2002 – 2003
    - President, MIT Club of Boston, 2001 – 2002
    - President-Elect, MIT Club of Boston, 2000 – 2001
    - VP of Programs, MIT Club of Boston, 1999 – 2000
    - Activities Super-Chair, MIT Club of Boston, 1998 – 1999
- MIT Enterprise Forum of Cambridge, Inc.
    - Past Chair, 2009 – 2011
    - In-NOW-vation Co-chair, 2010
    - Chair, 2007 – 2009
    - Vice Chair, 2005 – 2007
    - Executive Board Member, 2002 – 2011
    - Winter Workshop Co-Chair, February 2007 – conceived idea for conference, which sold-out and produced largest attendance numbers in recent memory
    - Spring Workshop Co-Chair, Spring 2004
    - Membership Committee Chair, Fall 2003 - 2006
    - 25th Anniversary Dinner Chair, Fall 2003
    - As Membership Chair and Board Member, started Special Interest Groups in 2004; a SIG won the MIT Presidential Citation award, the MIT Alumni Association's highest award for organizations, in 2006
- Estabrook Elementary School PTA
    - Advisory committee to the superintendent on PCB issue, 2010-2011
    - 4th Grade after-school science program co-organizer, 2010-2012
    - 4th and 5th Grade before-school Math Olympiad co-organizer, 2009-2013
    - 5th Grade BBQ and Yearbook Committee, 2011, 2013
    - Family Math Night volunteer, 2008-2012
- MIT Class of 1989
    - 35-year Reunion Gift Committee, Co-chair, 2024
    - Secretary, five consecutive 5 year terms, 1989 - 2014
    - 30-year Reunion Committee, 2019
    - 25-year Reunion Committee, 2014
    - 20-year Reunion Committee and Gift Committee, 2009
    - 15-year Reunion Committee and Gift Committee, 2004
    - 10-year Reunion Committee and Gift Committee, 1999

        - 5-year Reunion Committee, 1994
        - Interim Treasurer, 1993 - 1994
        - Instituted annual senior class career fair, now raising over $100,000 annually for senior class activities, Fall 1988
    - Strong, consistent record of leadership dating to high school
- Acting
    - ‘21’ (Sony Pictures), credited as “Chinatown Dealer,” 2007, Kevin Spacey's movie about the MIT Blackjack Team inspired by “Bringing Down the House” by Ben Mezrich, opened nationwide on March 28, 2008. 21 was the number one movie in US for two weeks and number one globally for one week. 21 also topped the DVD sales, Blu-ray sales and DVD rental charts.
    - Spring Lake Theater Company, first New York-area off-broadway production of “A Chorus Line,” played role of Mark, Summer 1990
- Former member of the MIT Blackjack team
- Producer for 10,000 Maniacs’ 2013 album “Music from the Motion Picture”
- Executive Producer for 10,000 Maniacs’ 2015 album “Twice Told Tales”
- Violist, violist, harpist, guitarist, singer, actor: played in many amateur/semi-professional groups including Firebird Pops Orchestra, Reading Symphony Orchestra, MIT Summer Philharmonic Orchestra, Carlisle Chamber Orchestra, Rivers Symphony Orchestra, Opera51 Orchestra, Freisinger Chamber Orchestra, Merrimack Valley Philharmonic, Longwood Symphony, MIT Symphony, and Somerville Community Chorus
- Violist, “Coming Home” album by Karen Phillips, 2014
- Google Scholar citations: 2224, h-index: 22, i10-index: 27 (since 2019: 539/12/12)
- Erdos-Bacon number of 7
- Erdos-Bacon-Sabbath number of 11