

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

SAMSUNG ELECTRONICS CO., LTD.

Petitioner

v.

ANONYMOUS MEDIA RESEARCH HOLDINGS, LLC.

Patent Owner

Case IPR2024-01348 (U.S. Patent No. 10,572,896)

Case IPR2024-01349 (U.S. Patent No. 10,719,849)

Case IPR2024-01350 (U.S. Patent No. 10,719,848)

EXPERT DECLARATION OF W. LEO HOARTY

TABLE OF CONTENTS

I.	INTRODUCTION	6
II.	BACKGROUND AND QUALIFICATIONS	6
III.	LEGAL STANDARDS	12
A.	Patentability	13
B.	Prior Art.....	17
C.	Inherency	18
D.	Person of Ordinary Skill In The Art.....	19
E.	Claim Construction.....	20
F.	Written Description	22
IV.	BASES FOR OPINIONS	23
V.	LEVEL OF ORDINARY SKILL IN THE ART	24
VI.	OPINIONS ON CLAIM CONSTRUCTION.....	25
VII.	OPINIONS ON WRITTEN DESCRIPTION SUPPORT	25
A.	The “BACKGROUND OF THE INVENTION”	26
B.	State of the Art	33
C.	The “SUMMARY”	43
D.	The “DETAILED DESCRIPTION”	47
E.	The Challenged Claims	56
VIII.	OPINIONS ON SPECIFIC ISSUES RAISED BY PETITIONER.....	61
IX.	CONCLUSION.....	80
	APPENDIX A	81

Exhibit List

Exhibit No.	Description
2001	Expert Declaration of W. Leo Hoarty
2002	<i>Curriculum Vitae</i> of W. Leo Hoarty
2003	Computer-generated Redline Between As-Filed Specification in U.S. Patent Application No. 11/139,330 (Ex. 1011) and Specification of U.S. Patent 10,572,896 (Application No. 16/403,204) (Ex. 2038)
2004	Joshua Grossnickle et al. THE HANDBOOK OF ONLINE MARKETING RESEARCH (2001)
2005	U.S. Patent No. 3,919,479
2006	U.S. Patent No. 4,230,990
2007	U.S. Patent No. 4,697,209
2008	U.S. Patent No. 4,677,466
2009	U.S. Patent No. 4,739,398
2010	U.S. Patent No. 5,019,899
2011	U.S. Patent No. 5,436,653
2012	U.S. Patent No. 5,581,658
2013	U.S. Patent No. 6,675,383
2014	U.S. Patent No. 6,633,651
2015	RESERVED
2016	U.S. Patent No. 6,990,453
2017	U.S. Patent No. 7,289,643
2018	U.S. Patent App. Pub. No. 2002/0026635

Exhibit No.	Description
2019	U.S. Patent No. 7,711,564
2020	WO 2002/065782
2021	WO 2004/051997
2022	U.S. Patent App. Pub. No. 2006/0041753
2023	U.S. Patent App. Pub. No. 2006/0075237
2024	WO 2004/040475
2025	WO 2004/030341
2026	WO 2004/044820
2027	WO 2005/006768
2028	U.S. Patent No. 7,694,318
2029	U.S. Patent No. 8,600,113
2030	U.S. Patent App. Pub. No. 2005/0193016
2031	U.S. Patent App. Pub. No. 2007/0271300
2032	U.S. Patent No. 7,738,704
2033	U.S. Patent App. Pub. No. 2006/0195860
2034	U.S. Patent No. 6,661,793
2035	U.S. Patent App. Pub. No. 2004/0189873
2036	U.S Patent No. 8,755,524
2037	U.S. Patent No. 5,912,815
2038	U.S. Patent 10,572,896

Exhibit No.	Description
2039	Computer-generated Redline Between Specification of U.S. Patent 10,572,896 (Application No. 16/403,204) (Ex. 2038) and Specification of U.S. Patent 10,719,849 (Application No. 16/660,624) (Ex. 1001 in '849 IPR)
2040	Computer-generated Redline Between Specification of U.S. Patent 10,572,896 (Application No. 16/403,204) (Ex. 2038) and Specification of U.S. Patent 10,719,848 (Application No. 16/660,615) (Ex. 1001 in '848 IPR)

I. INTRODUCTION

1. I have been retained as a technical expert by Anonymous Media Research Holdings LLC (“Patent Owner”) and asked to provide my objective and unbiased opinions concerning U.S. Patent Nos. 10,572,896 (“the ’896 patent”), 10,719,848 (“the ’848 Patent”) and 10,719,849 (the ’849 patent). Specifically, I have been asked to provide my opinions on the patentability of those patents in view of certain prior art references, the state of the art at the time of the claimed invention, and the knowledge of a person of ordinary skill in the art at the time of the claimed inventions. My opinions on these issues are provided in detail below.

II. BACKGROUND AND QUALIFICATIONS

2. In forming the opinions expressed in this Declaration, I have considered and relied upon my education, background, and experience. In addition, I have reviewed and relied upon additional materials in preparation of this Declaration, as described below.

3. I am currently the Principal of WLH Consulting LLC. *My curriculum vitae* is provided (as Exhibit 2002). I studied electronic engineering and computer science at the Ohio State University for two years.

4. I am the founder of six technology startups over the last 30 years. While creating new technology companies over the preceding 30 years, I have raised over \$600M of investment capital from the top venture capital firms as well as from top

technology companies such as IBM, Intel, Cisco, Samsung, among others. The companies that I founded have created jobs for thousands of technology workers.

5. Most of my work has focused on developing set-top technology and digital advertising support systems. Relevant to this matter, an interactive cable television technology company, called ICTV Inc., was founded by me in 1990, and renamed ActiveVideo Networks (AVN) in 2003. The company developed video-on-demand (VOD) as well as cloud-based systems providing virtual cable TV set-top services to deliver television programming as well as interactive program guides over managed (cable TV) and unmanaged (Internet) networks. My understanding is that AVN is currently providing service to 24 million cable TV set-top boxes in households serviced by Charter and TimeWarner systems throughout the U.S.

6. I founded the Dotcast company in 1999 and served as President and CTO. The company developed a set-top box and service, called MovieBeam, that received both over-the-air (OTA) and broadband (Internet) delivery of feature films stored on an internal hard drive in the set-top box. The OTA data was delivered by means of a patented process that I helped develop that embedded high-speed digital data in an analog TV signal of the then NTSC standard. The data was packaged as a sub-carrier injected orthogonally to the black and white (luminance) carrier and did not interfere with the television picture (which was the essence of the patented process). I petitioned the Federal Communications Commission (FCC) for

permission to apply the technology to existing broadcast television transmitters. After two years of field testing, the FCC granted permission and the service was then deployed in the top fifty television markets (DMAs) on a TV station in each market. Each MovieBeam set-top in the home automatically turned to the local modified TV station by means of its internal TV receiver and demodulated the data sub-carrier contained in the TV signal storing the incoming data on its internal hard drive. The set-top utilized a small indoor active TV antenna which I also designed and patented.

7. Movies were delivered slower than real-time on a 24/7 schedule (trickle-filled cache). Metadata about each movie was stored in the file header on the internal hard drive. This information included the usual fields of title, actors, director, and importantly, the rating. The set-top featured parental controls such as not playing R-rated films without a password, among other useful controls. Consumers selected programming from an internal user interface guide and purchased movies for rental under similar rules as Blockbuster. The set-top box also collected detailed media measurements including browsing activity and, of course, movie buy rates. Service was active from 2004 to 2007 in the top 50 television markets (DMAs) in the U.S. and was backed by all nine movie studios, led by the Walt Disney Company.

8. I have over twenty years of experience designing and developing speech recognition as well as automated content recognition (ACR) systems based

on both audio data sample and video data sample analysis. For example, I began research and development in speech recognition and natural language processing in the early 2000's in anticipation of use in the MovieBeam set-top though it remained only in prototype. In 2008, I consulted to Mediatronics Inc, a Los Angeles-based speech recognition consumer products development company engaged in creating the first fully speech recognition-based GPS navigation system for automotive use. Their 'VCommand' product required no keyboard nor display. The user simply spoke a navigation request to the device which then provided spoken turn-by-turn instructions as the user drove along. Although common today, the product was many years ahead of its competition.

9. In 2010 I was a founding member of the CKX Corporations' Viggle service which relied on audio ACR to identify the television program currently playing in the home of a Viggle participant. At the time, CKX was the owner of the popular reality TV program, American Idol, among other assets and was in the process of exploring expansion of its various franchises. I assembled a team that developed the Viggle app for mobile phones as well as the cloud-based service that provided the audio automated content recognition of audio segments captured by the Viggle mobile phone app. The users of the service won loyalty points redeemable for prizes based on detected TV programming.

10. I next spent several years in research and development in more advanced audio ACR techniques. I filed a non-provisional application in 2013 for a novel ACR system that was granted in 2015. I was consulting with the Vizio television manufacturer at the time and sold the patent to them. I was then contracted by the company to assemble a software development team and create the system based on the patent that would run in their smart TVs generating fingerprints from TV audio and sending the fingerprints to a cloud-based server system. I also wrote much of server-side software that receives the streaming fingerprints in real-time from the TVs and identifies content currently playing.

11. I also assisted in the parallel development of the video ACR system operating in Vizio smart TVs. My deep understanding of both the audio and video ACR technology allowed me to assist in integrating the two ACR systems in a cost-effective cloud-based system which today is operating in more than 25 million Vizio smart TVs.

12. The Vizio ACR technology is the basis of its audience measurement services that are widely used in the television industry for precision daily audience measurement, including by the venerable A.C. Nielsen Company. In other services provided by Vizio using its ACR technology, media related to currently playing programming is returned to the smart TV for display of supplemental information and for targeted advertising.

13. In 2018, I developed and patented a novel technology for video watermarking which places invisible data into the video signal for identification of the video segment. Under contract, I helped develop the software for a dynamic ad insertion system using video watermarks. The technology is currently in use by major broadcast networks under the name Project OAR. Client software runs on smart TVs that detect and decode video watermarks and then communicates over the Internet to Content Delivery Network (CDN) servers to download a specific TV ad to substitute for the existing ad based on certain combinations of zip code, demographic profiles, and time of day. Ad swapping actions occur in less than a tenth of a second. Information about the ad, including whether the ad was viewed in its entirety, or whether the person switched the channel, is passed back to the advertiser.

14. Additional technology I developed and patented for Vizio includes a TV ad attribution system using what is colloquially known as “tracking pixels.” The technology allows third-party verification of individual TVs displaying specific TV ads, a service of high value to advertisers.

15. Throughout my career I have received a variety of awards including, for example, the National Cable Television Association (NCTA)—Best New Technology 1996 for ICTV’s interactive TV and VOD service; Consumer Electronics Society (CES)—Best of Show in 2005 and 2006 for the MovieBeam set-

top and service; 2023 Emmy Award went to Vizio, Inc for their precision audience measurement service which utilizes my patented audio automated content recognition (ACR) technology for precision monitoring of household TV viewing, otherwise known as media measurement.

16. I am an inventor on 39 patents with three pending applications and my patent portfolio has been extensively cited by over a hundred third-party patents. My past work has been cited in many technology articles on cable television as well as on consumer electronics.

17. I am being compensated for services provided in this matter at a rate of \$475.00 per hour for all time spent on it, except for travel time. I am also being reimbursed for reasonable expenses, such as travel expenses. My compensation is not contingent on my opinions, on the outcome of any matter, or on any of the technical positions I explain in this declaration.

18. I have no financial interest in Patent Owner or the '896, '848 or '849 Patents.

III. LEGAL STANDARDS

19. I am not an attorney. For purposes of this Declaration, I have been informed about certain aspects of the law that are relevant to my opinions, as set forth below.

A. Patentability

20. I understand that for an invention claimed in a patent to be valid and patentable, it must be, among other things, new and not obvious in light of what was known and came before it. That which was known or came before the invention claimed is generally referred to as “prior art.”

21. I understand that there are two ways in which prior art may render a patent claim unpatentable. First, the prior art can be shown to “anticipate” the claim. Second, the prior art can show that the claim was obvious to a person of ordinary skill in the field of the invention at the time the invention was made.

22. I understand that, for a patent claim to be “anticipated” by the prior art, each and every element or requirement of the claim must be found, expressly or inherently, in a single prior art reference. Further, I understand that anticipation requires that the single prior art reference must disclose the elements of the claim arranged or combined in the same way as in the claim.

23. It is my understanding that whether a claimed invention is obvious is a question of law, with underlying factual inquiries that are determined from the perspective of a person of ordinary skill in the art at the time of the claimed invention.

24. It is my understanding that a claimed invention is obvious if the differences between the claimed invention and the prior art are such that the claimed

invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art (a “POSITA”).

25. It is my understanding that obviousness requires consideration of the following underlying factual inquiries:

- determining the scope and content of the prior art
- ascertaining the differences between the claimed invention and the prior art;
- resolving the level of ordinary skill in the art; and
- evaluating objective evidence or “secondary considerations” of non-obviousness, such as commercial success, long-felt but unresolved needs, failure of others, industry praise, and unexpected results.

26. It is my understanding that a reference must be analogous art to the claimed invention in order to be properly considered for use in an obviousness inquiry. It is my understanding that a reference is analogous art to the claimed invention if:

- the reference is from the same field of endeavor as the claimed invention (even if it addresses a different problem); or
- the reference is reasonably pertinent to the problem faced by the inventor (even if it is not in the same field of endeavor as the claimed invention).

27. It is my understanding that whether a reference is reasonably pertinent turns on how the problem faced by the inventor is reflected, either explicitly or implicitly, in the specification.

28. It is my understanding that the following are exemplary rationales that may support a conclusion of obviousness:

- Combining prior art elements according to known methods to yield predictable results;
- Simple substitution of one known element for another to obtain predictable results;
- Use of known techniques to improve a similar method or product in the same way;
- Applying a known technique to a known method or product ready for improvement to yield predictable results;
- Obvious to try – choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success;
- Known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations are predictable to one of ordinary skill in the art; and
- Some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.

29. In addition to the above, I understand that a person of ordinary skill in the art may resort to logic, judgment and common sense available to him or her at the time of the invention to find a claim feature obvious, even if the prior art relied on does not explicitly state or teach that feature. I understand that when common sense is relied on to find a feature obvious, the reasoning for that conclusion must be articulated with sufficient clarity.

30. I also understand that a patent claim is not proved obvious merely by demonstrating that each of its elements was, independently, known in the prior art. An obviousness determination requires finding both that a skilled artisan would have been motivated to combine the teachings of the prior art references to achieve the claimed invention, and that the skilled artisan would have had a reasonable expectation of success in doing so. I understand that it is important to be careful not to allow hindsight reconstruction of references to produce the claimed invention, without any explanation as to how or why the references would be combined to produce the claimed invention.

31. Regarding “objective evidence” or “secondary considerations,” I understand that for such evidence to be relevant to the obviousness of a claim, there must be a causal relationship (called a “nexus”) between the claim and the evidence. I also understand that such objective evidence may include: commercial success of products covered by the patent claims; long-felt need for the invention; failed

attempts by others to solve the problem addressed by the invention; copying of the invention by others in the field; unexpected results achieved by the invention; praise of the invention; the taking of licenses under the patent by others; expressions of surprise by experts and those skilled in the art at the making or use of the invention; and that the patentee proceeded contrary to the accepted wisdom of the prior art.

32. I understand that a person of ordinary skill is also a person of ordinary creativity. I further understand that the obviousness analysis need not seek out precise teachings directed to the specific subject matter of the challenged claim, but instead can take into account “ordinary innovation” that does no more than yield predictable results, which are inferences and creative steps that a person of ordinary skill in the art would be understood to employ.

33. I understand that sometimes it will be necessary to look to interrelated teachings of multiple patents and publications; the effects of demands known to the relevant field or present in the marketplace; and the background knowledge possessed by a person having ordinary skill in the art. I understand that all these issues may be considered to determine whether there was an apparent reason to combine the known elements in the fashion claimed by the patent at issue.

B. Prior Art

34. I understand that for purposes of this proceeding, prior art includes patents and printed publications in or outside the United States before the effective

filing date of the claimed invention. I also understand that prior art includes U.S. patents, published U.S. patent applications, and published international applications (pursuant to the Patent Cooperation Treaty) that themselves have an effective filing date before the effective filing date of the claimed invention, as long as they name another inventor.

35. I also understand that prior art may include any public disclosure, including any products in public use, on sale, or otherwise available to the public, before the effective filing date of the claimed invention. I understand that information that constitutes prior art generally must be information that was publicly disclosed before the effective filing date of the claimed invention. I understand that there may be exceptions to what is prior art but have been instructed that they are not relevant for purposes of my opinions here.

C. Inherency

36. I understand that if a prior art reference does not expressly set forth a particular claim element, the prior art may still disclose that element if it is “inherent” in its disclosure. To establish inherency, it is my understanding that evidence must make clear that the missing descriptive matter is necessarily present in the thing described in the reference. It is my understanding that inherency may not be established by probabilities or possibilities. In other words, it is my understanding

that the mere fact that a certain thing may result from a given set of circumstances is not sufficient to establish inherency.

37. I understand that under the principles of inherency, if a prior art device, in its normal and usual operation, would necessarily perform a claimed method, then the method will be considered to be anticipated by the prior art device. I also understand that when the prior art device is the same as a device described in the specification for carrying out the claimed method, it can be assumed the device will inherently perform the claimed process.

D. Person of Ordinary Skill In The Art

38. I understand that “a person of ordinary skill in the art” or POSITA is a hypothetical person who is presumed to have known the relevant art at the time of the invention. I understand that the term “person having ordinary skill in the art” or POSITA is sometimes used synonymously. Such a person would necessarily have the capability of understanding any scientific and engineering principles applicable to the pertinent art.

39. I understand that factors that may be considered in determining the level of ordinary skill in the art include, but are not limited to, the following:

- a. the educational level of the inventor;
- b. the type of problems encountered in the art;
- c. prior art solutions to those problems;

- d. the rapidity with which innovations are made;
- e. the sophistication of the technology; and
- f. the educational level of active workers in the field.

40. I understand that a person of ordinary skill in the art is not a real person, but rather is a hypothetical individual having the qualities reflected by the factors above. This hypothetical person has knowledge of all prior art in the relevant field as if it were arranged on a workshop wall and takes from each reference what it would teach to a person of ordinary skill in the art.

41. I understand that a person of ordinary skill in the art is a person of ordinary creativity, not an automaton. In many cases a person of ordinary skill will be able to fit the teachings of multiple prior art references together like pieces of a puzzle.

42. I understand that a person of ordinary skill in the art can collaborate with one or more other persons of ordinary skill in the art for one or more aspects with which the other person may have superior expertise, experience, and/or knowledge.

E. Claim Construction

43. It is my understanding that claim construction, or the interpretation of the claims of a patent, is ultimately a question of law. I have been informed that

claim terms are generally presumed to carry their ordinary and customary meaning. However, I understand that a patent applicant is his own lexicographer and is entitled to provide alternate definitions of words in common usage, but if the patent applicant uses a term that is in common usage without providing an alternate definition, then that term should generally be understood as having its common meaning.

44. I understand that a person of ordinary skill in the art is deemed to read a 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 specification and the “file history” of the patent developed during examination by the Patent Office.

45. I understand that the claims, the specification, and the file history are referred to as “intrinsic evidence,” and are to be read as a person of ordinary skill in the art at the time of the invention would have done at the time of the alleged invention.

46. In addition to the intrinsic evidence, I understand that the court or tribunal may also consider other relevant evidence, which is called “extrinsic evidence,” concerning relevant scientific principles, the meaning of technical terms, and the state of the art. Such evidence includes the testimony of expert witnesses. Although extrinsic evidence may be used to resolve ambiguities that may exist after reviewing the intrinsic evidence, I understand that extrinsic evidence cannot be used to contradict a claim construction that is clear from the intrinsic evidence.

47. I understand that some claims are independent, and that those independent claims are complete by themselves. Other claims refer to these independent claims, or other dependent claims, and are “dependent” on those claims. The dependent claims include all of the limitations of the claims on which they depend.

F. Written Description

48. I understand that a claimed invention must have sufficient written description support in the underlying specification upon which it relies, which means the specification must reasonably convey to those skilled in the art that the inventor had possession of the full scope of the claimed subject matter as of the filing date based on an objective inquiry into the four corners of the specification. The written description requirement is satisfied when the specification sets forth enough detail to allow a person of ordinary skill in the art to understand what is claimed and to recognize that the inventor invented what is claimed. I also understand that the level of detail required to satisfy the written description requirement varies depending on the nature and scope of the claims and on the complexity and predictability of the relevant technology. Factors used to evaluate the sufficiency of a disclosure include: 1) the existing knowledge in the particular field; 2) the extent and content of the prior art; 3) the maturity of the science or technology; and 4) the predictability of the aspect at issue. I further understand that the state of the knowledge existing at the

time must be considered in determining whether the written description requirement has been satisfied.

IV. BASES FOR OPINIONS

49. The opinions set forth in this Declaration are my own. In forming the opinions herein, I have considered the Petitions for *Inter Partes Review* filed against the '896 Patent (Paper No. 1 in IPR2024-01348, hereafter "the '896 Petition"), the Petition for *Inter Partes Review* filed against the '848 Patent (Paper No. 1 in IPR2024-01350, "the '848 Petition"), and the Petition for *Inter Partes Review* filed against the '849 Patent (Paper No. 1 in IPR2024-01349, "the '849 Petition") (collectively, "the Petitions"), the exhibits cited therein, and the additional materials cited in this Declaration (including the exhibits to the Patent Owner's Preliminary Responses cited herein). I also rely on my education, training, and experience.

50. My analysis of the materials relating to this matter is ongoing and I will continue to review any new material that is provided to me. This Declaration represents certain opinions that I have formed to date and the bases for those opinions. I reserve the right to revise, supplement, and/or amend my opinions based on new information presented to me and on my continuing analysis of the materials cited herein or later provided to me.

V. LEVEL OF ORDINARY SKILL IN THE ART

51. I understand that Petitioner has advanced the following definition of a person having ordinary skill in the art (“POSITA”):

A person of ordinary skill in the art at the time of the alleged invention of the '896 patent (“POSITA”) would have had an undergraduate degree in electrical engineering, computer engineering, computer science or a related field along with at least four years of work experience in the field of media-related software development or in the alternative, equivalent experience, such as six years of work or research experience in the field of media-related software development.

'896 Pet., 7 / '849 Pet., 7-8 / '848 Pet., 7-8.

52. For purposes of this Declaration only, I have applied the definition of a POSITA proposed by Petitioner. However, my application of that definition is not a reflection of my agreement or disagreement with it. I reserve the right to provide my own opinion on the level of ordinary skill should there be further proceedings.

53. As discussed above, I have two years of college education in the field of electrical engineering, as well as over a decade of industry experience designing and implementing ACR systems, including audio-sample-based and video-sample-based ACR systems. I have also been awarded numerous patents covering ACR technologies. My experience and qualifications as described in Section II above exceed that of a POSITA and qualify me to testify as to the knowledge of a POSITA under any reasonable definition.

VI. OPINIONS ON CLAIM CONSTRUCTION

54. I have applied the plain and ordinary meaning to a POSITA of all claim terms at issue. I understand that the Petitioner does not contend that any terms require construction in these proceedings.

VII. OPINIONS ON WRITTEN DESCRIPTION SUPPORT

55. I have been asked to offer my opinion on whether a person having ordinary skill in the art would have recognized that the inventors possessed the inventions claimed in the '896, '848, and '849 Patents based on the written description of U.S. Patent Application Nos. 11/139,330 ("the '330 application"). I have been informed that everything appearing after the "CROSS REFERENCE TO RELATED APPLICATIONS" in the specification of the '330 application also exists in the specification of U.S. Patent Application Nos. 12/954,604 ("the '604 application"), 13/914,870 ("the '870 application"), and 14/175,811 ("the '811 application") and within the specification of the '896, '848 and '849 patents. As such, my conclusion as to the '330 application applies equally to these other applications and patents. As I explain in greater detail below, it is my opinion that a POSITA would have understood that the inventors possessed the challenged claims based on the written description that is common to these applications and patents.

A. The “BACKGROUND OF THE INVENTION”

56. I begin my analysis with the “BACKGROUND,” which is the first section of the common written description appearing after the section titled “CROSS REFERENCE TO RELATED APPLICATIONS.”

57. This section discusses different types “media measurement systems” that existed long before the ’330 application was filed (May 26, 2005). Ex. 1011 (‘330 application) ¶¶ [0002]-[0005]. It is worth first noting that the term “media measurement” is an industry term that a POSITA would understand to refer to the monitoring of media consumption habits of audience members. Ex. 2004, 11 (“Media measurement provides the subscriber with estimates of the number of people who consumed various types of media (radio, TV, print, online, outdoor, etc.) and the characteristics of those people.”). The data acquired by these systems is then used to help broadcasters and advertisers measure the level of consumption of particular content (e.g., advertisements, TV shows), which can help determine future content placement. *Id.* at 11. For example, having knowledge and insights into the habits of the audience consuming media content allows advertisers to determine where, when, and how to more effectively advertise in the future (e.g., where to most effectively place an ad).

58. The first type of media measurement system discussed is referred to or known as a “channel-centric” media measurement system:

Traditional media measurement systems have focused on directly monitoring channels being utilized by audience members. However, as media consumption patterns have become more complex, channel-centric media measurement is inadequate for many purposes. It may be desirable to track usage of particular media content independent of channel. Furthermore, although “channel” identification in a traditional media measurement system may sometimes be limited to radio or television broadcast station, it is increasingly desirable to track usage of media across several types of media delivery vehicles including radio, television, CD, DVD, computer download, portable media players (e.g. MP3 players, iPod), and other vehicles. Furthermore, with respect to tracking consumption of advertisements, it may be inadequate to simply track channel tuning, because, for example, an audience member may mute a broadcast during commercial periods. Thus simply identifying a broadcast channel does not adequately track whether the audience member listened to a particular advertisement.

Ex. 1011 (‘330 application), 577 (¶ [0002]). “Channel-centric” media measurement typically involves monitoring the channel (e.g., radio station or television channel) to which an audience member is tuned. For example, the television channels that an audience member tunes to may be monitored. While this provides some insights into an audience member’s media consumption habits, those insights are limited to identifying the channel audience members are watching and when. For instance, channel-centric media measurement systems did not provide insights into whether the audience member takes certain play-altering actions (e.g., muting, skipping) or whether the audience member consumed media (e.g., audio or video) over a different delivery mechanism (e.g., radio broadcast rather than television broadcast).

59. The second type of media measurement system discussed in the “BACKGROUND” are those that relied on embedded code or tag to identify or track the content being consumed by audience members:

Some media measurement systems have used codes to “tag” and track particular content. However, such systems are limited in that they can only track content that has been properly encoded.

Ex. 1011 (‘330 application), 577 (¶ [0003]). These “tag” based media measurement systems addressed some of the drawbacks of channel-centric measurement, because “tag” based media measurement allows media content itself to be tracked rather than simply the channel being selected by the audience member. Specifically, an identifier (e.g., “tag”) is encoded or embedded within the media content (audio data or video data), and the identifier can be used to later identify and track consumption or distribution of the content. While this may seem effective in theory, as the passage above explains “such systems are limited in that they can only track content that has properly been encoded.” A POSITA would understand that while “tagging” allows for the tracking of particular content consumed by an audience member (not simply the identity of the channel being listened to and/or watched), a drawback of such systems is the fact that the content itself must be encoded. This means, for example, that the distributor of the content must thus cooperate to ensure that the coding is embedded with the content such that the coding can be used to later identify that

content. Those attempting to rely on the coding to identify content must also obviously be familiar with the meaning of the code and understand how to extract it.

60. The third type of media measurement system discussed in the “BACKGROUND” are sample-based content recognition systems:

With the development of more robust content recognition technologies, some content recognition systems have recently been deployed which do not rely on codes. For example, Philips, Shazam Entertainment, and others have marketed systems for identifying songs played into a mobile phone. Although such systems can be efficiently deployed in the context of song recognition, deploying such systems in the context of media measurement systems poses particular challenges. Continuous searching against a large database of media content can be computationally intensive. Furthermore, such systems, while increasingly robust, still return some erroneous results, particularly in high-noise environments.

Ex. 1011 (‘330 application), 577-78 (¶ [0004]). A POSITA would have understood that this passage is teaching that “robust content recognition technologies” had been developed and existed to identify content by comparing captured data samples to samples associated with known content (e.g., TV shows, commercials).¹ Such content recognition systems do not need to rely on tags,

¹ These “known” samples are also commonly referred to a “reference samples,” because they are referenced to determine the identity of the unknown sample.

because the content is being identified based on the audio or video data sample itself, rather than an embedded code or tag that is added or embedded.

61. A POSITA would have understood that for both audio-sample and video-sample-based content recognition, a data sample (audio data or video data) is taken and then features (commonly referred to as “fingerprints” or “signatures”) of the sample are extracted and compared to features of reference samples to identify content. Specifically, content recognition using audio samples works by analyzing the audio signal to create a unique fingerprint which is then compared against a database of known audio fingerprints to identify the source of the audio, like a song or commercial, by matching it to the closest match in the database. Essentially, audio-sample content recognition extracts distinctive features from the audio signal like frequency patterns, tempo, and amplitude to create a unique identifier that can be recognized even with background noise or slight variations in quality. Content recognition using video samples works in a very similar manner. It extracts distinctive features from the video signal like color patterns, and pixel arrangement, and motion within the video frame to create a unique fingerprint that can be matched against the database.

62. The passage above provides specific examples of commercially marketed content identification systems that rely on audio-sample content recognition to identify songs. However, a POSITA would not understand the passage

itself to be limited to audio-based systems. The first sentence refers to “robust content recognition technologies,” which is not specific to audio-based systems but also includes those that are video-based. Additionally, the passage later makes clear that content recognition is applied in “media measurement systems” where a “large database of **media** content” is searched against. A POSITA would understand that “media content” encompasses both audio- and video-data-based content. Indeed, the next passage (discussed below) that specifically references “audio or video data samples,” which conveys to a POSITA that the disclosed system applies to both types of sample-based content recognition systems.

63. The passage concludes that content recognition systems “can be computationally intensive” given that a large database of media content may be required for matching purposes, and those systems may return erroneous results in high-noise environments. A POSITA would understand that “erroneous results in high-noise environments” refers to the fact that content recognition systems may monitor audio and/or video signals being transmitted over, for example, radio broadcast or television signal broadcast, and that the quality of the signals can be diminished by extraneous factors, such as signals from other sources (perceived as “noise” to the ACR system). This diminished signal may be harder for content recognition systems to identify, given that aspects of the underlying data may be missing or distorted.

64. The “BACKGROUND” concludes with a passage discussing the intended use and purpose of the invention:

At the same time, the media measurement context provides opportunities to utilize data exogenous to a particular **audio or video data sample**. Such opportunities have thus far been insufficiently exploited for the purpose of efficiently **applying existing content recognition technologies** in the media measurement context. Thus, an improved media measurement system and method **is needed**.

Ex. 1011 (‘330 application), 578 (¶ [0005]). This passage is important for what it conveys to a POSITA about the purpose and application of the invention disclosed in the common written description.

65. A POSITA would readily recognize this to passage to teach that the purpose or goal of the invention is to “apply[] **existing** content recognition technologies”—including existing audio and video data sampling technologies—in conjunction with additional data beyond the content identification results generated by those existing systems. Indeed, a POSITA would understand that the reference to “audio ... data sample” is referring to the application of these known, existing content recognition systems to identify content by sampling an audio signal and comparing the sample to reference audio samples. A POSITA would likewise understand that “video data sample” is referring to the application of these robust known content recognition systems to identify content by sampling a video signal and comparing the sample to reference video samples. As I later explain, audio-data

and video-data-based content recognition systems were well-known and understood in the art to identity content, and in fact the common written description expressly discusses such prior systems and states that the disclosed system “is adapted to implement one or more known content recognition methods that identify content by extracting parameters from a media signal and applying an algorithm to those parameters to search for a content match. Those skilled in the art will recognize that many such methods and algorithms exist. Ex. 1011, 584-85 (¶ [0025]).

66. This passage from the “BACKGROUND OF THE INVENTION” thus conveys to a POSITA that the purpose of the invention is to apply existing content recognition technologies that identify content based on audio or video data samples, and then “utilize[s] data exogenous” to improve upon the content identification that they provide. The novelty of the invention is thus agnostic as to whether an audio-sample or video-sample based content recognition systems is being used to generate content identification results. Rather, the improvement would be understood to lie in the use of additional data to improve the content identification results generated based on these existing technologies.

B. State of the Art

67. I have been informed and understand that determining whether a patent application provides adequate written description support requires consideration of the state of the scientific knowledge at the time of the invention as understood from

the perspective of a POSITA. I provide a detailed discussion of the state of the art, with a particular focus on audio-sample and video-sample-based content recognition systems given that these are the focus of Petitioner’s written description position.

68. As of May 26, 2005—i.e., the day the ‘330 application was filed—audio- and video-sample-based content recognition systems were well-known and understood in the art by POSITAs. Before turning to external evidence, I first note that the common written description itself evidences this. The “BACKGROUND OF THE INVENTION” refers to “applying **existing** content recognition technologies in the media measurement context” and specifically mentions “audio and video sample[s].” Ex. 1011 (‘330 application), 578 (¶ [0005]). In other words, the “BACKGROUND” teaches that such systems existed in the art and proposes applying such technologies as part of the disclosed system.

69. This is only further evidenced by the “DETAILED DESCRIPTION” of the ’330 application, which teaches that “Content recognition system 124 is adapted to implement one or more **known** content recognition methods that identify content by extracting parameters from a **media** signal and applying an algorithm to those parameters to search for a content match. **Those skilled in the art will recognize that many such methods and algorithms exist.**” Ex. 1011 (‘330 application), 584-85 (¶ [0025]). The common written description further

discusses a variety of “known” content recognition systems, including audio-data-sample-based and video-data-sample-based content recognition systems. *Id.*

70. As one example, the common written description references WO 2004/044820, which itself describes audio-data- and video-data-based content recognition systems as follows:

The method comprises the steps of extracting a set of robust perceptual features from the multimedia signal, and converting the set of features into the fingerprint. For audio signals, the perceptual features are energies of the audio contents in selected sub-bands. **For image signals, the perceptual features are average luminances of blocks into which the image is divided. The conversion into a binary sequence is performed by thresholding, for example, by comparing each feature 20 sample with its neighbors.**

Ex. 2026, 2:14-20 (discussed at Ex. 1011 (‘330 application), 584-85 (¶ [0025])). A POSITA would understand that the bolded language is referring to video-sample-based content recognition, where a video data sample is taken from a video signal so that perceptual features (e.g., luminance, pixel pattern) can be used to match against a reference database of video data samples. Notably, the same reference goes on to discuss various considerations applicable to audio-sample-based content recognition, and explains that “[s]imilar considerations apply to image and video material and to other kinds of perceptual features being used for fingerprint extraction.” *Id.* at 3:17-18. This is consistent with a POSITA’s understanding at the

time, as further evidenced below.

71. Similarly, WO 2004/040475 is discussed by the common written description and teaches that video-sample-based content recognition was known, and that:

Whilst the above embodiments of the present invention have been described with reference to audio information streams, **it will be appreciated that the invention can be applied to other information signals, particularly multi-media signals, including video signals.**

For instance, the article “J.C. Oostveen, A.A.C. Kalker, J.A. Haitsma, “Visual Hashing of Digital Video: Applications and Techniques”, SPIE, Applications of Digital Image Processing XXIV, July 31 - August 3 2001, San Diego, USA, describes a suitable technique for extracting essential perceptual features from a moving image sequence.

As the technique relates to visual fingerprinting, the perceptual features relate to those that would be viewed by the HVS i.e. it aims to produce the same (or a similar) fingerprint signal for content that is considered the same by the HVS. The proposed algorithm looks to consider features extracted from either the luminance component, or alternatively the chrominance components, computed over blocks of pixels.

Ex. 2024, 13:19-32 (discussed at Ex. 1011 (‘330 application), 584-85 (¶ [0025])).
Consequently, not only was audio- and video-sample-based content recognition

well-known and understood, it was also well-known and understood that their principles of operation were similar.

72. While the common written description reflects the well-known and understood nature of both audio- and video-data-based content recognition, I also discuss dozens of relevant references below that span the preceding decades.

73. One of the earliest discussed is U.S. Patent No. 4,230,990 (“Lert”), which was filed in 1979. It explains:

As an alternative to adding codes to the program material, the prior art includes a method of program identification whereby the audio or video program signal is analyzed in such a way that the program supplies its own unique code. Method B, therefore includes all attempts to solve the problem of automated broadcast monitoring solely through the use of computerized pattern recognition.

Pattern recognition consists of two basic processes: feature extraction and classification. The feature extraction process is applied to the program signal to produce a digital signature of a given program: certain features of the program signal are measured, and these measured values are used to characterize that program. The analog program signal is normally digitized by being passed through an analog-to-digital converter, and program information (**either audio or video**) is sampled and processed using some non-linear transform (which can be done either digitally or in analog) to produce a digital data set which is essentially unique to a particular program. Such a data

set is commonly referred to as a “signature”, “feature set”, or “feature vector”, terms which are to be considered as equivalent and are used interchangeably in this application.

From each program which is to be identified when broadcast, a segment of some duration is selected and used in the feature extraction process to produce the program’s “reference” signature. In the field, program information from the broadcast signal of a monitored station can be sampled for a like duration and used in the same feature extraction process to form a “broadcast” signature of unknown program identity. In the classification process, broadcast signatures are compared mathematically to reference signatures in order to identify the occurrence [sic] of known programs. When a broadcast signature is sufficiently similar to a reference signature, the broadcast program is identified as the know [sic] program.

Ex. 2006, 3:45-4:17. A POSITA would understand Lert to be teaching that sample-based automated content identification was understood in the art, and further that it involves sampling an audio or video signal (e.g., radio station signal for audio, television broadcast signal for video) to extract features (“signatures”) for comparison to reference signatures, which allows the content of the signal to be identified. Lert thus evidences that more than 25 years before the ’330 application was filed, the operation and implementation of audio- and video-sample-based content recognition was well-known. It simply involved taking an audio or video data sample from an audio and/or video signal, respectively, and comparing the

sample to reference samples associated with known content.

74. Specifically, a POSITA would have understood that for both audio- and video-sample-based automatic content recognition, a data sample is taken and then features (commonly referred to as “fingerprints” or “signatures”) of the sample are extracted. Those extracted features are then compared to features of reference samples associated with known content to determine whether there is a sufficient match.

75. Content recognition based on audio data samples operates by analyzing the audio signal to create a unique fingerprint which is then compared against a database of known audio fingerprints to identify the source of the audio (e.g., a song or commercial) by matching it to the closest match in the database. Basically, audio-sample-based content recognition extracts distinctive features from the audio signal (e.g., frequency patterns, tempo, and amplitude) to create a unique fingerprint that can be recognized even with background noise or slight variations in quality.

76. Video-sample-based content recognition operates in a very similar manner. It extracts distinctive features (e.g., color patterns, pixel arrangement, motion) from within the video frame to create a unique fingerprint that can be compared against a database of known video fingerprints to identify the source of the video, such as a TV show or commercial.

77. A POSITA would have understood that when the final passage of the “BACKGROUND” refers to the application of existing content recognition systems to “audio or video data sample[s],” it is referring to the application of these types of known audio-sample and video-sample content recognition systems.

78. In addition to Lert, there are many other references spanning the decades preceding the ’330 application that also evidence that audio- and video-sample-based content recognition systems were well-known and understood. To demonstrate this point, I have provided a table in Appendix A that lists references (by row) in chronological order by year, beginning in 1975 and ending in 2005. The “Relevant Disclosures” columns include exemplary quotations from each reference discussing audio- and/or video-sample-based content recognition systems. The purpose of Appendix A is to illustrate how well-known and well-understood both audio- and video-sample based content recognition are. I have bolded portions of the quoted passages to call attention to disclosures that may be considered most relevant for this purpose.

79. While these references evidence how well-known and understood audio- and video-sample-based content recognition was, these references also further demonstrate that POSITAs understood that systems built upon existing audio-sample content recognition technologies could typically be applied video-sample content recognition technologies and, likewise, that systems built upon

video-sample content recognition could typically be applied to audio-sample content recognition technologies. As some specific examples, U.S. Patent Application Publication No. 2006/0075237 (filed October 31, 2003), focuses on audio-sample content recognition but states “**similar considerations apply to image and video material** and other kinds of perceptual features being used for fingerprinting extraction” and “[a]lthough the invention has been described with reference to audio fingerprinting, **it can also be applied to other multimedia signals such as images and motion video.**” Ex. 2023, ¶¶ [0004]-[0005], [0009], [0027].

80. U.S. Patent Application Publication No. 2006/0041753 (filed Aug. 11, 2003) teaches that “[a]lthough embodiments of the method and arrangement have been described with reference to audio fingerprint extraction, the invention is not restricted thereto” and is applicable to “a video fingerprint extracting method **in which the fingerprint is derived from the mean luminance values of image blocks into which each image is divided**”). Ex. 2022, ¶[0019].

81. U.S. Patent Application Publication No. 2005/0193016 (filed Feb. 17, 2005) primarily focuses on embodiments that involve audio-data fingerprint matching but goes on to explain that “it should be understood that various embodiment of the media content database system could be used with any other type of media content or other types of media items” **including “video.”** Ex. 2030, ¶[0050].

82. U.S. Patent No. 6,990,453 (filed Apr. 20, 2001) focuses primarily on audio fingerprint matching. *See* Ex. 2016, Abstract (“A method for recognizing an audio sample locates an audio file that most closely matches the audio same from a database indexing a large set of original records.”). However, it also teaches that “[w]hile the following discussion refers primarily to audio data, it is to be understood that the method of the present invention can be applied to **any type of media** samples and media files, including but not limited to, text, audio, **video**, **image**, and any multimedia combination.” *Id.* at 5:36-46.

83. U.S. Patent Application Publication No. 2007/0271300 (filed May 22, 2007), teaches that: “In general, audio signature-based media content identification techniques are computationally efficient and require less memory than known video signature-based techniques. However, **either audio or video signature-based techniques** may be employed, and because **many such techniques are known in the art**, further discussion of these techniques is not provided herein.” Ex. 2031 ¶ [0031].

84. Many of the references discussed above also evidence that audio and video fingerprinting operate in fundamentally the same way and that they are typically described together without distinction by those of skill in the art. *See e.g.*, Ex. 2022 ¶ [0020] (“The invention can be summarized as follows. Fingerprints are bit strings extracted from **a media signal (e.g., an audio or video clip)** to identify

said media signal. Typically, they are derived from a perceptual property of the signal, for example, the **spectral energy distribution of an audio fragment or the luminance distribution of a video image.**”); Ex. 2006, 3:45-4:17.

85. To summarize my opinion on the “BACKGROUND OF THE INVENTION,” it clearly conveys that the invention relies on existing audio- and video-sample-based content recognition systems and that such systems were well-known and understood in the art. The “BACKGROUND” also teaches that the novel improvement lies not in the use of those existing content recognition technologies, but in the use of additional data beyond the content identification results generated by such systems to deduce additional useful information about audience consumption (e.g., play-altering actions, more accurately identify consumed content by “scrubbing” a play stream, and/or identification of a channel content is being consumed on).

C. The “SUMMARY”

86. It is my opinion that, in addition to the “BACKGROUND OF THE INVENTION,” the “SUMMARY” further conveys to a POSITA that the inventors possessed the invention as applied to both audio and video data samples. Immediately following the final passage of the “BACKGROUND,” the “SUMMARY” recites:

Some embodiments of the present invention provide a media

measurement system and method that enhances recognition (e.g. in terms of accuracy or efficiency) of the content of a media sample by analyzing information exogenous to the sample. Some embodiments of the present invention provide a media measurement system and method that utilizes audience data to enhance content identifications. Some embodiments analyze media player log data to enhance content identification. Other embodiments of the present invention analyze sample sequence data to enhance content identifications. Other embodiments analyze sequence data to enhance content identification and/or to establish channel identification. Yet other embodiments provide a system and method in which sample construction and selection parameters are adjusted based upon identification results. Yet other embodiments provide a method in which play-altering activity of an audience member is deduced from content offset values of identifications corresponding to captured samples. Yet other embodiments provide a monitoring and measurement system in which a media monitoring device is adapted to receive a wireless or non-wireless audio signal from a media player, the audio signal also being received wirelessly by headphones of a user of the monitoring device.

Ex. 1011, 578 (¶ [0006]). A POSITA would understand that the first sentence of this passage teaches that the disclosed invention relies on existing audio- and video-sample-based content recognition systems and “enhances” their recognition using information external to the sample itself. For one, this statement immediately follows, and is consistent with, the “BACKGROUND” statement that “the **media**

measurement context provides opportunities to utilize data exogenous to a particular **audio or video data sample**,” and teaches recognizing the content of a “media sample,” which a POSITA would understand refers to both audio and/or video data samples. Ex. 2034, 3:66-4:1 (explaining that a “media sample[s] may be audio or video samples”). A POSITA would understand that if the invention had been intended to be limited to audio sample content recognition, it would refer to “audio sample” rather than “media sample.” By using the word “media sample” the inventors clearly conveyed the applicability of the invention to both audio and video samples. This is even more apparent to a POSITA given that the immediately preceding passage from the “BACKGROUND OF THE INVENTION” expressly references content recognition as applied to an “audio or video data sample.”

87. A POSITA would understand that “media” refers to both audio and video and “media sample” refers to both audio and video data samples. It would further be understood that both “audio and video data sample[s]” are referenced in the immediately preceding “BACKGROUND” passage as being “media” and, as I discuss in greater detail below, the disclosed invention is agnostic as to whether a media measurement system relies on audio or video data samples—the invention improves both existing audio- and video-sample-based content recognition technologies in precisely the same way by using additional data beyond the results generated by those systems. As such, the reference to applying the invention to

“media samples” is sufficient in and of itself to convey that the inventors possessed the invention as applied to audio and video data samples, but even more so when viewed in light of the final passage in the “BACKGROUND OF THE INVENTION.”

88. The remaining portion of the “SUMMARY” is consistent with this understanding. That portion refers to a “media” measurement system, “sample sequence data” and “sequence data,” which encompasses both audio- and video-sample-based data. For instance, “sample sequence data” can be audio sample sequence data or video sequence sample data, and the “SUMMARY” does not limit the sequence data to one or the other. Likewise, the reference to a “media” measurement system indicates to a POSITA that the system measures audio- and video-data based media.

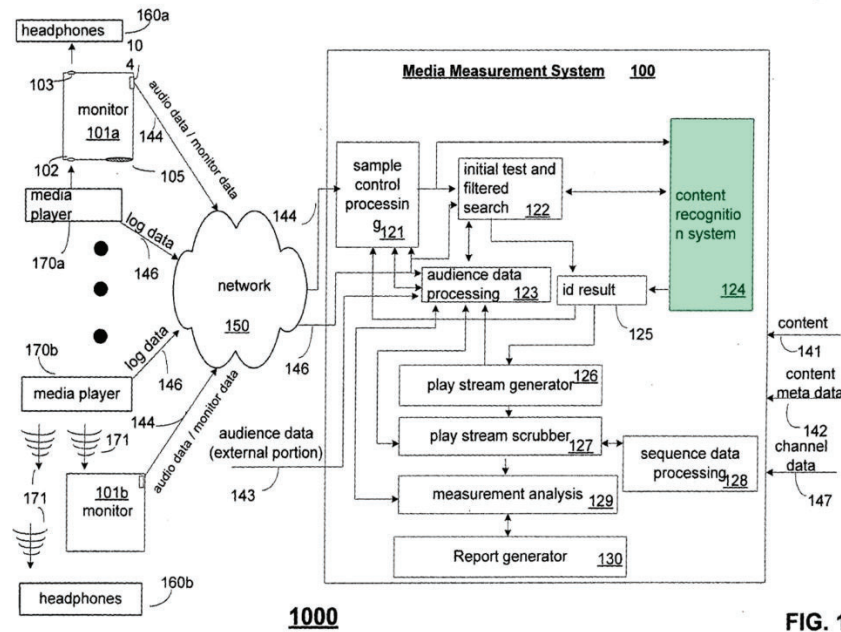
89. I note that the final sentence does mention an “audio signal” and not “video signal.” However, a POSITA would understand that is merely one embodiment because the sentence begins with “[y]et **other** embodiments.” A POSITA would not understand the reference to one example in the “SUMMARY” to mean that the invention is limited to only that one example, and in fact the first sentence of the “SUMMARY” makes clear that the invention applies to any type of media sample.

90. As I discussed in greater detail below, a POSITA would understand that the disclosed invention is agnostic as to how the content identification results are generated—it simply takes those results (whether based on audio and/or video data sample matching) and deduces additional information about audience member consumption habits (e.g., play-skipping, channel identification, more accurate identification of consumed content via “scrubbing”) using additional data. A POSITA would thus understand that, in addition to the “BACKGROUND,” the “SUMMARY” demonstrates that the inventors possessed the invention as applied to video data samples.

D. The “DETAILED DESCRIPTION”

91. A POSITA would understand that the systems disclosed in the common written description—including in the “DETAILED DESCRIPTION”—rely on content identification results generated using existing content recognition systems. Figure 1 provides an example of the invention as applied to content identification results generated by audio-sample-based content recognition systems, but as demonstrated below a POSITA would understand the system applies equally and in the same manner to video-sample-base content recognition systems.

92. In reference to Figure 1 (annotated below), media monitoring devices 101 capture audio data from media players 170 and transmit the data to media measurement system 100 over network 150. Ex. 1011, 580-81 (¶¶[0016]-[0019]).



The audio data may include audio sampled by monitor 101b or the sampling can be performed by sample control processing unit 121, and then content recognition of the sample is performed by content recognition system 124 (green) by matching the sample against a reference stored in a database. *Id.* at 583-84 (§§ [0023]-[0025]).

93. Notably, the common written description does not provide details on how the content recognition system identifies the content of the captured samples. The written description instead teaches that “[c]ontent recognition system 124 is adapted to implement **one or more known** content recognition methods that identify content by extracting parameters from a **media** signal and applying an algorithm to those parameters to search for a content match. Those skilled in the art will recognize that **many such methods and algorithms exist.**” *Id.* at 584 (§[0025]). A POSITA would understand that a “**media** signal” includes both audio and video signals and

thus that the specification teaches that content recognition systems that extract parameters from a media signal (i.e., an audio *or* video signal) to identify content were well-known in the art and could be used with the disclosed invention. This is because a “media signal” includes audio and video signals. Indeed, a POSITA would understand that if the inventors had intended for this sentence to refer only to audio, they would have referenced “audio signals” rather than “media signals.”

94. The written description provides numerous examples of prior art references (including Exhibits 2024 and 2026) that teach audio- and/or video-sample-based content recognition, and discusses various considerations in deciding which fingerprint matching algorithm to select. *Id.* at 584-85 (¶[0025]). I discuss Exhibits 2024 and 2026 extensively above in Section VII.B to show that each reference describes both audio-sample and video-sample based content recognition. A POSITA would understand that it is the content identification results output by these well-known, existing content recognition systems that the disclosed invention builds upon, using the techniques discussed immediately below.

95. In one embodiment shown in Figure 4 (annotated below), a “Raw Play Stream” 410 of content identification results (green) is generated, which includes content identifiers obtained using existing content recognition technology. *Id.* at 590 (¶[0040]). For example, the content identifier of row 1 indicates that there was a

match with a reference sample associated with “song 51,” while row 2 indicates that there was no sufficient match with any reference sample.

410			450			470			
Raw Play Stream			Clean Play Stream			Channel Data for “radio 35”			
row	log time	content ID	row	log time	content ID	row	start time	end time	content ID
1	t_0	song 51	1	t_0	song 51	1	C(0)	C(0)+6	song21
2	t_0+1	none	2	t_0+1	song51	2	C(0)+6	C(0)+8	ad49
3	t_0+2	song51	3	t_0+2	song51	3	C(0)+8	C(0)+13	song35
4	t_0+3	song51	4	t_0+3	song51				
5	t_0+4	movie 81	5	t_0+4	song51				
6	t_0+5	song51	6	t_0+5	song51				
7	t_0+6	song51	7	t_0+6	song51				
8	t_0+7	song72	8	t_0+7	song72				
9	t_0+8	song72	9	t_0+8	song72				
10	t_0+9	none	10	t_0+9	song72				
11	t_0+10	song72	11	t_0+10	song72				
12	t_0+11	song72	12	t_0+11	song72				
13	t_0+12	song72	13	t_0+12	song72				
14	t_0+13	song65	14	t_0+13	song65				
15	t_0+14	song65	15	t_0+14	song65				
16	t_0+15	TVprgm58	16	t_0+15	song65				
17	t_0+16	song65	17	t_0+16	song65				
18	t_0+17	song65	18	t_0+17	song65				
19	t_0+18	song65	19	t_0+18	song65				
20	t_0+19	song65	20	t_0+19	song65				
21	t_0+20	song21	21	t_0+20	song21				
22	t_0+21	song21	22	t_0+21	song21				
23	t_0+22	song21	23	t_0+22	song21				
24	t_0+23	song21	24	t_0+23	song21				
25	t_0+24	song21	25	t_0+24	song21				
26	t_0+25	song21	26	t_0+25	song21				
27	t_0+26	ad49	27	t_0+26	ad49				
28	t_0+27	ad49	28	t_0+27	ad49				
29	t_0+28	song81	29	t_0+28	song35				
30	t_0+29	song35	30	t_0+29	song35				

480			
Channel Data for “album 12”			
row	start time	end time	content ID
1	C(0)	C(0)+7	song51
2	C(0)+7	C(0)+13	song72
3	C(0)+13	C(0)+23	song65
4	C(0)+23	C(0)+30	song29

490				
Clean Play List				
row	start log time	end log time	content ID	Channel ID
1	t_0	t_0+7	song51	album12
2	t_0+7	t_0+13	song71	album12
3	t_0+13	t_0+20	song65	album12
4	t_0+20	t_0+26	song21	radio35
5	t_0+26	t_0+28	ad49	radio35
6	t_0+28	t_0+30	song34	radio35

ssd = “sample sequence data”

FIG. 4

The process of obtaining a content identifier using existing content recognition technology (e.g., the identity of a song based on an audio data sample or a TV show based on a video data sample) was well-known and understood as applied to both audio and video data samples, as explained above in Sections VII.A-B. A POSITA would understand that the technical innovation provided thus does not lie in applying these existing content recognition technologies to generate a content identification result. Rather, it lies in what happens next.

96. As one example, after existing content recognition technologies are used to generate content identification results (green) as part of a “raw play stream,”

that play stream is then “scrubbed” to create “Clean Play Stream” 450 by correcting matches that are incorrect (red) or supply the identification of missing content (purple) based on, for example, an expected pattern of sample sequence results. Ex. 1011, 590-91 (¶¶[0040]-[0042]).

97. This feature is agnostic as to whether the content identification results were generated based on an audio or video data sample. Indeed, in either case the system simply relies on the content identifiers (e.g., the names of songs or TV programs) with no regard as to whether that content was identified based on audio- or video-sample-based content recognition. To illustrate this point, if the exact same scrubbing process were applied to a video-sample-based content recognition system, the “content ID” of Raw Play Stream 410 would identify video content (e.g., movie81, TVpgrm58) based on video data, rather than identifying audio-only content (e.g., song51) or audio-video content (e.g., TVpgrm58) based on audio data. However, the “scrubbing” process described would be the same, because that process begins with the content identification results (green) in the Raw Play Stream 410. In other words, the process is agnostic as to how those results are generated. A POSITA would thus understand that whether audio- or video-sample-based content recognition is applied is incidental to the disclosed invention. The “scrubbing” feature above applies in the same way—and is just as effective—as applied to content identification results generated based on video data rather than audio data.

98. The “DETAILED DESCRIPTION” disclose another feature that is similarly agnostic as to how the content identifiers are initially generated. In reference to Figure 5 (annotated below), a clean play stream of content identification results (green) is generated. Ex. 1011, 593 (¶[0047]). However, log times (“time at which capture of the analyzed sample began”) and content offsets (“time offset from the beginning of a particular piece of content”) are also associated with each result and can be used to deduce play-altering actions of audience members (e.g., skipping forward or backward, or switching content). *Id.* at 593-95 (¶¶[0047]-[0052]).

510 Clean Playstream			
row	log time	content ID	content offset
1	t ₀	TVprgm 27	(0)
2	t ₀ +1	TVprgm 27	(0)+1
3	t ₀ +2	none	n/a
4	t ₀ +3	none	n/a
5	t ₀ +4	none	n/a
6	t ₀ +5	TVprgm 27	(0)+2
7	t ₀ +6	TVprgm 27	(0)+3
8	t ₀ +7	TVprgm27	(0)+9
9	t ₀ +8	TVprgm27	(0)+10
10	t ₀ +9	TVprgm27	(0)+9
11	t ₀ +10	TVprgm27	(0)+10
12	t ₀ +11	none	n/a
13	t ₀ +12	TVprgm27	(0)+29
14	t ₀ +13	TVprgm27	(0)+30
15	t ₀ +14	ad23	(0)
16	t ₀ +15	none	n/a
17	t ₀ +16	ad25	(0)+2
18	t ₀ +17	TVprgm28	(0)
19	t ₀ +18	TVprgm28	(0)+1
20	t ₀ +19	TVprgm28	(0)+2

560 Clean Play List				
row	start log time	end log time	content ID	content offset
1	t ₀	t ₀ +2	TVprgm27	(0)
2	t ₀ +2	t ₀ +5	none	n/a
3	t ₀ +5	t ₀ +7	TVprgm27	(0)+2
4	t ₀ +7	t ₀ +9	TVprgm27	(0)+9
5	t ₀ +9	t ₀ +11	TVprgm27	(0)+9
6	t ₀ +11	t ₀ +12	none	n/a
7	t ₀ +12	t ₀ +14	TVprgm27	(0)+29
8	t ₀ +14	t ₀ +15	ad23	(0)
9	t ₀ +15	t ₀ +16	none	n/a
10	t ₀ +16	t ₀ +17	ad25	(0)+2
11	t ₀ +17	t ₀ +20	TVprgm28	(0)

570 Channel Data for TV C4			
row	start time	end time	content ID
1	C(0)	C(0)+31	TVprgm27
2	C(0)+31	C(0)+34	ad23
3	C(0)+34	C(0)+37	ad24
4	C(0)+37	C(0)+40	ad25
5	C(0)+40	C(0)+50	TVprgm28

590 Clean Play List with Deduced Actions				
row	log time	end log time	content ID and <deduced action>	content offset
1	t ₀	t ₀ +2	TVprgm27	(0)
2	t ₀ +2	t ₀ +5	<pause for 3 units>	n/a
3	t ₀ +5	t ₀ +7	TVprgm27	(0)+2
4	t ₀ +7	t ₀ +7	<skip frwrd 5 units>	n/a
5	t ₀ +7	t ₀ +9	TVprgm27	(0)+9
6	t ₀ +9	t ₀ +9	<skip bckwrd 2 units>	n/a
7	t ₀ +9	t ₀ +11	TVprgm27	(0)+9
8	t ₀ +11	t ₀ +12	<fast frwrd 18 units>	n/a
9	t ₀ +12	t ₀ +14	TVprgm27	(0)+29
10	t ₀ +14	t ₀ +15	ad23	(0)
11	t ₀ +15	t ₀ +16	<fast frwrd 7 units>	n/a
12	t ₀ +16	t ₀ +17	ad25	(0)+2
13	t ₀ +17	t ₀ +20	TVprgm28	(0)

FIG. 5

As the “DETAILED DESCRIPTION” from the common written goes onto explain, “[h]ad the media consumption from one sample to the next been continuous, one

would expect that the second offset value (in row 8) would have been (0)+4 rather than (0)+9.” Ex. 1011, 594 (¶[0050]). Because they are not, “a boundary between media consumption events can be deduced.” *Id.* This embodiment is thus able to create a “Clean Play List with Deduced Actions” 590 based on the content identification results obtained using existing content recognition technologies.

99. This methodology applies in the same way, and with the same effectiveness, to content identification results generated based on video data samples. Indeed, just like audio samples have corresponding log times (“time at which capture of the analyzed sample began”) and content offsets (“time offset from the beginning of a particular piece of content”), video samples may likewise have them.

100. The “DETAILED DESCRIPTION” discloses another feature that is similarly agnostic as to how the content identifiers are initially generated. In reference to Figure 4 (annotated below), the written description teaches that the channel that an audience member is consuming content on can be identified using content identification results. Ex. 1011, 591 (¶[0043]).

410 Raw Play Stream			450 Clean Play Stream			470 Channel Data for "radio 35"				480 Channel Data for "album 12"				490 Clean Play List				
row	log time	content ID	row	log time	content ID	row	start time	end time	content ID	row	start time	end time	content ID	row	start log time	end log time	content ID	Channel ID
1	t_0	song 51	1	t_0	song 51	1	$C(0)$	$C(0)+6$	song21	1	$C(0)$	$C(0)+7$	song51	1	t_0	t_0+7	song51	album12
2	t_0+1	none	2	t_0+1	song51	2	$C(0)+6$	$C(0)+8$	ad49	2	$C(0)+7$	$C(0)+13$	song72	2	t_0+7	t_0+13	song71	album12
3	t_0+2	song51	3	t_0+2	song51	3	$C(0)+8$	$C(0)+13$	song35	3	$C(0)+13$	$C(0)+23$	song65	3	t_0+13	t_0+20	song65	album12
4	t_0+3	song51	4	t_0+3	song51					4	t_0+20	t_0+26	song21	4	t_0+20	t_0+26	song21	radio35
5	t_0+4	movie 81	5	t_0+4	song51					5	t_0+26	t_0+28	ad49	5	t_0+26	t_0+28	ad49	radio35
6	t_0+5	song51	6	t_0+5	song51					6	t_0+28	t_0+30	song29	6	t_0+28	t_0+30	song34	radio35
7	t_0+6	song51	7	t_0+6	song51													
8	t_0+7	song72	8	t_0+7	song72													
9	t_0+8	song72	9	t_0+8	song72													
10	t_0+9	none	10	t_0+9	song72													
11	t_0+10	song72	11	t_0+10	song72													
12	t_0+11	song72	12	t_0+11	song72													
13	t_0+12	song72	13	t_0+12	song72													
14	t_0+13	song65	14	t_0+13	song65													
15	t_0+14	song65	15	t_0+14	song65													
16	t_0+15	TVprgm58	16	t_0+15	song65													
17	t_0+16	song65	17	t_0+16	song65													
18	t_0+17	song65	18	t_0+17	song65													
19	t_0+18	song65	19	t_0+18	song65													
20	t_0+19	song65	20	t_0+19	song65													
21	t_0+20	song21	21	t_0+20	song21													
22	t_0+21	song21	22	t_0+21	song21													
23	t_0+22	song21	23	t_0+22	song21													
24	t_0+23	song21	24	t_0+23	song21													
25	t_0+24	song21	25	t_0+24	song21													
26	t_0+25	song21	26	t_0+25	song21													
27	t_0+26	ad49	27	t_0+26	ad49													
28	t_0+27	ad49	28	t_0+27	ad49													
29	t_0+28	song81	29	t_0+28	song35													
30	t_0+29	song35	30	t_0+29	song35													

ssd = "sample sequence data"

FIG. 4

According to this feature, the "Clean Play Stream" (element 450) is compared to known "Channel Data for 'radio 35'" (element 470) and known "Channel Data for 'album 12'" (element 480), which indicates that the rows 1-20 are associated with "album 12" (yellow) and rows 21-30 are associated with "radio35" (red). As with the other discussed features, this feature is also agnostic as to whether the content identification results are audio- or video-sample-based. The only distinction when applying this feature to content identification results generated by video-sample-based content recognition systems is that the "content ID" of Raw Play Stream 410 would identify video content (e.g., movie81, TVprgm58) based on video data, rather than identifying audio-only content (e.g., song51) or audio-video content (e.g.,

TVpgrm58) based on audio data, and the “channel” data would be associated with video content (e.g., a television channel or a movie) rather than audio-only content (e.g., radio35).

101. As explained above, each of these disclosed features relies on existing content recognition systems—both audio-sample and video-sample based—that were well-known and understood in the art and that are incidental to the novelty of the invention. The “BACKGROUND” and “SUMMARY” convey to a POSITA that the invention is applicable to audio- and video-sample-based content recognition systems, and a POSITA would understand that the invention applies equally to, and provides the same benefit for, both kind of systems. Indeed, the very first statement in the “DETAILED DESCRIPTION” is:

The following description is presented to enable any person skilled in the art to make and use the invention, and is provided in the context of particular applications and their requirements. **Various modifications to the exemplary embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the invention.** Thus, the present invention is not intended to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles and features disclosed herein.

Ex. 1011, 580 (¶[0015]). A POSITA would understand that the teachings of the “DETAILED DESCRIPTION” applies to both audio-sample-based and video-

sample-based media measurement systems, and that the applicability of the teachings of the Figures 1-6 embodiment to video-sample-based content recognition systems is “readily apparent” and nothing more than the application of the generic principles of that embodiment to then-existing video data sampling systems referenced in the “BACKGROUND” and “SUMMARY.”

E. The Challenged Claims

102. The Challenged Claims of each of the '896, '848 and '849 Patents are fully supported by the common written description.

103. Independent claims 1 of the '896 Patent, '848 and '849 Patents are exemplary of the independent claims in their respective patents and are reproduced below. Each reference to “video” is in bold type.

'896 Patent

1. A media measurement method comprising:

using one or more computers to generate a play stream of content identification results corresponding to a sequence of submitted data samples, the submitted data samples corresponding to **video** data samples captured at a monitoring device associated with a monitored audience member; using one or more computers to utilize sample sequence data and content offset data to deduce play-altering actions of the monitored audience member, the sample sequence data corresponding to a sequence of the content identification results, the content offset data including content offsets associated with the content identification results, a content offset indicating a time position of a content identification result within identified content, wherein to utilize comprises, for a particular plurality of submitted data samples and corresponding content identification results,

comparing a progression of log times associated with capture timing of each of the particular plurality of submitted data samples and a progression of content offsets of the corresponding content identification results; and

using one or more computers to utilize deduced play altering actions of the monitored audience member to generate a media measurement report.

'848 Patent

1. A media measurement method comprising:

(a) receiving over a network, at one or more computers, data for a sequence of **video** data samples, the **video** data samples comprising representations of **video** data captured at a media monitoring device, the plurality of **video** data samples being submitted over the network;

(b) using the one or more computers to query an electronic database of a plurality of **video** data representations and corresponding content identifiers;

(c) generating a raw play stream, the raw play stream comprising a sequence of content identification results corresponding to the sequence of **video** data samples; wherein:

the sequence of content identification results is obtained by querying the electronic database to attempt to determine respective likely matches between respective video data samples in the sequence of **video** data samples and respective **video** data representations in the electronic database;

a content identification result of the sequence of content identification results comprises either: (i) a content identifier associated in the electronic database with a respective **video** data representation that is determined to be a respective likely match with a respective one of the **video** data samples; or (ii) an indication of the absence of a respective likely match between a respective **video** data sample and a **video** data representation in the electronic database; and

the raw play stream includes either: at least two different content identifiers obtained from the electronic database; or at least one content identifier obtained from the electronic database and at least one indication of the absence of a respective likely match between a respective **video** data sample and a **video** data representation in the electronic database;

(d) scrubbing the raw play stream by analyzing sample sequence data of the raw play stream to determine whether to change a result of the sequence of content of identification results in view of a pattern of the sample sequence data of the raw play stream compared to an expected pattern of sample sequence data; and

(e) generating a clean play stream from the raw play stream by making any changes to the raw play stream that are determined to be made by the scrubbing.

'849 Patent

1. A media measurement method comprising:

using one or more computers to generate a play stream of content identification results corresponding to a sequence of data samples, the data samples comprising representations of **video** data captured at a media monitoring device, wherein content identifiers of the play stream's content identification results are obtained by using corresponding data samples to search a computerized database of known content;

in response to obtaining the content identifiers, using one or more computers to utilize a sequential order of at least two different obtained content identifiers in the play stream to identify a channel corresponding to the data samples, the at least two different obtained content identifiers identifying different media content items; and using one or more computers to utilize an identified channel corresponding to the data samples to generate a media measurement report.

104. I understand that Petitioner contends that the common written description does not demonstrate possession of the challenged claims because of the inclusion of the word “video” rather than “audio” in the claims. Petitioner is mistaken for several reasons.

105. For one, the common written description teaches the invention applies “existing” content recognition systems based on “audio or video data sample[s]” and such systems were well-known and understood. *Supra* Section VII.A-B. Second, the “SUMMARY” teaches that the invention applies to “media samples,” which refers to both audio and video data samples. *Supra* Section VII.C. Third, the “DETAILED DESCRIPTION” teaches that algorithms for matching media signals (i.e., both audio and video signals) were well-known in the art and that applying the principles of operation of the audio embodiments to other embodiments would be “readily apparent” and the novel feature disclosed in the detailed embodiments is entirely agnostic as to how the content identification results are generated. *Supra* Section VII.D. A POSITA would thus not need or expect the inventors to disclose how content identification results are generated based on audio- or video-data samples, because that was within the knowledge of a POSITA based on existing technology.

106. Each independent claim of the ’896, ’848 and ’849 Patents referenced above relies on content identification results generated based on “video data” and/or

“video data samples,” which is nothing more than the application of existing content recognition technologies well-known and understood in the art. In other words, the recitation of “video data” and “video data samples” in the claims merely convey that the content identification results are generated based on video data samples. This aspect of the claims is thus clearly supported by the written description, as this involves nothing more than the application of existing video-sample-based content recognition technologies that are expressly mentioned and discussed in the common written description. Indeed, the “BACKGROUND OF THE INVENTION” expressly conveys that the novel features of the disclosed system may apply existing “audio or video sample” based content recognition technologies, and the “SUMMARY” teaches that the system applies to “media sample[s]” (audio or video data samples). The “DETAILED DESCRIPTION” even discusses several references that describe audio- and video-sample-based content recognition technologies, though I note that a POSITA would not require this discussion to understand that the inventors possessed the claimed inventions. All other aspects of the claim are agnostic to whether the “content identification results” are generated based on audio or video data, and thus they are also supported by the “DETAILED DESCRIPTION,” which provides specific examples of applying these same steps based on content identification results obtained based on audio data samples. Indeed, as explained above in Section VII.D, deducing play-altering actions using content

offsets and log times, “scrubbing,” and identifying a channel as taught in the “DETAILED DESCRIPTION” operate in the same way, and provides the same benefit, as applied to video data samples.

VIII. OPINIONS ON SPECIFIC ISSUES RAISED BY PETITIONER

107. As I explain in Section VII above, the common written description conveys to a POSITA that the invention applies “existing” and “robust” content recognition technologies based on “audio or video data sample[s],” that the invention applies to “media sample[s]” (which includes both audio and video data samples), and that the improvement provided lies in the use of additional data beyond the results provided by those well-known and understood technologies. A POSITA would have thus not required the inventors to disclose a detailed embodiment showing how the invention applies to content identification results generated using a video-sample-based content recognition system to demonstrate that the inventors possessed the claimed invention. I understand, however, that Petitioner has taken a different position on this issue. I address those positions in more detail below.

108. The common written description’s teaching that it applies “existing” and “robust” content recognition technologies, including those based on “audio or video data sample[s],” is sufficient in and of itself to convey that the inventors possessed the invention as applied to video data samples. Ex. 1011 (‘330

application), 578 (¶ [0005]). I understand that Petitioner disagrees and takes the following position:

Even in this instance, it is unclear if the term “video data sample” is used in the context of identifying content by sampling video data or if the term, as used in the context of the sentence (“the media measurement context provides opportunities to utilize data exogenous to a particular audio or video data sample”), is used to reference video content which includes an audio component, and the audio component is sampled for content recognition (since audio-sampling-based content recognition is the only type of content recognition described in the ’330 application). Indeed, the ’330 application does not describe any content recognition using a “video data sample.”

’896 Pet., 26-27 (citing Ex. 1002 ¶75) / ’849 Pet., 27 (EX1002, ¶ 72) / ’848 Pet., 22-23 (EX1002, ¶ 72).

109. I first note that Petitioner does not cite any evidence to support this assertion, other than nearly identical statements from its expert. In any event, there is no basis for a POSITA to interpret or understand the reference to a “video data sample” as referring to the audio portion of video content. I have referenced numerous references demonstrating that both audio- and video-sample-based content recognition systems were well-known and understood in the art at the time of the invention, as reflected in Section VII.B (discussing exhibits 2005-2035). This discussion describes Exhibit 2026, which is discussed in the common written description and demonstrates that audio- and video-sample-based content

recognition was well-known and understood. It was well-understood that content recognition based on “audio data samples” involved analyzing audio data extracted from an audio signal to extract fingerprints (e.g., frequency patterns, tempo, and amplitude) and “video data samples” involved analyzing video data extracted from a video signal to extract fingerprints (e.g., pixel patterns, color patterns, luminance, motion). In fact, I note that patents that appear to be assigned to Samsung demonstrate that it uses the term “video data” to refer to data sampled from a video signal, rather than data sampled from an audio signal. See e.g., Ex. 2036, Abstract, 3:47-49 (discussing “[u]ser media data, i.e., audio or video data that is saved inside an MPEG file”); Ex. 2037, 4:15-27 (discussing an apparatus for “coding video data” that includes “input means for receiving two or more frames of video data representative of image sequences”).

110. I have decades of experience in the ACR industry, and I have never seen or heard anyone describe audio-based content recognition as relying on, or being applied to, “video data samples.” Instead, the reference to “video data samples” would be understood to refer to the data from a video signal.

111. I note that the challenged independent claims include the same “video data sample” (or simply “video data” in the ‘849 patent) language referenced in the “BACKGROUND.” See e.g., ’896 patent, cl. 1, 9 (“the submitted data samples corresponding to *video data samples* captured at a monitoring device”); ’848 Patent,

cl. 1, 9 (“receiving over a network, at one or more computers, data for a sequence of *video data samples*”); ’849 Patent, cls. 1, 6, 10 (“the data samples comprising representations of *video data* captured at a media monitoring device.”). If, as Petitioner states, the term “video data sample” in the “BACKGROUND” could somehow be understood to refer to the audio portion associated with video content, that would mean that “video data samples” and “video data” in the claims would have that same meaning. If that were the true, then that would mean that the common written description even more clearly provides written description support for the challenged claims, because the “DETAILED DESCRIPTION” provides a detailed embodiment as applied to audio data associated with video content (e.g., a TV show). Ex. 1011, 582 (¶[0022]).

112. It is also my opinion that a POSITA would not understand “video data sample” in the “BACKGROUND OF THE INVENTION” to refer to the audio data associated with video content (e.g., a movie or TV show) because the same sentence already refers to an “audio ... data sample.” There would be no need to refer to a “video data sample” at all if the intent of the inventors was to refer to data taken from an audio signal, because the reference to “audio ... data sample” already conveys that those existing content recognition systems analyzed audio data of any type (including audio associated with video content, such as a movie).

113. Petitioner is also mistaken that a POSITA would understand “video data sample” to refer to audio data content recognition because audio-sampling-based content recognition is allegedly “the only type of content recognition described in the ’330 application” and allegedly “the ’330 application does not describe any content recognition using a ‘video data sample.’” ’896 Pet., 26-27 / ’849 Pet., 27 / ’848 Pet., 23. This disregards the fact that the “BACKGROUND” refers to a “video data sample” and that the reference to “media sample[s]” in the “SUMMARY” encompasses both audio and video data samples. The reference to “existing” and “robust” content recognition systems that analyze “audio or video data sample[s]” **does** in fact disclose a content recognition system that is video sampling based. A POSITA would thus not understand the common written description to only disclose audio-data-based sampling.

114. A POSITA also would not understand the “DETAILED DESCRIPTION” to limit what is taught by the “BACKGROUND” and “SUMMARY.” Instead, a POSITA would understand that the “DETAILED DESCRIPTION” is intended to provide specific examples of embodiments of the invention, but that the invention is not necessarily limited to those embodiments. This is particularly true given that the very first passage in the “DETAILED DESCRIPTION” explains that its disclosed embodiments are not intended to be limiting:

The following description is presented to **enable any person skilled in the art to make and use the invention**, and is provided in the context of particular applications and their requirements. Various modifications to the exemplary embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the invention. **Thus, the present invention is not intended to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles and features disclosed herein.**

Ex. 1011, 580 (¶ [0015]).

115. It is also my opinion that if “video data sample” was intended to refer to the audio data associated with video content (i.e., the audio associated with a movie), a POSITA would expect the “DETAILED DESCRIPTION” to use the term “video data” or “video data sample” consistent with that meaning. However, the “DETAILED DESCRIPTION” never discusses the audio being analyzed as “video data.” A POSITA would understand that the reason the written description never refers to the analysis of audio associated with a movie as involving a “video data sample,” is because such analysis relies on “audio data samples.” “Video data samples” are those taken from a portion of a video signal and used to create video data fingerprints (e.g., luminance-based fingerprints).

116. As discussed extensively in Section VII.C above, the “SUMMARY” conveys to a POSITA that the invention applies to content recognition technology

based on video data samples, particularly when it states “[s]ome embodiments of the present invention provide a media measurement system and method that enhances recognition (e.g. in terms of accuracy or efficiency) of the content of a **media** sample by analyzing information exogenous to the sample.” Ex. 1011, 578 (¶ [0006]). A POSITA would understand that “media” refers to both audio and video data samples and the immediately preceding passage from the “BACKGROUND” expressly teaches that the invention applies to content recognition systems based on audio and video data samples.

117. Petitioner does not appear to address or discuss the “SUMMARY” of the common written description. However, this particular disclosure conveys that the invention was not intended to be limited to audio samples. If the inventors had intended the invention to be limited to audio samples, the first sentence of the “SUMMARY” would refer to “audio samples,” not “media sample[s].” The use of the term “media” rather than “audio” or “video” indicates that the invention applies to both types of data samples. A POSITA would thus understand the “SUMMARY” to convey possession of a video-data embodiment. This would be even more apparent to a POSITA when viewed in the context of the immediately preceding passage from the “BACKGROUND” teaching that the invention applies to “existing” robust content recognition technologies that analyze “audio or video data sample[s].”

118. Petitioner appears to suggest that the number of times the written description refers to “audio” as compared to “video” evidences that the inventors did not possess the claimed invention. ’896 Pet., 26 / ’849 Pet., 26 / ’849 Pet., 22. While “audio” is certainly discussed more than “video” in the common written description, a POSITA would still readily recognize that the “BACKGROUND OF THE INVENTION” expressly discusses the invention as applicable to both audio- and video-sample-based content recognition technologies and that the “SUMMARY” teaches that the invention applies to “media samples,” which includes both audio and video data samples. Consequently, a POSITA would understand the common written description to clearly teach that the invention is applicable to both audio- and video-sample-based content recognition systems. A POSITA would not understand additional discussion of the invention as applied to video-samples to be required to demonstrate possession, because the invention is agnostic as to whether the content identification results are generated based on audio or video data samples. Whether the underlying samples are audio data or video data samples is incidental to carrying out or understanding the claimed invention because the novelty relates to what the invention does *with* those results rather than merely how, or from what media data-type, those results are generated.

119. Petitioner contends that a POSITA would not have recognized the inventors to have possessed the invention as applied to video data samples because

there are alleged “obstacles,” “challenges,” and “substantive differences” between audio- and video-sample-based content recognition and the “DETAILED DESCRIPTION” does not provide a detailed example of the disclosed system as applied to “video data samples.” ’896 Pet., 20-21, 28 / ’849 Pet., 20-21, 28-29 / ’848 Pet., 19, 23-24. Petitioner is again mistaken. There is simply no question that audio- and video-sample-based content recognitions were well-known and understood, as I extensively demonstrate in Section VII.A-B. In addition to the common written description stating that these technologies “existed” and were known, I have provided dozens of examples further evidencing this fact.

120. Because audio- and video-sample-based content recognition technologies were well-known and understood, a POSITA would not need an explanation or detail as to *how* to generate content identification results based on video rather than audio samples. Nor would a POSITA require or demand that the inventors explain how video data sampling works to determine that they possessed the invention as applied to video-sample-based content recognition systems.

121. I note that the common written description does not even explain how audio data samples are used to generate content identification results. Instead, the written description simply refers to prior art references showing that both audio- and video-fingerprint matching techniques and algorithms were known in the art. Ex. 1011, 584-585 (¶[0025]). A POSITA would understand that the reason the common

written description does not disclose how to perform audio- or video-sample-based content recognition, is because such technologies were well within the knowledge of a POSITA and the written description appropriately assumes as much. As such, the express statements that the disclosed system “appl[ies] **existing** content recognition technologies” based on an “audio or **video** data sample” (Ex. 1011, 578 (¶ [0005])) and “*enhances* recognition (e.g., in terms of accuracy or efficiency) of the content of a *media* sample by analyzing information exogenous to the sample” (Ex. 1011, 578 (¶ [0006])), clearly conveys that the inventors possessed the invention as applied to both audio and video data samples. Ex. 2001 ¶XX.

122. There seems to be a misplaced suggestion by Petitioner that video ACR is categorically more difficult to implement than audio ACR. This could not be further from the truth. I have implemented both audio and video ACR systems and I am intimately familiar with the math and algorithms required for each. As Neumeier’s video ACR patent is cited, I will start with that as an example. It utilizes relatively simple math to average the value of 10x10 groups of pixels into average red, average green, and average blue values for each of 25 locations within the video frame. The resulting average pixel values in the 25 pixel locations within a video frame form a set of 75 values per video frame. This set is known in linear algebra as a vector and called a “fingerprint” in ACR terminology.

123. There may not be a simpler process for creating video fingerprints, whereas the math required to convert audio to an equivalent vector (fingerprint) may require many complex signal processing steps. As an example, audio may be converted to monaural and then down sampled to, for example, 8kHz. It is then segmented into, for example, 100 millisecond blocks of audio samples. The blocks of audio data samples may then be shaped with a gaussian windowing function and then each 100msec block of audio may be overlapped by 50% with the previous block before being applied to an auto-correlation process which results in a polynomial output with a certain number of coefficients, in this case, 35. The output of the polynomial may be applied to the Levinson-Durban algorithm to stabilize the coefficients which are then processed through a normalization process of immittance spectral frequencies (ISF) which produces an all-zero output (with respect to the z-axis of a unit circle, a well-known step in signal processing.)

124. The resulting 35 values of the polynomial are now applied as values of a 35th-order vector like the 75th-order vector of the equivalent video ACR process. Both audio and video vectors are fingerprints that allow similar searching methodologies in respective reference databases using nearest neighbor (Euclidean distance) search means to search within the known media content to attempt to identify the unknown media.

125. An additional potential complexity of the audio ACR process is in the search process of a reference database of tagged audio ACR fingerprints. For example, television video is commonly transmitted as 30 discrete frames per second providing discrete frame boundaries when searching a reference database. However, audio is commonly sent as a continuous stream, making matching to reference data potentially more difficult as there is no defined starting point of the audio frame in the reference database nor from the client TV outputting fingerprints used for lookup. It is beyond the scope of this brief discussion of audio versus video ACR complexities to discuss search methodologies, but it is known to the POSITA this also adds to the complexity of audio ACR systems. Consequently, it is incorrect to suggest that audio-sample-based content recognition was categorically more “difficult” or “complex” than video-sample-based content recognition.

126. A POSITA as of May 26, 2005 would have been very familiar with both audio- and video-sample based content recognition technologies and would have known how to implement and apply them to generate a content identifier. A POSITA did not need any guidance or express teachings in the specification to understand how to generate an underlying content identification result based on a video data sample using known ACR technologies, just as a POSITA did not need any guidance or express teachings in the specification to understand how to generate an underlying content identification result based on an audio data sample using known ACR

technologies. As previously noted, the common written description does not explain how audio-sample-based content recognition is used to generate content identification results, because such technologies were known in the art just as video-sample-based technologies were.

127. Petitioner is also mistaken that “a POSITA would not have understood that the ’330 application’s disclosures of embodiments relying on audio data would have been interchangeable with embodiments that instead use video data.” ’896 Pet., 28 / ’849 Pet., 28-29 / ’848 Pet., 24. The disclosed invention is agnostic as to whether the content identification results are generated based on audio or video data samples, as explained in great detail in Section VII.D and VII.E. A POSITA would understand that the novel features disclosed apply in the same way and provide the same benefit, regardless of whether those identification results are audio- or video-sample-based. Indeed, numerous prior art references demonstrate that the applicability of inventions based on audio-sample based content recognition often applies equally to video-sample-based content recognition. *Supra* Section VII.B.

128. Petitioner is also mistaken that a POSITA would not understand how to apply the disclosures of the common written description to video data samples because “capturing audio would operate in different environments and be subject to different obstructions, challenges, and limitations to a system that relies on capturing video.” ’896 Pet., 28 / ’849 Pet., 29 / ’848 Pet., 19. The disclosed features rely on

existing audio- and video-sample based content recognition technologies, and capturing and matching audio and video samples was well-known and understood in the art using such technologies. In other words, these “different environments” and “different obstructions, challenges and limitations” described by Petitioner were already addressed by those existing systems. As I explained before, the “BACKGROUND” and “SUMMARY” make it clear that the inventors are not inventing a novel way of audio- or video-based fingerprint matching but are instead applying those existing technologies to generate content identification results that are then used as the basis for the invention’s novel combination of features.

129. Even assuming “different obstructions, challenges and limitations” between audio- and video-sample-based content recognition systems, they have no effect on how the invention applies to content identification results generated by either type of system.

130. Petitioner appears to suggest that video-sample-based content recognition was not well-known or understood. ’896 Pet. 20-21 (stating that “government-funded efforts at the time demonstrated around 10-30% accuracy rate for video content recognition) (citing Ex. 1030-1032). The only evidence cited to support this statement is a TRECVID-sponsored *object recognition* competition. The TRECVID-sponsored *object recognition* competition was a government-sponsored event to test the ability of software systems to recognize specific attributes

in video content. The first phase of the competition tested for shot changes in video content where a shot is a single camera roll or segment of video also known as a scene. Ex. 1030, 3-4. In video ACR systems, scene changes are irrelevant to the tasks of an ACR system. Without specialized software search analysis, shot changes are not detectable in any video ACR system that I am aware of.

131. Another challenge of the TRECVID competition was detecting storyline changes, such as during a news program where one story may be of an automobile accident then followed by a report on flooding from a recent storm. Ex. 1030, 6-7. The goal would be to detect the change in the storyline from automobiles to weather. Yet another challenge was detecting semantic changes in scenes such as outdoor versus indoor video scenes or scenes of people versus inanimate objects. It should be observed that both video as well as audio ACR systems are typically only matching fingerprints where the content of fingerprints cannot be identified, that is, there is no means to “playback” video or audio fingerprints and see or hear the original content from which they were derived. There is not enough information left in fingerprints to decode them, rather the fingerprints are highly compressed representations of the original content containing only enough unique patterns to match each other in short segments of fingerprints.

132. Video object recognition, as described above, is fundamentally different than content recognition based on video sample matching, which is simply

generating a video sample from a video signal and comparing it to a database of reference video samples associated with known content. The dozens of references discussed in Section VII.B, as well as the common written description expressly discussing references that describe video-sample-based content recognition, evidence that video-sample-based content recognition was well-known and understood.

133. Petitioner also suggests that a POSITA would not have understood how the invention could be implemented based on a video-sample-based content recognition system because the common written description is alleged to be silent as to “how the disclosed audio-based embodiment would have been implemented in the context of video in light of any unique challenges posed by video data.” ’896 Pet., 21 / ’849 Pet., 21 / ’848 Pet., 19. It is unclear what “unique challenges” the Petitioner is referring to, as it does not discuss or identify what them. In fact, it seems that the Petitioner is confused as to the fundamentals of how video ACR functions as evidence by the attempt to compare video ACR to the TRECVID competition, which is clearly an “apples to oranges” mistake.

134. In any event, a POSITA would readily understand how to apply the disclosed invention to video-samples using existing content recognition technologies, and further that the invention would apply in the same way as it does to results generated based on audio data samples.

135. Petitioner also contends that a POSITA would not have understood the inventors to have possessed the invention as applied to video-data samples because “multiple disclosures in the ’330 application describe peculiarities related to audio content like ‘mut[ing] a broadcast,’ ‘receiv[ing] a wireless or non-wireless audio signal,’ ‘captur[ing] audio energy,’ and ‘content types (e.g. speaking audio versus music audio),’ that have no clear analog in the video context.” ’896 Pet. 28 / ’849 Pet., 29 / ’848 Pet., 24. Petitioner is again mistaken, because a POSITA would understand that the vast majority of the other disclosures in the “DETAILED DESCRIPTION” apply equally to content identification results generated based on video-sample-based content recognition. Stated another way, a POSITA would understand that the fact that the common written description provides additional detail to account for issues specific to the “audio context” says nothing about whether the common written description also sufficiently demonstrates how the disclosed invention applies in the “video context.”

136. Petitioner selectively identifies these “peculiarities” from the “DETAILED DESCRIPTION,” but does not account for the fact that a POSITA would clearly understand and recognize the vast majority of the disclosures to be applicable to video-sample-based content recognition. For example, the disclosures from the “DETAILED DESCRIPTION” describing the “scrubbing” feature, the use of log time and content offsets to deduce play altering actions, and the use of

sequence data to identify a channel, all apply similarly regardless of whether the content identification results are generated based on audio or video data samples.

137. I understand that Petitioner asserts that new subject matter was added to the specification of the '896 Patent (the '204 application). '896 Pet., 36-38 / '849 Pet., 35-38 / '848 Pet., 30-32. Specifically, Petitioner contends that the following passage is new subject matter because it conveys that the invention may be applied to video data samples:

In alternative embodiments to what is illustrated and described in FIGS. 1-6, a record is created of individual media usage over time by means of an open-ended measurement system that employs portable and stationary detection devices to generate digital fingerprints of the video signal portion of the media content itself and identifies media played in the environment by means of matching such fingerprints against a database of known media. Such a video-based system may be built upon direct electrical connections and optical sensors (rather than microphones). Thus, various alternatives to the exemplary methods discussed in the context of the audio data-based embodiments illustrated in FIGS. 1-6 rely instead on video data.

'896 Patent, 14:17-29. Petitioner is mistaken, because the common written description already conveys this concept. The common written description teaches that the disclosed invention applies “existing” audio- and video-sample based content recognition technologies to generate content identification results and, further, that the invention applies to “media sample[s]” (i.e., audio and video

samples). Ex. 1011, 578 (¶¶[0005], [0006]). A POSITA would thus understand, based on the common written description, that the disclosures relating to the detailed embodiment of Figures 1-6 are applicable to content identification results generated based on video data samples.

138. A POSITA would understand that passage added to the '896 patent specification does nothing more than elaborate or clarify what is already conveyed by the common written description. A POSITA would thus recognize that while that passage was indeed “added,” it did not add “new” matter as alleged by Petitioner.

139. I also note that even if the added passage was assumed to include some new matter, a POSITA would still understand the common written description to convey possession of the claimed inventions of the '896, '848, and '849 Patents. In other words, because the common written description conveys possession of the claimed inventions, the passage added to the '896 patent later cannot somehow undo that understanding.

IX. CONCLUSION

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

Dated: December 27, 2024



W. Leo Hoarty

APPENDIX A

Date	Exhibit	Exemplary Disclosures
1975	2005	“A process for automatic electronic recognition and identification of programs and commercial advertisements broadcast on television and radio wherein a digitally sampled reference signal segment derived from either the audio or video portion of the original program content to be identified is compared with successive digitally sampled segments of the corresponding audio or video portion of a broadcast signal in a correlation process to produce a correlation function signal. . . .” Abstract. “This invention utilizes a process for completely automatic electronic detection and identification of recorded commercial advertisements, or other specialized program content, in television or radio broadcasts. The addition of codes or other alterations to the broadcast material is completely unnecessary. In accordance with the invention a pattern recognition technique is utilized which technique permits an audio or a video signal which constitutes the program or commercial advertisement which is to be identified to effectively act as its own code. That is, a segment of the audio signal, for example, of each program or commercial advertisement which it is desired to recognize is stored in a suitable memory storage device to form a unique set of stored waveform patterns (i.e., in effect, a set of fingerprints or signatures, one for each of the commercial advertisements that are involved) against which all future monitored broadcast information is compared.” 2:57-3:17.
1979	2006	“As an alternative to adding codes to the program material, the prior art includes a method of program identification whereby the audio or video program signal is analyzed in such a way that the program supplies its own unique code. Method B, therefore includes all attempts to solve the problem of automated

Date	Exhibit	Exemplary Disclosures
		broadcast monitoring solely through the use of computerized pattern recognition.” 3:47-54. “It should be noted at this point that the term "program signal" is used throughout this application to designate either the composite video signal or the audio signal, or even both, where appropriate. (These components of the program signal are not shown in FIG. 1.) Thus, the raw program signal 8 can include a video component 7 and/or an audio component 9; similarly, the final program signal 16 can refer to either a composite video signal 15 or an audio signal 17 (or both) which include cue signals.” 9:53-62.
1984	2007	“A method and apparatus for identifying programs such as television programs received from various sources detects the occurrence of predetermined events such as scene changes in a video signal and extracts a signature from the video signal. The signatures and the times of occurrence of the signatures are stored and subsequently compared with reference signatures to identify the program. The signatures may be compared in pairs to increase resolution, and the time interval between events or signatures may also be used to identify the program, either by themselves or in conjunction with the signatures.” Abstract “Therefore, in accordance with a preferred embodiment of the invention, there is provided a system having a home unit that monitors the viewing conditions in the home, for example, whether the television set is on or off, whether any home video recorder (e.g., a video cassette recorder or VCR) is on or off and the video signal being received. By monitoring the aforementioned signals and functions, the system determines whether no viewing is occurring, whether a television broadcast is being viewed, whether a

Date	Exhibit	Exemplary Disclosures
		television broadcast is being recorded, whether viewing and recording is occurring simultaneously, or whether a previously recorded program is being viewed. Once the viewing mode has been established, the program being viewed is identified by extracting a characteristic signature from the video signal and storing it for subsequent comparison with reference signatures stored in a central office.” 3:1-17. Referring now to the drawing, with particular attention to FIG. 1, there is illustrated a block diagram of the home unit of the system according to the invention generally designated by the reference numeral 10. The system 10 receives signals to be identified from a television receiver 12, and in many instances from a home video recorder, such as a video cassette recorder or a VCR 14. In a typical home installation, the receiver 12 and the VCR 14 receive signal from a receiving antenna 16, and in many instances from a cable television system 18. These signals may be applied directly to the receiver 12 and to the VCR 14, but it is convenient to employ a switching network 20 that selectively switches the receiver 12 and VCR 14 to the antenna 16 and to the cable 18 in order to permit the viewing and recording of broadcasts received from the antenna 16 or from the cable 18. In addition, the switching network 20 may connect the VCR 14 to the receiver 12 to permit prerecorded or previously recorded tapes to be played back by the VCR and viewed by the receiver 12. Video signals representative of the program applied to the receiver 12, as well as video signals representative of the signals being recorded by the VCR 14, are applied to a video processing circuit 22, for example, by a video switching circuit 24 which permits either the receiver 12 or the VCR 14, or both, to be monitored. The video processing circuit 22 includes a detector and a low pass

Date	Exhibit	Exemplary Disclosures
		filter (discussed subsequently in greater detail), and provides a signal representative of the envelope of the video signal to an analog-to-digital converter 25 which generates a digital representation of the video envelope. The digitized video envelope signal is applied to an events detector 26 which detects predetermined events that occur in the digitized video signal, and causes the signature of the digitized video signal to be extracted by signature extraction circuitry 28 upon the occurrence of a predetermined sequence of events. The signature thus generated is stored in a data storage system 30 along with the time that the signature was extracted. The extraction times are provided by a clock 32 which provides a digital representation of time to the data storage system 30 via, for example, a combining circuit 34. The clock 32 may be a real time clock or a relative time clock that is periodically reset from a central location. The function of the events detector 26, the signature extractor circuitry 28, the data storage 30 and the clock 32 can be performed by individual circuits as shown or by a microprocessor based system. The home unit 10 is interrogated at periodic intervals, for example, once a day, by a central office unit, generally designated by the reference numeral 40 (FIG. 2), which compares the signatures from the various home units with reference signatures in order to identify the signatures from the home unit.” 4:27-5:12 “Alternatively, the audio signal may be monitored by an audio processing circuit 38 and applied to the analog-to-digital converter 25 to generate an Event 2 when the audio signal departs from a predetermined level.” 10:29-35
1985	2008	“A method and apparatus for identifying repetitively broadcast programs, such as commercial advertisements broadcast on television and radio, detects the occurrence of predetermined events in a

Date	Exhibit	Exemplary Disclosures
		video signal and/or in an audio signal, detects a stability condition and extracts a signature from the video and/or audio signal. The signatures and the times of occurrence of the signatures are stored and subsequently compared with reference signatures to identify the program and its time of broadcast. The reference signatures may be obtained by identifying repetitively extracted, unidentified signatures and indexing a recording of the video and/or audio signals to enable the system user to identify the unidentified broadcast program and store the previously unidentified signatures as additional reference signatures in memory.” Abstract. “Referring now to the drawing, with particular attention to FIG. 1, there is illustrated a block diagram of a station monitoring unit of the system according to the invention designated generally by the reference character 10. The station monitoring unit 10 receives television broadcast signals from a receiving antenna 12. While the station monitoring unit 10 is depicted and generally described herein for monitoring a television broadcast signal which includes both video and audio signals, the principles of the present invention disclosed herein are equally applicable to a broadcast radio signal. The received signals are applied to a receiver 14 which includes a tuner and amplifier. The receiver 14 is tuned to a desired broadcasting station to be monitored to provide a desired program signal 16 at the output of receiver 14. The program signal 16 is applied to audio processing circuitry 18, video processing circuitry 20 and a recorder 22, such as a video cassette recorder or a VCR 22. The program signal 16 is recorded on a video tape 24 within the VCR 22. The audio processing circuitry 18 includes an energy detector and a plurality of bandpass filters, and provides a detected energy signal 25 and, a plurality of signals 26 representative of an audio spectrum that may be, for

Date	Exhibit	Exemplary Disclosures
		example, in a range between 330 Hz and 2.54 kHz. An ASA-16 speech preprocessor integrated circuit manufactured by Interstate Electronics Corporation of Orange, Calif., may be employed to provide the bandpass filtering of audio processing circuitry 18. The audio spectrum signals 26, for example, 12 bandpass signals, are applied to a multiplexer 28 which provides the analog spectrum signals to an analog-to-digital (A/D) converter 30. The A/D converter 30 generates digital representations of the audio spectrum signals or digitized audio signals 32. The digitized audio signals 32 are applied to both an audio signature extraction circuit 34 and to an events detector 38. The detected energy signal 25 is applied to the events detector 38. The events detector 38 has an output that is coupled to an enable input 36 of the audio signature extraction circuitry 34. The audio signature extraction circuitry 34 extracts an audio signature from the digitized audio signals 32 when enabled by the events detector 38 which is discussed subsequently in greater detail. The video processing circuitry 20 includes a detector and a low-pass filter having, for example, a cut-off frequency of approximately 0 to 2 kHz, and provides a signal representative of the envelope of the video signal to an analog-to-digital (A/D) converter 40 which generates a digital representation of the video envelope. The digitized video envelope signal is applied to both a video signature extraction circuit 42 and the events detector 38 which detects predetermined events that occur in the digitized video signal, and causes the signature of the digitized video signal to be extracted by video signature extraction circuitry 42. The signature thus generated is stored in a data storage device or memory 44 along with a time that the signature was extracted. The extraction times are provided by a clock 46 which provides a digital representation of time to the data storage device 44 via, for example, a combining circuit 48.

Date	Exhibit	Exemplary Disclosures
		The clock 46 may be a real time clock or a relative time clock that is periodically reset from an accurate clock, for example, a timing signal transmitted by the National Bureau of Standards. The clock 46 provides the digital representation of time to a video date and time code generator 50 which applies the digital representation of time to the VCR 22 for storage on the video tape 24. The function of the events detector 38, the audio signature extraction circuitry 34, the video signature extraction circuitry 2, the data storage 44 and the clock 46 can be per2, formed by individual circuits as shown or by a microcomputer, a personal computer, or various microprocessor based systems. Referring to FIG. 2, there is shown a block diagram illustrating a signature recognition and postprocessing unit designated generally by the reference character 54. The station monitor unit 10 (FIG. 1), extracts and stores signatures and times of extraction thereof, and applies the signatures to the signature recognition and post-processing unit 54 which compares the extracted signatures from the various station monitoring units 10 with a plurality of reference signatures taken from identified programs in order to identify particular broadcast programs corresponding to the extracted signatures received from the station monitoring units 10. The station monitoring units 10 may be interrogated at periodic intervals by a data communication circuit 56 that accesses the various station monitoring units 10 via a suitable communications system, for example an RS-232 Serial communication line, a common computer bus or a plurality of telephone lines 58. The extracted signatures thus collected are compared by a signature recognition computer 60 to a plurality of reference signatures stored in a memory or reference signature data base 62. The signature recognition computer 60 performs the collection and classification of the signatures received from the station

Date	Exhibit	Exemplary Disclosures
		monitoring units 10, as well as the generation of reference signatures to be stored in the reference signature data base 62. The signature recognition computer 60 compares the extracted signatures to the reference signatures stored in data base 62 and classifies and stores the signatures in a recognition file 64 which is coupled to a post-processing computer 66. The signature recognition computer 60 utilizes successive recognized extracted signatures to identify a particular program. The recognized extracted signatures from the various station monitoring units 10 are applied to the post-processing computer 66. The post-processing computer 66 produces a file or log 70 of the identified programs, commercial advertisements or other repetitively broadcast programs from each monitored station and can include the starting time of each identified or recognized program and the duration of the program. In addition, such analytical operations can provide various logical checks to flag incomplete or otherwise defective broadcasts of repetitively broadcast programs, and thereby enable the detection of possible false signal recognitions.” 3:37-5:19
1986	2009	“The present invention can identify a segment of a broadcast signal by pure continuous pattern recognition on a real-time basis. No codes need be inserted into the broadcast signal prior to transmission and no ‘cues’ or ‘signalling [sic] events’ (e.g., fades-to-black or scene changes) that occur in the broadcast need be used in the recognition process. The broadcast signal can be either a radio broadcast, for which audio information is obviously used for the recognition process, or a television broadcast, for which audio information, video information, or both, can be used, and can be an over-the-air signal, a cable television signal, or a videotape signal.” 3:35-55.

Date	Exhibit	Exemplary Disclosures
1988	2010	“The disclosure relates to an apparatus and method for creating digital signatures from frames of selected video segments such as TV commercials by a digitizing, compression and selection process which produces a multi-digit signature from each frame. The signatures are stored in a memory database. In the recognition mode broadcast video transmissions are similarly digitized frame by frame to generate signatures which are then compared with the signatures in the database to find matches.” Abstract. “Briefly, the system of the invention operates by forming a library or database of commercials which can be searched with respect to a real time TV commercial being received at any time. The system is designed to monitor constantly, and without need for operator intervention. The library or database is prepared as follows: 1. Each pixel (pel) of each frame of a recorded commercial is digitized to form an 8-bit word representing the pixel's luminosity (8-bit quantization yields 256 possible levels) and each TV frame or digital words is stored in a frame memory (buffer) 2. Each frame of digitized pixels is arrayed into 64 super pixels, each super pixel being made-up of a group of 16×16 (256) pixels, the super pixel is simply an average of all 256 pixels. 3. The 64 super pixels per frame are compared with each other according to a pattern which produces a 32-bit signature word or template. 4. Selected signatures of each commercial are stored in a memory which is the database system. The recognition process, briefly, comprises forming signatures of every frame of a real time video or other TV transmission, and comparing each such signature with the entire library (database) of stored signature to find

Date	Exhibit	Exemplary Disclosures
		matching signatures. When signatures are matched and a commercial is identified, the name of the commercial, the data and time, the broadcast channel and other pertinent desired data are presented.” 4:14-55.
1992	2011	“Broadcast segment recognition systems and methods are provided in which a signature representing a monitored broadcast segment is compared with broadcast segment signatures in a data base representing known broadcast segments to determine whether a match exists. Criteria for determining the validity of such a match are provided.” Abstract “Each of the RF broadcast converters 24 receives television broadcast signals over a respective channel and demodulates the received signals to provide baseband video and audio signals. The video and audio signals are thereafter supplied to the segment recognition subsystem 26, wherein frame signatures for each of the video and audio signals are generated which are thereafter compared to stored key signatures to determine if a match exists. For purposes of clarity, video and audio signatures are separately termed "subsignatures" herein. The segment recognition subsystem also produces cues which represent signal events, such as a video fade-to-black or an audio mute. The cues as well as match information are supplied to the control computer 30 for use in determining whether the received signal represents a new segment or commercial of interest, determining whether to capture video and audio information for use at the central site in identifying a new segment of interest, assessing the validity of questionable matches, and for grouping match information for storage in a database.” 10:12-31.
1993	2012	“Audio processing circuitry 28 includes an energy detector and a plurality of band pass filters (not shown) and

Date	Exhibit	Exemplary Disclosures
		provides a detected energy signal 35 and a plurality of audio signals 36 each representing the energy of signal 26 within a specified frequency band. Audio spectrum signals 36 are applied to the multiplexer 38, which connects each signal 26 in turn to the A/D converter 40, thereby producing the digitized multiplexed audio signals 42. Digitized audio signals 42 are applied to both an audio signature extraction circuit 44 and to an events detector 48. Events detector 48 produces an output signal 46 that enables audio signature extraction circuit 44 to extract an audio signature from digitized audio signals 32. Video processing circuitry 30 provides a signal representative of the video signal envelope to an A/D converter 50, which generates a digital representation of the video envelope. The digitized video envelope signal is applied to both a video signature extraction circuit 52 and events detector 48, which causes the signature of the digitized video signal to be extracted by video signature extraction circuitry 52 responsive to the detection of predetermined events that occur in the digitized video signal. The signatures from audio signature extraction circuit 44 and video signature extraction circuit 52 are stored in a data storage device 54 together with the corresponding extraction time and date. Extraction times are provided by a clock 56 that provides a digital representation of time to data storage device 54 by way of a combining circuit 58. Clock 56 also provides the digital representation of time to a video date-time code generator 60 that creates a digital representation of time for storage on medium 34 by recorder 32. Station monitor unit 20 extracts and stores signatures and times of extraction thereof. At some convenient time (not in real-time), the stored signatures are applied to the signature and recognition post-processing unit 64, which compares the extracted signatures from the various station monitoring units 20 with a plurality of reference signatures taken from programs to identify

Date	Exhibit	Exemplary Disclosures
		particular corresponding broadcast programs. Several station monitoring units 20 may be interrogated at periodic intervals by the data communications circuit 66, which accepts the stored signatures from station monitoring units 20 by way of several communication lines 68. The extracted signatures are compared by a signature recognition computer 70 to a library of reference signatures stored in a reference signature database memory 72. Signature recognition computer 70 collects and classifies signatures received from several station monitoring units 20 and also generates the reference signatures stored in reference database memory 72. Computer 70 compares the extracted signatures to the reference signatures from database memory 72 and classifies and stores them in a recognition file 74, which is coupled to a post-processing computer 76. Signature recognition computer 70 searches for instances of successive recognition of extracted signatures to identify a particular program. These successive recognized extracted signatures are provided to post-processing computer 76, which produces a file or log 80 containing records of the identified programs, commercial advertisements or other repetitively broadcast programs from each monitored broadcast station together with a starting time and the program duration. Post-processing computer 76 also provides extensive analytical services to logically check and flag incomplete or otherwise defective broadcast identifications and thereby enable user detection of false signal recognitions. Post-processing computer 76 also assigns a temporary code to identify each non-recognized or unmatched extracted signature to permit later manual review of the program recorded in medium 34 for manual identification. A data entry terminal 78 is provided for updating reference signature database 72.” 5:9-6:3
1997	2013	“As shown in FIG. 3, three of the non-invasive signal

Date	Exhibit	Exemplary Disclosures
		sensors 30' are employed to acquire representations of various signal components from the television receiver 22' whenever it is turned on. These three non-invasive signal sensors may include (i) a suitable sync signal pickup 60 which may be of the type disclosed in U.S. Pat. No. 5,889,548, which is disposed adjacent the television receiver 22', and which picks up a representation of the vertical and horizontal synchronization pulses of the television receiver 22', (ii) a video signal pickup 62 which may be of the type disclosed in the aforementioned U.S. Pat. No. 5,889,548 and which is arranged to pick up a representation of the video signal being displayed by the television receiver 22', and (iii) an audio signal pickup 64, which may be a microphone as taught by Thomas et al. in U.S. Pat. No. 5,481,294 and which is used to acquire the audio portion of the displayed program signal for subsequent matching operations.” 10:27-43. “It will be understood to those skilled in the art that both the video signal pickup 62 and the audio signal pickup 64 may employ appropriate circuitry (not separately shown) to match their output levels to the industry standard output levels of the signals obtained from the VCR video output 58 and the VCR audio output 56 against which they are matched.” 10:61-67.
1998	2014	“According to the present invention there is provided a method of detecting video sequences comprising the steps of receiving an incoming video stream consisting of sequences of successive frames, creating on-the-fly digital signatures in accordance with a predetermined algorithm, comparing said digital signatures with a plurality of stored digital signatures created in accordance with said predetermined algorithm, and identifying a candidate sequence of said incoming video

Date	Exhibit	Exemplary Disclosures
		stream as corresponding to said stored sequence in the event of a positive match of said digital signatures in accordance with predetermined criteria...” 1:38-48. “The video sequence is in digital form, either as a result of analog-to-digital conversion of an analog broadcast signal or the signal being in digital format.” 1:58-60.
2001	2016	“A method for recognizing an audio sample locates an audio file that most closely matches the audio sample from a database indexing a large set of original recordings. Each indexed audio file is represented in the database index by a set of landmark timepoints and associated fingerprints. Landmarks occur at reproducible locations within the file, while fingerprints represent features of the signal at or near the landmark timepoints. To perform recognition, landmarks and fingerprints are computed for the unknown sample and used to retrieve matching fingerprints from the database.” Abstract. “The present invention provides a method for recognizing an exogenous media sample given a database containing a large number of known media files. It also provides a method for generating a database index that allows efficient searching using the recognition method of the invention. While the following discussion refers primarily to audio data, it is to be understood that the method of the present invention can be applied to any type of media samples and media files, including, but not limited to, text, audio, video, image, and any multimedia combinations of individual media types.” 5:36-46.
2001	2017	“A more dynamic signature deriving process is discussed with respect to FIG. 1. With reference to FIG. 1, an input signal is segmented in step 20. The signal may be an audio, video, or image signal, and may be divided into sets such as

Date	Exhibit	Exemplary Disclosures
		segments, frames, or blocks, respectively.” 3:59-63. “FIG. 4 is a flow diagram illustrating a method to determine an identification of an unknown content item. In step 50, a signal set (e.g., image block, video frame, or audio segment) is input into a system, e.g., a general-purpose computer programmed to determine signatures of content items.” 6:14-19. “The following discussion details another method, which uses Trellis Coded Quantization (TCQ), to derive a content signature from a content item. Whereas the following discussion uses an image for an example, it will be appreciated by one of ordinary skill in the art that the concepts detailed below can be readily applied to other content items, such as audio, video, etc.” 11:14-19.
2001	2018	“In accordance with a further aspect of the present invention, a system for measuring the use of a video recorder operably connected to a television receiver comprises source tuning means, television receiver signal acquiring means, and first and second comparing means. The television receiver receives a plurality of transmitted television signals. The source tuning means tunes to a source program signal corresponding to a transmitted television signal. The television receiver signal acquiring means acquires a signal displayed on the television receiver. The first comparing means compares a signal from the video recorder to the source program signal. The second comparing means compares the signal from the video recorder to the signal acquired by the television receiver signal acquiring means.” [0027] “As shown in FIG. 3, three of the non-invasive signal

Date	Exhibit	Exemplary Disclosures
		sensors 30' are employed to acquire representations of various signal components from the television receiver 22' whenever it is turned on. These three non-invasive signal sensors may include (i) a suitable sync signal pickup 60 which may be of the type disclosed in copending application Ser. No. 08/654,309 filed on May 28, 1996, which is disposed adjacent the television receiver 22', and which picks up a representation of the vertical and horizontal synchronization pulses of the television receiver 22', (ii) a video signal pickup 62 which may be of the type disclosed in the aforementioned copending application Ser. No. 08/654,309 filed on May 28, 1996 and which is arranged to pick up a representation of the video signal being displayed by the television receiver 22', and (iii) an audio signal pickup 64, which may be a microphone as taught by Thomas et al. in U.S. Pat. No. 5,481,294 and which is used to acquire the audio portion of the displayed program signal for subsequent matching operations.” [0049]
2002	2019	“FIG. 1 is a diagram of a system configuration of linked media objects. In this configuration, an identifier links audio objects to metadata via an electronic network, such as the Internet, a wireless network, or a broadcast network. As depicted in FIG. 1, an embedding process may be used to encode an identifier in an audio object or its container. In some cases, an embedding process encodes the identifier in the audio file (e.g., a tag in a file header or footer), in the audio signal (a digital watermark), or in the physical packaging. The identifier may also be derived as a function of the audio signal or other information in the file or physical packaging (e.g., track information on a CD).” 4:2-13. “While specifically discussed in the context of audio objects, the fingerprinting process applies to other types of multimedia content as well, including still images,

Date	Exhibit	Exemplary Disclosures
		video, graphics models, etc. For still images and video, the identifier can be derived dynamically from a compressed or uncompressed version of the image or video signal. The fingerprinting process may be tuned to generate a specific identifier based on the type of file format. For example, the process extracts the file format from the file (e.g., from a header or footer), and then uses a fingerprinting process tailored for that type of file (e.g., a hash of a compressed image or video frame).” 10:22-33.
2002	2020	“It is a general object of the invention to provide a robust hashing technology. More particularly, it is a first object of the invention to provide a method and arrangement for extracting a limited number of hashing bits from multimedia content.” 2:9-11. “Robust hashing of audio signals will be described first. The audio signal will be assumed to be mono audio that has been sampled at a sample frequency of 44.1 kHz (CD-quality). If the audio is stereo, there are two options: either hash signals are extracted for the left and the right channel separately, or the left and the right channel are added prior to hash signal extraction.” 4:24-28. “Even if we only have a short piece of audio (of the order of seconds), we would like to determine which song it is. As audio can be seen as an endless stream of audio-samples, it is necessary to subdivide audio signals into time intervals or frames and to calculate a hash word for every frame.” 4:29-32. “Robust hashing of image or video signals will now be described. Again, the robust hashes are derived from specific features of the information signal. The first

Date	Exhibit	Exemplary Disclosures
		question to be asked is in which domain to extract said features which determine the hash word.” 6:31-33.
2002	2021	“ The use of digital video recorders (DVRs) and personal video recorders (PVRs) such as the TiVo™ system further complicates collection of viewing behavior information because viewers in households with these types of recording devices can rapidly change between live viewing of a program, a somewhat delayed viewing of a program, fast forwarding and rewinding a program, pausing a program, and recording a program for later viewing while watching another program live. The use of processing techniques based on recognition of program signatures has been employed in some cases where it is difficult to obtain accurate tuning information and in cases where it may otherwise be difficult to obtain viewing time information and channel information associated with the currently viewed program content.” [0007]. Fig. 2 is a more detailed block diagram of an example of the digital information monitor/analyzer 14 shown in Fig. 1. In the example of Fig. 2, the digital information monitor/analyzer 14 includes a time stamp generator 52, a realtime clock 54, an event record generator 56, and an access detector 58. The access detector 58 has a read detector 60 and a write detector 62 associated therewith. The event record generator 56 receives a source input 63, which may, for example, be one of the attachments 16 shown in Fig. 1. Of course, the source input 63 does not necessarily have to be provided by one of the attachment 16. For example, the source input 63 may instead be provided by apparatus and techniques that detect a source by analyzing or decoding audio content, video content, etc. In particular, signature detection systems and methods (many of which are well known), systems and methods that insert a tag or other identifier (which

Date	Exhibit	Exemplary Disclosures
		are also known), etc. may be used instead. [0035]
2003	2022	“The invention can be summarized as follows. Fingerprints are bit strings extracted from a media signal (e.g. an audio or video clip) to identify said media signal. Typically, they are derived from a perceptual property of the signal, for example, the spectral energy distribution of an audio fragment or the luminance distribution of a video image. A method and arrangement for extracting a fingerprint is here disclosed which is robust with respect to shifts of the perceptual property.” [0020] “Although embodiments of the method and arrangement have been described with reference to audio fingerprint extraction, the invention is not restricted thereto. Applicant’s International Patent Application WO 02/065782, already cited above, discloses a video fingerprint extracting method in which the fingerprint is derived from the mean luminance values of image blocks into which each image is divided.” [0019]
2003	2023	“Fingerprints, in the literature sometimes referred to as hashes or signatures, are binary sequences extracted from multimedia contents, which can be used to identify said contents. Unlike cryptographic hashes of data files (which change as soon as a single bit of the data file changes), fingerprints of multimedia contents (audio, images, video) are to a certain extent invariant to processing such as compression and D/A & A/D conversion. This is generally achieved by extracting the fingerprint from perceptually essential features of the contents.” [0002] “It is an object of the invention to provide an improved method and arrangement for extracting a fingerprint

Date	Exhibit	Exemplary Disclosures
		from multimedia contents. It is a particular object of the invention to provide a method and arrangement for extracting a fingerprint from an audio signal that is substantially invariant to speed changes of the audio signal.” [0010] “Although the invention has been described with reference to audio fingerprinting, it can also be applied to other multimedia signals such as images and motion video.” [0027]
2003	2024	“Methods and apparatus are described for matching a set input fingerprint blocks, each fingerprint block representing at least a part of an information signal, with fingerprints stored in a database that identify respective information signals.” Abstract. “In many applications, identification of multimedia signals, including audio and/or video content, is desirable. However, multimedia signals can frequently be transmitted in a variety of file formats. For instance, several different file formats exist for audio files, like WAN, MP3 and Windows Media, as well as a variety of compression or quality levels. Cryptographic hashes such as MD5 are based on the binary data format, and so will provide different fingerprint values for different file formats of the same multimedia content. This makes cryptographic hashes unsuitable for summarising multimedia data, for which it is required that different quality versions of the same content yield the same hash, or at least similar hash.” 1:14-22. “Whilst the above embodiments of the present invention have been described with reference to audio information streams, it will be appreciated that the invention can be applied to other information signals, particularly multi-

Date	Exhibit	Exemplary Disclosures
		media signals, including video signals.” 11:19-22.
2003	2025	“The invention can be summarized as follows. Fingerprints are bit strings extracted from a media signal (e.g. an audio or video clip) to identify said media signal. Typically, they are derived from a perceptual property of the signal, for example, the spectral energy distribution of an audio fragment or the luminance distribution of a video image. A method and arrangement for extracting a fingerprint is here disclosed which is robust with respect to shifts of the perceptual property.” 5:3-8. “Although embodiments of the method and arrangement have been described with reference to audio fingerprint extraction, the invention is not restricted thereto. Applicant’s International Patent Application WO 02/065782, already cited above, discloses a video fingerprint extracting method in which the fingerprint is derived from the mean luminance values of image blocks into which each image is divided.” 4:25-29.
2003	2026	“It is an object of the invention to provide an improved method and arrangement for extracting a fingerprint from multimedia contents. It is a particular object of the invention to provide a method and arrangement for extracting a fingerprint from an audio signal that is substantially invariant to speed changes of the audio signal.” 2:21-24. “Although the invention has been described with reference to audio fingerprinting, it can also be applied to other multimedia signals such as images and motion video. While speed changes are often applied to audio signals,

Date	Exhibit	Exemplary Disclosures
		affine transformations such as shift, scaling and rotation, are often applied to images and video.” 6:18-21.
2003	2027	“Signature-based program identification apparatus and methods for use with digital audio and/or video systems are disclosed. The disclosed systems and methods develop signatures of digital audio and/or video content from a digital data stream at two or more sites and compare the signatures to one another to identify the content.” Abstract “Fig. 1 is a block diagram of an example system that uses signatures to identify digital audio/video content.” [0010] “Fig. 2 is flow diagram of an example manner in which the system shown in Fig. 1 may be used to identify digital audio/video content consumed at a consumption site.” [0011]
2004	2028	“A television signal substitution system that replaces known video segments such as advertisements with selected replacement advertisements. Fingerprint data of known advertisements can be stored in a fingerprint database. When new fingerprint data is available, the fingerprint data can be automatically or manually transmitted to subscribers. Various techniques can be used to identify advertisements based on the fingerprint data.” Abstract “It is an aspect of the present invention to provide an improved system for detecting segments in video programming and replacing the detected segments with a replacement portion. The above aspects can be obtained by a method that

Date	Exhibit	Exemplary Disclosures
		includes (a) receiving an input video signal; (b) automatically identifying a segment in the input video signal locally using fingerprint data of both the segment and the input video signal; and (c) generating an output video signal comprising the input video signal with the segment replaced with a replacement portion.” 2:52-62. “An input video signal 100 is input into a detection and replacement system 102. The input video signal 100 can be any type of signal, e.g., analog, digital, single channel, composite, etc. and may also include audio or other data.” 4:61-64.
2004	2029	“A method for fingerprinting a video involving identifying motion within the video and using a measure of the identified motion as a fingerprint. Once videos are fingerprinted, these fingerprints can be used in a method for identifying video. This involves creating a motion fingerprint for unidentified videos; comparing the fingerprints of the known and unknown videos, and identifying whether the unknown video is a copy of the known video based on the step of comparing.” Abstract “According to one aspect of the present invention, there is provided a method for fingerprinting video comprising identifying motion within the video and using a measure of the identified motion as a motion fingerprint. By motion fingerprint it is meant a record or file that is representative of movement as a function of time within all or part of a video. By using motion within a video as a means for providing a fingerprint, there is provided a secure and effective mechanism for identifying video that does not require the addition of data to the video and that can be applied even when the video format is changed or in

Date	Exhibit	Exemplary Disclosures
		the presence of noise.” 3:45-55. “According to another aspect of the invention, there is provided a system for fingerprinting a video, the system being operable to identify motion within the video and create a fingerprint using a measure of the identified motion.” 3:64-67.
2004	2035	“The present invention is directed to a system, method, and computer readable storage for detecting segments such as advertisements in a video signal and replacing the detected segments with replacement portions such as targeted advertisements.” [0003] “It is an aspect of the present invention to provide an improved system for detecting segments in video programming and replacing the detected segments with a replacement portion.” [0011] “The above aspects can be obtained by a method that includes (a) receiving an input video signal; (b) automatically identifying a segment in the input video signal locally using fingerprint data of both the segment and the input video signal; and (c) generating an output video signal comprising the input video signal with the segment replaced with a replacement portion.” [0012] “FIG. 1 is a block diagram illustrating a general principle of the present invention, according to an embodiment of the present invention.” [0042] “An input video signal 100 is input into a detection

Date	Exhibit	Exemplary Disclosures
		and replacement system 102. The input video signal 100 can be any type of signal, e.g., analog, digital, single channel, composite, etc. and may also include audio or other data. The detection and replacement system 102 receives the input video signal 100 and identifies advertisements (the process of identifying will be described below in more detail). When an advertisement is identified, the advertisement can be replaced with a replacement advertisement.” [0043]
2005	2030	“A media content database is created and maintained by comparing the media items from two related media streams. Segments of media items from the streams are sampled and stored. Samples from one stream are matched against samples from the other stream to identify any repeating content in the streams. A reviewer observes the matched samples representing the repeating content and provides identifying meta-data for that content. The identified media content is then added to the media content database, where it can be used at a later time in an application that requires media content recognition.” Abstract “To address the problem of existing content acquisition solutions, embodiments of the invention attempt to match content sampled from a media stream with content sampled from another media stream (which may be a different time period of the same media stream). For example, audio items from a radio broadcast can be matched against audio items from a previous period of the same broadcast. In this way, the radio broadcast itself is used as the source of data to avoid having to extract the characteristic information from a number of master recordings for each audio item.” [0012]

Date	Exhibit	Exemplary Disclosures
		“Alternatively, if the media stream 102 is a video broadcast, each media item may comprise a television show or segment, an advertisement, or any other type of video or audio/video content typical of such broadcasts.” [0026]
2005	2031	“Methods and apparatus for media source identification and time shifted media consumption measurements are disclosed. A disclosed method identifies a media source by generating first signature information based on media presented via a media delivery device and comparing the first signature information to second signature information. The second signature information is derived from a library of signature information local to the media delivery device. The disclosed method generates a collection of matching signature information based on the comparison of the first and second signature information, and identifies a source of the media presented via the media delivery device based on the collection of matching signature information.” Abstract “The following describes example methods, apparatus, and articles of manufacture that may be used to identify sources of media content and measure time shifted consumption (e.g., viewing, listening, etc.) of media content. In general, source identification refers to determining which of a plurality of media sources (e.g., a television program, an audio program, web pages, a video game, etc.) is presenting or delivering media content. The methods, apparatus, and articles of manufacture disclosed herein are particularly well-suited to identify media sources in media delivery systems in which time shifted consumption of media content may occur. In general, time shifted media consumption occurs when an individual consumer causes one or more media sources (e.g., a PVR, a DVD player, a DVR, a VCR, an STB, a game console, and/or any other devices capable of reproducing audio and/or video

Date	Exhibit	Exemplary Disclosures
		information) to present or deliver media content at a time convenient to the consumer (e.g., viewing broadcast or pre-recorded content at a time that is subsequent to the time at which the content was recorded and/or broadcast).” [0018] “An example method identifies a local media source (i.e., a PVR, a DVD player, a DVR, a VCR, an STB, a game console, and/or any other device capable of reproducing and/or presenting audio and/or video information) by generating signature information based on media (e.g., a television program, a radio program, etc.) presented via a media delivery device (e.g., a television set, a radio, etc.), and comparing the signature information to additional signature information derived from a library of signature information local to the media delivery device.” [0019]
2005	2032	“In general, in one aspect, the disclosure describes a method for detecting a known video entity within a video stream. The method includes receiving a video stream and continually creating statistical parameterized representations for windows of the video stream. The statistical parameterized representation windows are continually compared to windows of a plurality of fingerprints.” Abstract “FIG. 1 illustrates an exemplary content delivery system 100. The system 100 includes a broadcast facility 110 and receiving/presentation locations. The broadcast facility 110 transmits content to the receiving/presentation facilities and the receiving/presentation facilities receive the content and present the content to subscribers. The broadcast facility 110 may be a satellite transmission facility, a head-end, a central office or other distribution center. The broadcast facility 110 may transmit the content to the

Date	Exhibit	Exemplary Disclosures
		receiving/presentation locations via satellite 170 or via a network 180. The network 180 may be the Internet, a cable television network (e.g., hybrid fiber cable, coaxial), a switched digital video network (e.g., digital subscriber line, or fiber optic network), broadcast television network, other wired or wireless network, public network, private network, or some combination thereof.” 4:26-40. “The content delivery network 100 may deliver many different types of content. However, for ease of understanding the remainder of this disclosure will concentrate on programming and specifically video programming.” 4:50-53.
2005	2033	“In general, in one aspect, the disclosure describes a method for detecting and acting on a known video entity within a video stream. The method includes receiving a video stream and continually creating statistical parameterized representations for windows of the video stream. The statistical parameterized representation windows are continually compared to windows of a plurality of fingerprints. Each of the plurality of fingerprints includes associated statistical parameterized representations of a known video entity. A known video entity in the video stream is detected when a particular fingerprint of the plurality of fingerprints has at least a threshold level of similarity with the video stream.” Abstract. “FIG. 1 illustrates an exemplary content delivery system 100. The system 100 includes a broadcast facility 110 and receiving/presentation locations. The broadcast facility 110 transmits content to the receiving/presentation facilities and the receiving/presentation facilities receive the content and present the content to subscribers. The broadcast

Date	Exhibit	Exemplary Disclosures
		facility 110 may be a satellite transmission facility, a head-end, a central office or other distribution center. The broadcast facility 110 may transmit the content to the receiving/presentation locations via satellite 170 or via a network 180. The network 180 may be the Internet, a cable television network (e.g., hybrid fiber cable, coaxial), a switched digital video network (e.g., digital subscriber line, or fiber optic network), broadcast television network, other wired or wireless network, public network, private network, or some combination thereof.” [0039] “The content delivery network 100 may deliver many different types of content. “ [0041]