

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

Patent No. 10,963,911

DECLARATION OF TIMOTHY WAHLERS

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	BACKGROUND AND QUALIFICATIONS	2
III.	MATERIALS CONSIDERED AND SUMMARY OF OPINIONS.....	11
IV.	PERSON OF ORDINARY SKILL IN THE ART	14
V.	THE '911 PATENT.....	17
VI.	CLAIM CONSTRUCTION	32
VII.	OVERVIEW OF THE PRIOR ART	33
A.	Overview of Steuer750 (EX1005) and Steuer389 (EX1006)	33
B.	Overview of Eldering (EX1006)	37
VIII.	DISCLOSURE OF CERTAIN CLAIM LIMITATIONS IN THE '911 PATENT CLAIMS AND ITS PARENT APPLICATIONS.....	47
A.	The Priority Applications Do Not Support The Video Data Features of The Challenged Claims	48
1.	The Priority Applications Do Not Provide Support for the Claimed "Video Data Being Divisible Into Samples" and Related Video Features	48
a)	The '836 Provisional Application Does Not Support the Claimed Video Data Features.....	53
b)	The '330, '604, '870, and '811 Priority Non- Provisional Applications Do Not Provide Adequate Support For The Challenged Claims	67
c)	The Combination of The '836 Provisional Application's Disclosure with the Priority Non- Provisional Applications Does Not Support the Video Data Features of the Challenged Claims	80
2.	Additional Materials Referenced in The Priority Non- Provisional Applications Do Not Support the Video Data Features of the Claims of the '911 Patent.....	82

3.	New Disclosures Added to The '204 Application That Issued as The '896 Patent	85
IX.	THE PRIOR ART DISCLOSES AND/OR SUGGESTS ALL RECITED FEATURES OF CLAIMS 1-17 OF THE '911 PATENT.....	89
A.	Grounds 1 and 2: Ground 1: The Combination of Steuer750 and Eldering Discloses and/or Suggests Claims 1-17 Ground 2: The Combination of Steuer389 and Eldering Discloses and/or Suggests Claims 1-17	89
1.	Claim 1	90
a)	A media measurement method comprising:	90
b)	receiving, at one or more computers, video data from a monitoring device, the video data being divisible into samples; and	94
c)	determining, by the one or more computers, a next sample of the video data for attempted identification of a media content item corresponding to the next sample by adjusting a length of a time window of the next sample based at least in part upon a result of an attempt by the one or more computers to identify a first prior sample as corresponding to a media content item stored in a computerized database of reference content;	114
d)	wherein the time length window of the next sample is changed relative to a time length window of the first prior sample based on the attempt to identify the first prior sample indicating a result that is different than a result of an attempt to identify second prior sample;	122
e)	submitting, by the one or more computers, the determined next sample for attempted content identification; and	124

f)	generating, by the one or more computers, a play stream of content identifications based on results of attempts to identify a series of samples comprising the determined next sample.	126
2.	Claim 2.....	129
a)	The method of claim 1 wherein determining comprises selecting a format of a sample.....	129
3.	Claim 3.....	130
a)	The method of claim 2 wherein: selecting a format includes selecting video parameters corresponding to requirements of a content recognition algorithm; and.....	130
b)	determining comprises, for a sample subsequent to an n'th sample, selecting video parameters corresponding to one of a plurality of available content recognition algorithms based at least in part upon an identification result corresponding to the n'th sample.....	131
4.	Claim 4.....	133
a)	The method of claim 2 wherein: selecting a format includes selecting video parameters corresponding to settings of a content recognition algorithm; and	133
b)	determining comprises, for a sample subsequent to an n'th sample, selecting a setting from a plurality of settings of a content recognition algorithm based at least in part upon an identification result corresponding to the n'th sample.	134
5.	Claim 5.....	136
a)	The method of claim 1 wherein channel data is used in determining the next sample, the channel data relating to recently identified samples.....	136
6.	Claim 6.....	137

a)	The method of claim 1 wherein content data is used in determining the next sample, the content data relating one or more recently identified samples.	137
7.	Claim 7.....	138
a)	The method of claim 6 wherein content data includes at least one of content type and content length.	138
8.	Claim 8.....	139
a)	The method of claim 1 wherein sample sequence data is used in determining the next sample, the sample sequence data relating to recently identified samples.	139
9.	Claim 9.....	140
a)	The method of claim 1 wherein log data is used in determining the next sample, the log data relating to a media player of an audience member using the monitoring device.	140
10.	Claim 10.....	142
a)	The method of claim 1 wherein location data is used in determining the next sample, the location data relating to a media player of an audience member using the monitoring device.	142
11.	Claim 11.....	143
a)	A media measurement method comprising:	143
b)	receiving, at one or more computers, video data from a monitoring device, the video data being divisible into samples; and	143
c)		145

- (1) determining, by the one or more computers, a next sample for attempted identification based at least in part upon a result of an attempt by the one or more computers to identify prior samples against a computerized database of known content.....145
- (2) wherein determining the next sample comprises selecting a relative time location of the next sample within the video data such that the relative time location of the next sample is selected to be between a relative time location of the first prior sample and a relative time location of the second prior sample based on a result of attempting to identify a first prior sample against the computerized database of known content being different from a result of attempting to identify a second prior sample against the computerized database of known content; 145
- d) submitting, by the one or more computers, the determined next sample for attempted content identification; and 153
- e) generating, by the one or more computers, a play stream of content identifications based on results of attempts to identify a series of samples comprising the determined next sample. 154
- 12. Claim 12 155
 - a) The method of claim 11 further comprising, for a series of samples of the video data: determining a sample by selecting an $(n \pm r)$ 'th sample for to be submitted for content recognition wherein r is determined at least in part based upon an identification result corresponding to the n 'th sample. 155
- 13. Claim 13 158

a)	The method of claim 12 wherein the magnitude of r is increased from a prior value if the identification result corresponding to the n'th sample does not indicate an identification.	158
14.	Claim 14.....	159
a)	The method of claim 12 wherein the magnitude of r is increased from a prior value if the identification result corresponding to the n'th sample indicates absence of any media content.....	159
15.	Claim 15.....	161
a)	The method of claim 14 further comprising: depending upon identification results corresponding to an n'th and an (n±r)'th sample, determining whether any samples captured between the n'th and the (n±r)'th sample should be submitted for a content recognition attempt.....	161
16.	Claim 16.....	163
a)	The method of claim 15 further comprising: submitting one or more samples captured between the n'th and the (n±r)'th sample for a content identification attempt if the identification results corresponding to the n'th and (n±r)'th sample do not match.....	163
17.	Claim 17.....	165
a)	The method of claim 11 wherein: if the result of attempting to identify the first prior sample against the computerized database of known content does not differ from the result of attempting to identify the second prior sample against the computerized database of known content, then the relative time location of the next sample within the video data is selected to be after relative time locations of the first and second prior samples.....	165
X.	CONCLUSION.....	170

I, Timothy Wahlers, declare as follows:

I. INTRODUCTION

1. I have been retained by Samsung Electronics Co., Ltd. (“Petitioner”) as an independent expert consultant in this proceeding before the United States Patent and Trademark Office (“PTO”) regarding U.S. Patent No. 10,963,911 (“the ’911 patent”) (EX1001). I have been asked to consider whether a person of ordinary skill in the art would have understood that the named inventors of the ’911 patent were in possession of the subject matter recited in certain claims of the ’911 patent at the time the priority applications were filed based on the disclosures of the priority applications. I have also been asked to consider whether prior art references disclose or suggest the features recited in claims 1 through 17 (“the challenged claims”) of the ’911 patent. My opinions are set forth below.

2. I am being compensated at my standard consulting rate for the time I spend on this matter. My compensation is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of this or any other proceeding. I have no other interest in this proceeding.

II. BACKGROUND AND QUALIFICATIONS

3. In this section, I have summarized my educational background, career history, publications, and other relevant qualifications. My Curriculum Vitae, which states my qualifications more fully, is filed as a separate Exhibit (EX1003).

4. I am the Owner and Principal at The Wahlers Group which offers technology, management, and expert witness consulting services to the digital video, cable, satellite, software, and consumer electronics industries. With over 35 years of industry experience, I specialize in digital video platforms, client and cloud-based platforms, content management via cloud systems, video-on-demand, and content streaming.

5. In 1981, I earned my B.Sc. in Nuclear Engineering from the University of Wisconsin, Madison. During my studies, I conducted post-graduate research in the field of fusion energy and microcomputer process control. I developed microprocessor monitoring software for the linear plasma experimental laboratory and conducted reactor core fuel analysis for the University Nuclear Reactor.

6. After completing my B.Sc., I worked in the field of nuclear engineering for 5 years, focusing on nuclear plant process computer monitoring and control. I developed software for various process computer systems and emergency response facilities. I also designed and developed PC-controlled laser disc application

software for a simulated journey through a nuclear containment environment. Additionally, I contributed to the development of an exhibit for the electric industry, which was showcased at the Chicago Science and Industry Museum. This work ignited my interest in digital media platforms, multimedia, and their use in education and entertainment.

7. In 1987, at the beginning of the digital media era, I joined OptImage Interactive Services as the Chief Engineer. OptImage was a joint venture of N.V. Philips, R.R. Donnelly, Sun Microsystems, and Microware. Over a span of 10 years, I was promoted to Vice President, Chief Technology Officer, and eventually President. My team was tasked with developing pioneering technology and tools for CD-ROM and CD-I, VideoCD, DVD, and Blu-ray digital MPEG video multimedia platform standards, which helped to advance the interactive television industry. During my tenure at OptImage, I led the development of software tools, interfaces, scripting languages, and applications that influenced the development of many of the early interactive video platforms. I have served and contributed to many industry standards bodies. I also led a project for an industry consortium that produced the world's first commercially sold MPEG A/V encoding solution.

8. In 1996, I founded Vidiom Systems Corporation, a company specializing in professional services and digital video products. As president and

sole owner, I steered the company's technical direction, cultivated business relationships, and oversaw the development of our marketed products and service contracts. Vidiom provided software development, integration, and testing services to leading consumer electronics, cable, satellite, and telecommunications companies across North America and Europe. My multifaceted role involved direct client engagement, serving as an architect, analyst, software developer, project lead, and management consultant. Our expertise centered on software for various platforms including PCs, mobile devices, set-top boxes, and TVs, as well as head-end systems for video and data delivery over broadband and IP networks. A key focus was the development of Electronic Program Guides and media-centric applications, including those synchronized with video delivery, audience measurement, and video-on-demand services.

9. From 1998 to 2001, while at Vidiom, I was contracted by CableLabs to serve as the Director of OpenCable Software for the cable industry. In this pivotal role, I led the development of the OpenCable Application Platform (OCAP) Java standard, functioning as its architect, analyst, and program lead. OCAP was designed as a standard platform for cable industry set-top boxes, enabling interoperability and supporting features such as program guides and DVR functionality. The platform's impact was substantial, achieving widespread adoption with deployments in over 20

million televisions and set-top boxes, significantly shaping the landscape of cable technology.

10. Vidiom grew to become the preeminent service provider in this space with over 110 employees and 60 subcontractors before I sold the company to the European entity, Advanced Digital Broadcast (ADB) in 2006, where I continued to serve as the CEO of Vidiom.

11. In 2008, I reacquired Vidiom's assets from ADB to establish EnableTV, a new venture that built upon Vidiom's mission and client base while expanding into cutting-edge technologies. EnableTV developed innovative products for the evolving digital television landscape, including: 1) an interactive television dashboard, 2) the OCAP Software Stack, 3) the VCERT testing platform, 3) an EnableTV application storefront, and 4) a next-generation content browser as a part of an IPTV program guide. These technologies positioned EnableTV at the forefront of the digital television industry, serving our established clientele while pushing boundaries in interactive and IP-based TV solutions.

12. In 2012, I sold the EnableTV assets and established The Wahlers Group, a strategic management, business development, and tactical consultancy specializing in digital video, interactive television platforms, and cloud-based content management platforms.

13. I have given many invited lectures and courses around the world on topics related to the software development lifecycle, processes, and DevOps. I have spoken at various conferences, including CES, NAB, Cable Show, and IBC. As part of my management consultancy, I have conducted training sessions for clients in software development and delivery.

14. In 2018, I co-founded Quantum Radius and currently serve as a board member. I provide technical, intellectual property, financing, and operations management services to the company. Quantum Radius is developing distributed, ultra-low-latency video processing technology for the VR/AR ecosystem. The technology is designed for simulation, medical, government, remote piloting, and eSports markets.

15. I have served on various company Boards of Directors and Technical Advisory Committees, including MyPhonix, MystroTV, RapidView Inc., and Quantum Radius.

16. I am honored to have received an industry award recognizing my contributions to my field, the *2006 ITVT Award for Leadership in Interactive Television*, <https://itvt.com/previous-awards/>, for my groundbreaking work in the digital video and interactive television industry.

17. Through my many decades of experience in software development and establishing companies in the media technology space, Automatic Content Recognition and audience measurement have frequently been recurring areas of interest. Partners and clients alike have developed ACR solutions that employ watermarking and fingerprinting techniques. The services offered through my many endeavors in the media technology industry have included software programming, program management, business development, partnering, and investment/merger opportunities, and those services have been provided to numerous prominent entities in the field of ACR, including:

- a. Philips (1997 – 2008): I worked for OptImage, a company mostly owned by Philips, from 1987 to 1996 before starting my own company, Vidiom Systems. In the late 1990s, Philips had a content identification group that I was familiar with. They made significant advances in content recognition technology. We looked into integrating their technology with ours, focusing on targeting the US cable industry. This group was spun off from Philips in 2008 as Civolution.

- b. Civolution; (2011 - 2013): We pursued the development of a demo of a 2nd screen application synchronized to video playing on a cable system OCAP Set-top.
- c. Enswers, Inc., which specialized in video/image fingerprinting and was later acquired by Gracenote (2012 – 2013): I evaluated Enswers ACR technology for investor, Stage 1 Ventures. The ACR technology utilized digital video frame fingerprinting within the TV. Various business relationships were considered utilizing EnableTV core technology to enhance Enswers offering for use in the cable industry and Verizon FIOS set-top boxes.
- d. CableLabs (1997 – 2012): One of the largest customers of Vidiom and EnableTV. I served as Director of OpenCable, Software in 1998 to 2000, and was contracted to provide a team, to assist the cable industry with software standards and reference implementations, from 2000 to 2012. I led the initial phase of the OCAP software platform definition and development. In that role, most industry vendors (including those involved in ACR technology), explored technology relationships with the cable industry and implementations on the OCAP platform.

- e. OpenTV/Nagra/Kudelski (1997 – 2008): A very large and first customer of Vidiom. I built a team under contract to develop tools and applications for OpenTV. We maintained their entire documentation collection from 1997 to 2001. ACR applications, mostly involving watermarking, were investigated on multiple occasions.
- f. The Nielsen Company (1995 – 2014): Nielsen was a major contributor to cable industry projects I led or participated in. I reviewed and commented on a proposed audience measurement data protocol specification. We explored the use of our technology to assist in Nielsen audience measurement solutions by tagging and watermarking audio and video content. We developed prototype software to generate media viewing log data for external delivery to Nielsen's platform.
- g. FourthWall Media (2010 – 2013): I worked with FourthWall on a project exploring television audience data collection on the OCAP platform.
- h. Gracenote, acquired by Nielsen in 2017 (2004): Explored joint projects and applications in media metadata technology. Using

ACR to feed 2nd screen applications. The focus was on tagging and watermarking.

- i. IBM (1996-1999): Multimedia platform work. Content identification, watermarking of both analog and digital video.

III. MATERIALS CONSIDERED AND SUMMARY OF OPINIONS

18. The opinions contained in this Declaration are based on the documents I reviewed and my professional judgment, as well as my education, experience, and/or knowledge regarding technologies relating to, among other things, networking technologies.

19. In forming my opinions expressed in this Declaration, I reviewed the following materials (and any other materials that I reference in this Declaration):

EX1001	U.S. Patent No. 10,963,911
EX1003	Curriculum Vitae of Timothy Wahlers
EX1004	File History of U.S. Patent Application No. 16/660,633 (U.S. Patent No. 10,963,911)
EX1005	U.S. Publication No. 2005/0267750 to Steuer <i>et al.</i>
EX1006	U.S. Patent No. 8,677,389 to Steuer <i>et al.</i>
EX1007	U.S. Publication No. 2006/0195860 to Eldering <i>et al.</i>
EX1008	Computer-generated Redline Between Specification of U.S. Publication No. 2005/0267750 to Steuer <i>et al.</i> (EX1005) and Specification of U.S. Patent No. 8,677,389 to Steuer <i>et al.</i> (EX1006)
EX1009	Computer-generated Redline Between Claims 1-17 of U.S. Patent No. 10,963,911 (EX1001) and Claims 1-17 of U.S. Patent No. 8,677,389 (EX1006)
EX1010	File History of U.S. Provisional Application No. 60/574,836
EX1011	File History of U.S. Patent Application No. 11/139,330
EX1012	File History of U.S. Patent Application No. 12/954,604

EX1013	File History of U.S. Patent Application No. 13/914,870
EX1014	File History of U.S. Patent Application No. 14/175,811
EX1015	File History of U.S. Patent Application No. 16/403,204 (U.S. Patent No. 10,572,896)
EX1017	Computer-generated Redline Between As-Filed Specifications in U.S. Patent Application No. 11/139,330 (EX1011) and U.S. Patent Application No. 12/954,604 (EX1012)
EX1018	Computer-generated Redline Between As-Filed Specifications in U.S. Patent Application No. 11/139,330 (EX1011) and U.S. Patent Application No. 13/914,870 (EX1013)
EX1019	Computer-generated Redline Between As-Filed Specifications in U.S. Patent Application No. 11/139,330 (EX1011) and U.S. Patent Application No. 14/175,811 (EX1014)
EX1020	U.S. Patent Publication No. 2002/0083060
EX1021	WO2004/044820
EX1022	WO2004/030341
EX1023	WO2004/040475
EX1028	Comparison between video-related paragraph in U.S. Provisional Application No. 60/574,836 and video- related paragraph in U.S. Patent Application No. 16/403,204
EX1030	TRECVID 2004 – An Overview (February, 2005)
EX1031	Guidelines for the TRECVID 2004 Evaluation (February, 2005)
EX1032	TRECVID-2005: Search Task (Dublin City University and NIST)

20. In support of my opinions, I have taken into account how a person of ordinary skill in the art (as defined below in Section IV) would have understood the claims and the specification of the '911 patent at the time of the alleged invention. My opinions take into account how a person of ordinary skill in the art would have understood the '911 patent and the state of the art at the time of the alleged invention.

21. Based on my experience and expertise, it is my opinion (i) that a person of ordinary skill in the art would not have understood that the named inventors of the '911 patent were in possession of the subject matter recited in certain claims of the '911 patent at the time the priority applications were filed based on the disclosures of the priority applications and (ii) that certain references disclose and/or suggest all the features recited in the challenged claims of the '911 patent, as I discuss in detail below.

IV. PERSON OF ORDINARY SKILL IN THE ART

22. I have considered the time of the alleged invention for the '911 patent to be around the mid-2019 timeframe, including up to May 3, 2019, the filing date of U.S. Patent Application No. 16/403,204 ("the '204 application"), which issued as U.S. Patent Application No. 10,572,896 ("the '896 patent"). (EX1001 at Cover.) U.S. patent application no. 16/660,633 ("the '633 application") which issued as the '911 patent, is a continuation of the '204 application, which added new matter not contained in any of the prior applications. I may refer to this time frame in this declaration as the time of the alleged invention, the relevant time, or the time of the '911 patent.

23. Based on my knowledge and experience, I understand what a person of ordinary skill in the art ("POSITA") would have known at the time of the alleged invention of the '911 patent. In my opinion, based on my review of the '911 patent, the technology, the educational level and experience of active workers in the field, the types of problems faced by workers in the field, the solutions found to those problems, the sophistication of the technology in the field, and drawing on my own experience, a person of ordinary skill in the art would have had an undergraduate degree in Electrical Engineering, Computer Engineering, Computer Science or a related field along with at least 4 years of work experience in the field of media-

related software development, or in the alternative, equivalent experience, such as 6 years of work or research experience in the field of media-related software development. (*See, e.g.*, EX1001 at 1:20-62 (discussing the background of the '911 patent), 1:66-2:22 (discussing a summary of the disclosure).) More education can substitute for practical experience and *vice versa*. For example, graduate level education could supplement relevant work experience and significant work experience could supplement education experience.

24. My analysis of the '911 patent and my opinions in this declaration are from the perspective of one of ordinary skill in the art, as I have defined it above, at the time of the alleged invention for that patent.

25. I note that my opinions in this declaration do not change even if the time of the alleged invention for the '911 patent is considered to be around the late 2010 timeframe, including up to the November 24, 2010 filing date of U.S. Patent Application No. 12/954,604 ("the '604 application"), which I understand is an intervening application to which the '633 application claims priority.

26. I further note that my opinions in this declaration do not change even if the time of the alleged invention for the '911 patent is considered to be around the mid-to-late 2008 timeframe, including up to the September 11, 2008 filing date of

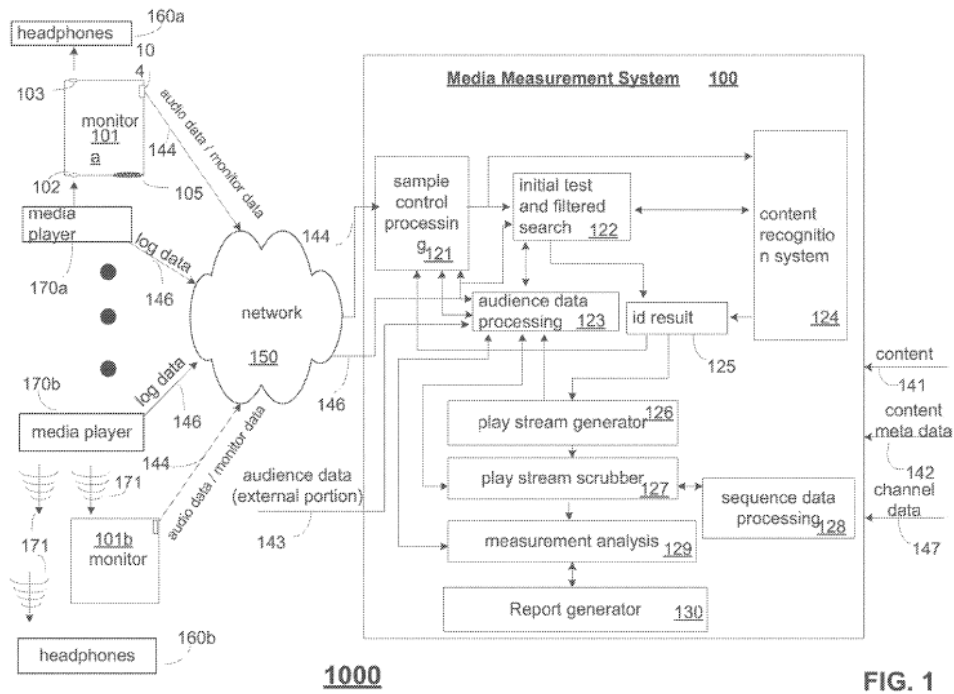
U.S. Patent Application No. 12/283,353 (“the ’353 application”), which I understand is another intervening application to which the ’633 application claims priority.

27. I further note that my opinions in this declaration do not change even if the time of the alleged invention for the ’911 patent is considered to be around the mid-2004 timeframe, including up to the May 27, 2004 filing date of U.S. Provisional Patent Application No. 60/574,836 (“the ’836 provisional application”), which I understand is the earliest application to which the ’633 application claims priority.

V. THE '911 PATENT

28. The '911 patent is directed to media usage monitoring and measurement systems and methods. (EX1001 at Title, Abstract, 1:66-2:22.) The media measurement method, in particular, relates to recognizing content being consumed by a user and performing additional actions with the recognized content, such as creating a clean play list for the user, identifying a source of the content, and deducing play-altering activities a user engages in (e.g., fast-forwarding, rewinding).

29. As shown in Figure 1 below, the '911 patent discloses its media measurement methods in the context of capturing audio data from content consumed by a user whereby a monitoring device, such as monitors 101a/101b, captures the audio data. (EX1001 at 3:5-44, FIG. 1.) The audio data, or a signature calculated from the audio data, is then sent via a network to a system 100 to be recognized by comparing it against data stored in a database containing data representing known content. (EX1001 at 3:45-4:6, 4:22-57.) System 100 identifies the content and generates content identification results using the content recognition system 124. (EX1001 at 4:58-6:23.)



(EX1001 at FIG. 1.)

30. Figure 2 below illustrates a media measurement method 200 in which the method steps may be carried out by media measurement system 100 of Figure 1 above. (EX1001 at 7:1-5, FIG. 2.)

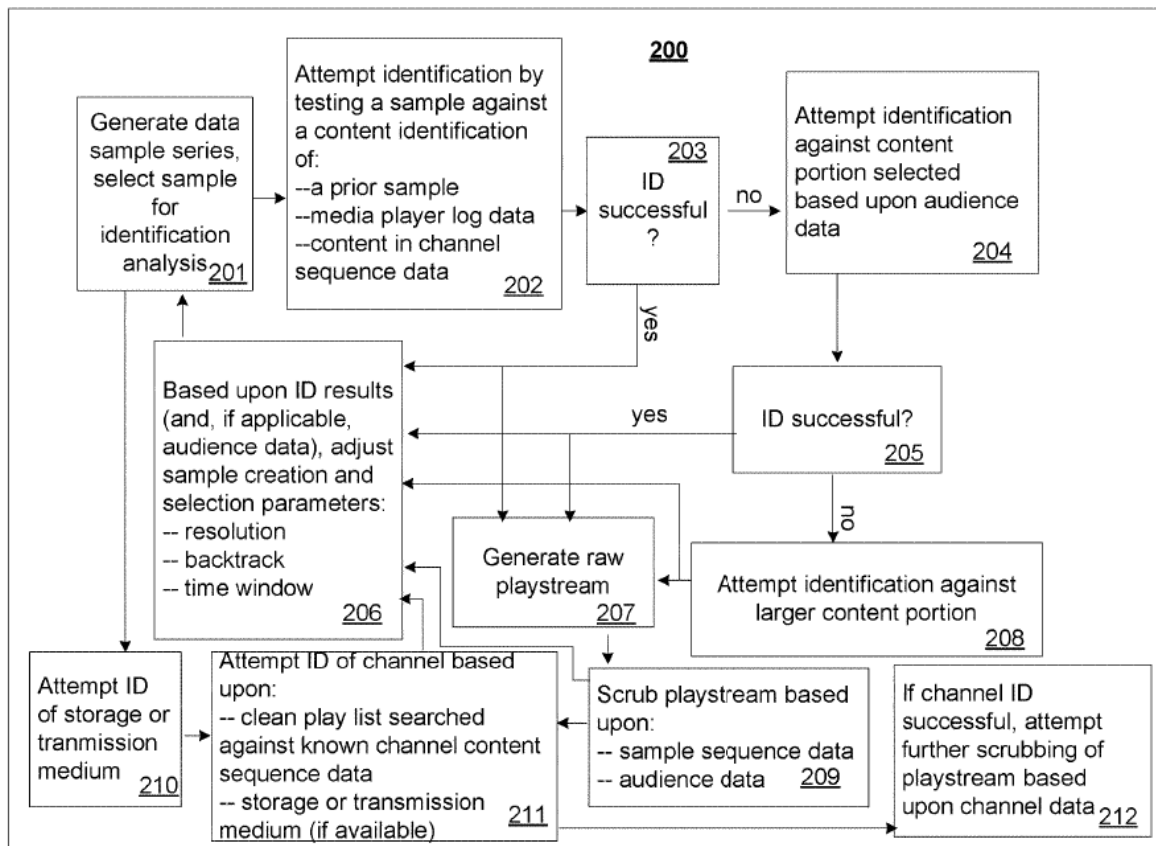


FIG. 2

(EX1001 at FIG. 2.)

31. With reference to Figure 2, the '911 patent describes a content recognition process using audio content samples. (EX1001 at 7:1-6.) For example, Figure 2 describes using prior content identification results in the audio content recognition process to “adjust sampling control and selection parameters for one or more subsequent samples.” (EX1001 at 7:49-51.) The parameters that may be adjusted include the audio sample length (also referred to as “time window” of each audio sample) or the selection of a previously unanalyzed intervening audio sample

for audio content identification (“backtrack selection” or “backtrack analysis”). (*Id.*, 7:51-54, 59-64 (describing backtracking analysis in step 206 where “if an identification result does not match the result of the prior sample identified, then step 206 might initiate a back track analysis and select an intervening sample that has not yet been analyzed.”), 8:1-10 (describing adjustment of audio sample length); *see also id.*, 7:5-13 (describing that in step 201, a series of audio data samples is received from a monitoring device and a “length, or ‘time window’ of each sample” is determined); *see also id.*, 13:46-14:16 (description of Figure 6 similarly explaining adjustment of sampling control and selection parameters like the backtracking analysis or changing the time length window of samples).).

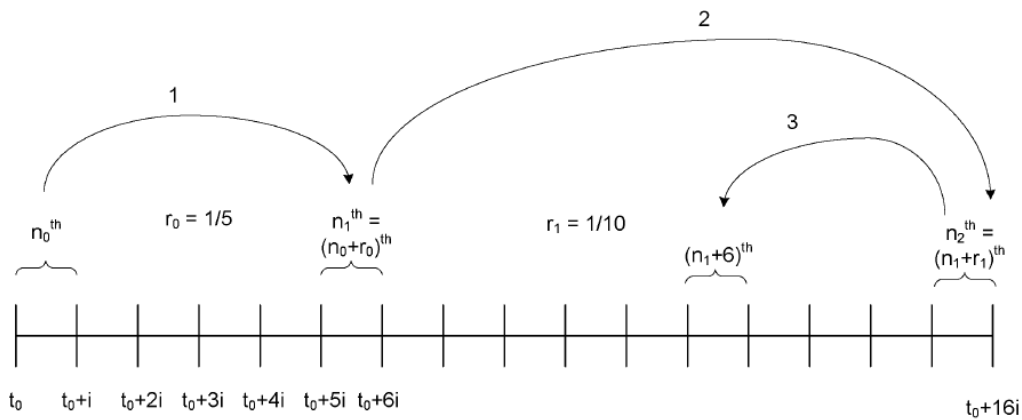
At step 206, an identification result is used to **adjust sampling control and selection parameters for one or more subsequent samples**. Parameters that might be adjusted include resolution, **backtrack selection, sample time window**, or recognition algorithm selected. For example, if a sample has been identified successfully, and the identification indicates a match with the prior sample, the sampling resolution might be decreased from $\frac{1}{5}$ to $\frac{1}{10}$, meaning that, in a sequence of samples, nine samples

rather than four are skipped before selecting the next sample to analyze for identification. **As another example, if an identification result does not match the result of the prior sample identified, then step 206 might initiate a back track analysis and select an intervening sample that has not yet been analyzed. Step 201 would then pass the sample selected from backtracking to step 202 for attempted identification.** As another example, resolution might be increased when recent identification results suggest an extended period of silence or absence of known content.

As yet another example, **once media content for a limited series of samples has been identified, the time window size for each sample might be decreased from, for example, 5 seconds to 3 seconds. This would mean that samples submitted for attempted identification would be shorter in length.** This can create efficiencies because once the content is identified, the process needs only to detect a change in content, and this can, in some

cases, be accomplished utilizing a smaller amount of audio
data, thereby preserving system resources.

(EX1001 at 7:49-8:10.)¹



600

FIG. 6

(EX1001 at FIG. 6.)

¹ I sometimes add emphases to text of a document that I quote (as I do here) in this declaration. All such emphasis is added unless I state or explain otherwise.

FIG. 6 illustrates timeline structure for data samples generated by system and methods such as module 121 of FIG. 1 and steps 201 and 206 of FIG. 2.

As illustrated, sample set 600 reflects a division of audio data into sixteen samples, three of which are selected for analysis. The samples selected for recognition analysis are referenced generally as the “ n_0^{th} ,” “ n_1^{th} ,” and “ n_2^{th} ” samples. Samples between the selected samples are not separately numbered, but are marked off between vertical lines along the horizontal length of timelines structure 600 as illustrated. **Each sample has a sample time window length “i.”** Furthermore, the sample set has varying “resolutions” including $r_0=1/5$ between the n_0^{th} and n_1^{th} samples and $r_1=1/10$ between the n_1^{th} and the n_2^{th} samples. In other words, “resolution” here refers to the portion of samples being selected during a period of time. In the present example, a process portion such as steps 201 and 206 of process 200 of FIG. 2 has, based on the identification results of the n_0^{th} and n_1^{th} samples, adjusted

the resolution from $\frac{1}{5}$ to $\frac{1}{10}$. This could occur, for example, as previously described, because the n_0^{th} and n_1^{th} samples match the same content ID and process 200 determines that the likelihood that the next selected sample using a resolution of $\frac{1}{5}$ would also match justifies decreasing the resolution to $\frac{1}{10}$ so that the next sample selected for analysis is the n_2^{th} . **However, if the n_2^{th} sample does not match the n_1^{th} sample, then process 200 can make a backtrack decision and go back in the series of samples to select, for example, the $(n_1+6)^{\text{th}}$ sample for content recognition analysis in order to pinpoint more closely where in the play stream the content change occurred. The order in which samples are selected for analysis in this example is illustrated by arrows 1, 2, and 3 in FIG. 6.**

Although not illustrated in FIG. 6, as previously described, steps 201 and 206 of method 200 can also determine based on identification results the time window length can be adjusted. In the context of FIG. 6,

this would correspond to changing the value of “i,” the
time window length of a particular sample.

(EX1001 at 13:46-14:16.)

32. Next, the content identification results are sent to play stream generator 126 to generate a raw play stream 410, as shown in Figure 4 below. (EX1001 at 6:24-37, 8:13-34, 9:25-45.) The raw play stream 410 “itemizes a sequence of analyzed samples and relates that sequence to corresponding identification results.” (EX1001 at 9:25-45.) Once the raw play stream 410 is generated, the media measurement method includes performing additional processing and analysis of the raw play stream 410, such as scrubbing the play stream to generate a clean play stream 450, identifying a channel associated with the recognized content, or deducing play-altering activities as shown in the clean play list 590 in figure 5. (EX1001 at 6:24-37, 8:13-34, 8:42-50 (describing the process of scrubbing as including “reviewing for missing or incorrect data”), 9:46-51 (describing the scrubbing process as including “analyzing the sequence of data on Raw Play Stream, it is possible to locate and correct any identifications that are likely to be incorrect or to supply missing identifications where the system has not otherwise been able to provide an identification”), 9:52-10:50, 11:8-28.)

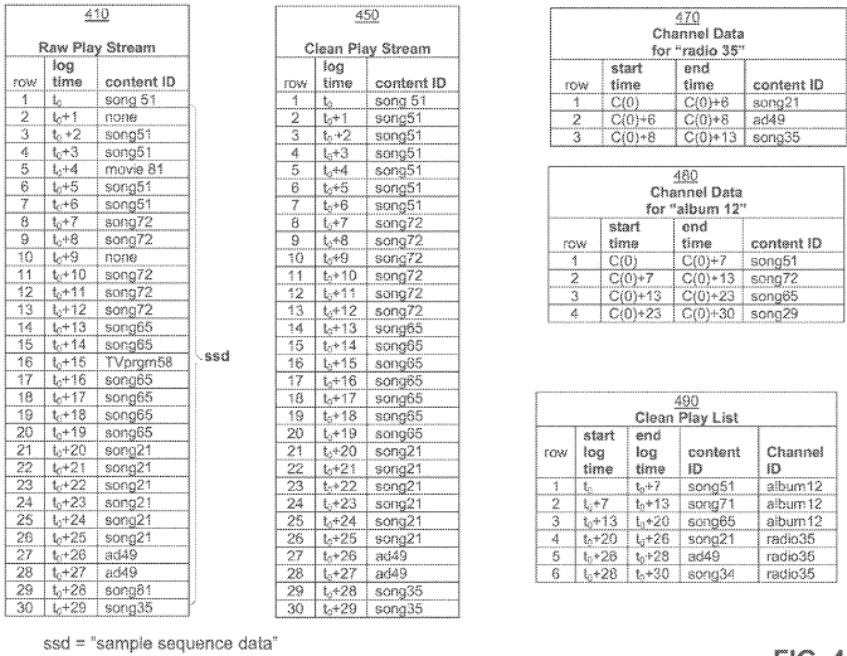


FIG. 4

(EX1001 at FIG. 4.)

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

580 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

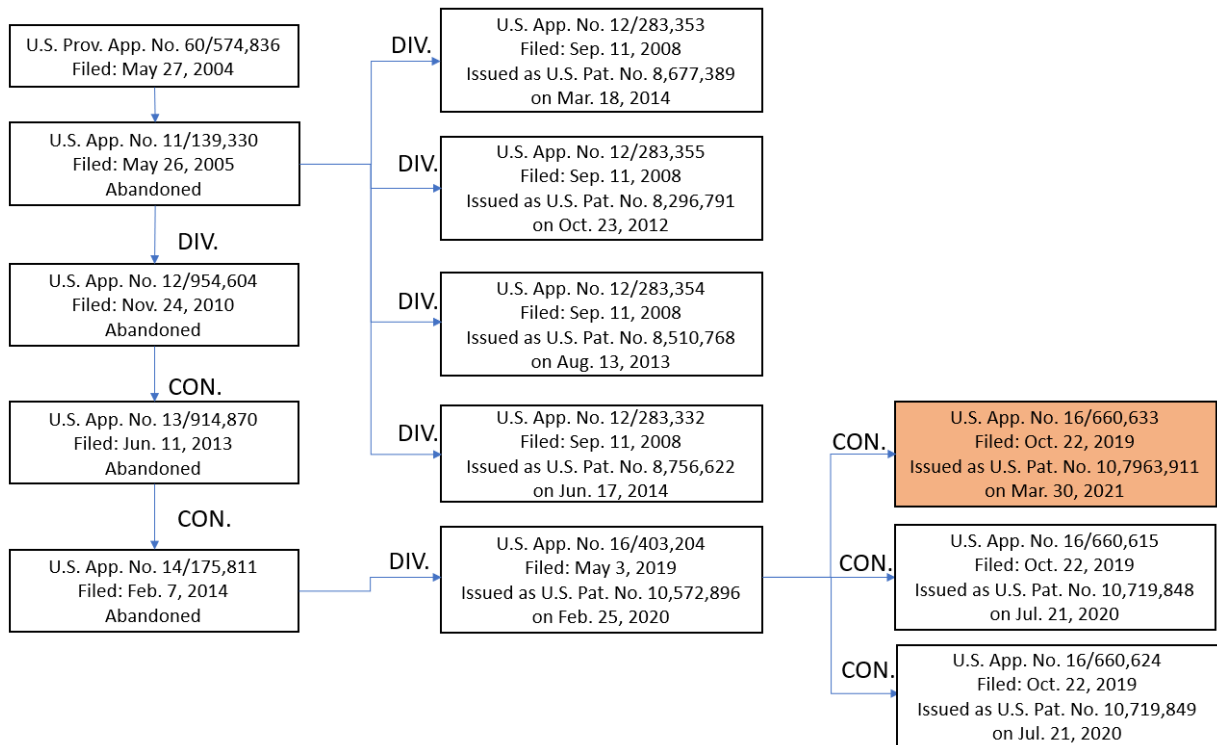
(EX1001 at FIG. 5.)

33. Throughout the '911 patent, the disclosures describe only embodiments related to recognizing, analyzing, or processing samples of audio data and do not describe any features that specifically rely on using video data samples (to identify content). (*See, e.g.*, EX1001 at 1:37-39, 2:17-22, 3:13-24, 4:23-33, 6:13-17, 7:14-21, 11:47-52.) Even where the '911 patent discusses video-related content (e.g., television, TV, movie, etc.), it does so only in the context of recognizing, analyzing, or processing the audio data portion of the captured video (with audio) content. (*See, e.g.*, EX1001 at 1:56-62, 4:23-33, 7:14-21, 11:47-52.)

34. The challenged claims, on the other hand, include claim limitations directed to recognizing, analyzing, or processing “**video data** being divisible into samples.” (EX1001 at claim 1). However, as I have discussed above and further discuss below, the ’911 patent and its priority applications only describe embodiments involving recognizing, analyzing, or processing audio data samples. (See my discussions and explanations below in Section VIII; *see also generally* EX1001 at 2:61-15:10 (Detailed Description section discussing aspects related to audio data).)

35. Priority Applications of the ’911 patent: I understand that the ’911 patent issued from U.S. Patent Application No. 16/660,633 (“the ’633 application”) (EX1004), which was filed on October 22, 2019. (EX1001 at Cover.) I also understand that the ’633 application is a continuation of U.S. patent application no. 16/403,204 (“the ’204 application”). I understand that the ’204 application is a divisional of U.S. patent application no. 14/175,811 (“the ’811 application”). I also understand that the ’811 application is a continuation of U.S. patent application no. 13/914,870 (“the ’870 application”). I understand that the ’870 application is a continuation of U.S. patent application no. 12/954,604 (“the ’604 application”). I also understand that the ’604 application is a divisional of U.S. patent application no. 11/139,330 (“the ’330 application”). I understand that the ’330 application is a

is a non-provisional application claiming priority to U.S. provisional application no. 60/574,836 (“the ’836 provisional application”). The chart in Demonstrative A below helps visualize the purported priority chain for the ’911 patent.



Demonstrative A²

² The chart in Demonstrative A also shows additional applications that I understand are related to the ’911 patent. For example, the chart shows other applications (U.S.

36. Based on my review, the priority applications of the '911 patent do not contain disclosures adequately describing recognizing, analyzing, or processing video data samples of the video portion of the video content, in the ways set forth in the challenged claims. (*See* my discussion below in Section VIII.) Thus, as I discussed above, I have considered the time of the alleged invention for the '911 patent to be around the mid-2019 timeframe, including up to May 3, 2019, the filing date of the '204 application, which added new matter not contained in any of the prior applications. The '633 application, which issued as the '911 patent, is a

patent application nos. 16/660,615 and 16/660,624) that I understand claims priority to the '204 application, which issued as the U.S. Patent No. 10,572,896—I understand that application no. 16/660,615 issued as U.S. Patent No. 10,719,848 and application no. 16/660,624 issued as U.S. Patent No. 10,719,849. The chart also shows four divisional applications (U.S. patent application nos. 12/283,353, 12/283,355, 12/283,354, and 12/283,332) that I understand claim priority to the '330 application—I understand that these divisional applications issued as U.S. Patent Nos. 8,677,389, 8,296,791, 8,510,768, and 8,756,622, respectively.

continuation of the '204 application. (*See* my discussion above in Section IV and my discussion below in Section VIII.)

VI. CLAIM CONSTRUCTION

37. I have given the terms in the challenged claims their plain and ordinary meanings, as would have been understood by a person of ordinary skill in the art at the time of the alleged invention, taking into consideration the language of the claims, the specification, and the prosecution history of the '911 patent. I have applied these understandings of the claim terms in my analysis and in forming my opinions where appropriate in this Declaration.

VII. OVERVIEW OF THE PRIOR ART³

A. Overview of Steuer750 (EX1005) and Steuer389 (EX1006)

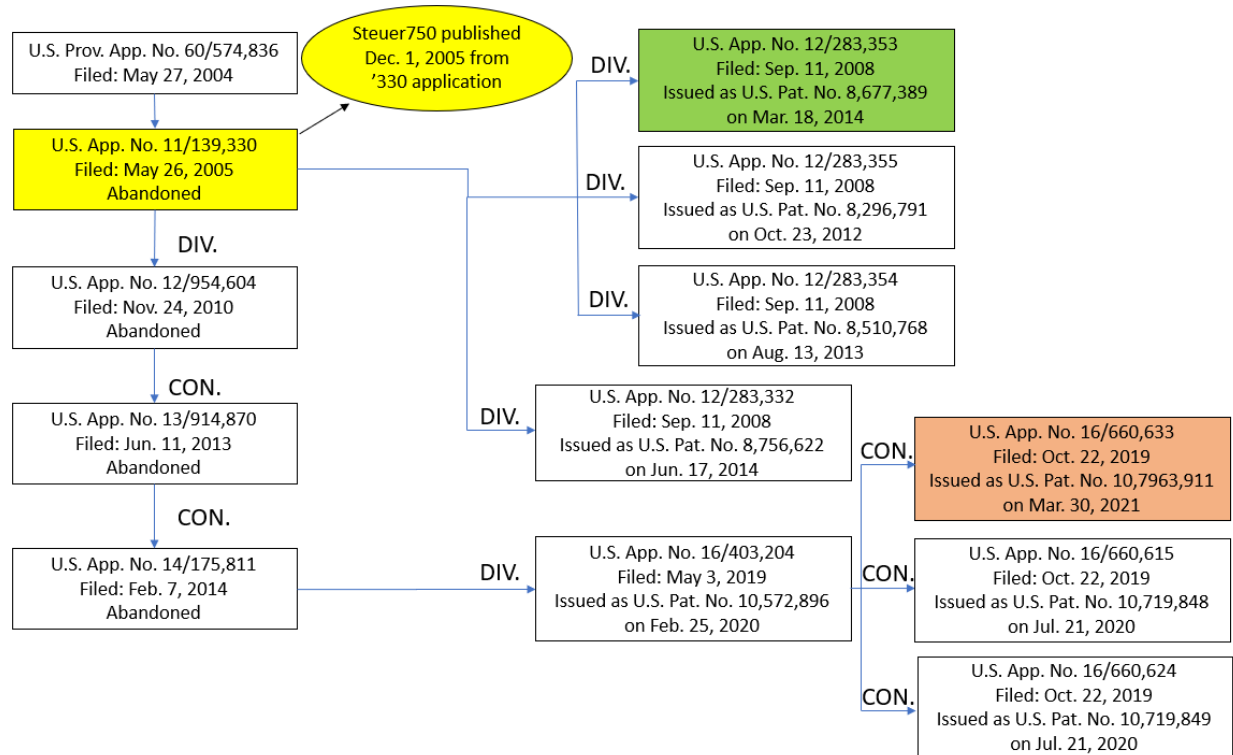
38. I understand that Exhibit 1005 is U.S. Publication No. 2005/0267750 to Steuer *et al.*, which I refer to as “Steuer750.” (EX1005 at Cover.) I understand that Steuer750 is a publication of the ’330 application, to which the ’911 patent claims priority. (*See* my discussion above in Section V.)

39. I understand that Exhibit 1006 is U.S. Patent No. 8,677,389 to Steuer *et al.*, which I refer to as “Steuer389.” (EX1006 at Cover.) I understand that Steuer389 is a patent that is part of the same family as the ’911 patent family and Steuer750, as it claims priority to the ’330 application. (*See* the demonstrative above in Section V.)

40. Demonstrative B below, which is an edited version of Demonstrative A in Section V above, shows a chart of the ’911 patent’s priority applications and

³ These summaries of the prior art in Section VII are applicable to my analysis below in Section IX. Accordingly, I incorporate these summaries, which reflect how a person of ordinary skill in the art would have understood the references, into my analysis below where appropriate (*see, e.g.*, Section IX below).

related applications where the Steuer750 publication and its related '330 application are highlighted in yellow and the Steuer389 patent is highlighted in green:



Demonstrative B

41. Thus, like the '911 patent, Steuer750 and Steuer389 are directed to measurement systems and methods for media usage monitoring involving recognizing content being consumed by a user and performing additional actions with the recognized content, such as creating a clean play list for the user, identifying a source of the content, and deducing play-altering activities a user engages in (e.g., fast-forwarding, rewinding). (EX1005 at Title, Abstract, [0006]; EX1006 at Title, Abstract, 1:60-2:15.) Steuer750 and Steuer389 both disclose the “adjusting a length

of a time window of the next sample” features of independent claim 1 and the backtrack analysis features of independent claim 11 of the ’911 patent, albeit in the context of audio samples. (*See, e.g.*, EX1005 at [0037]-[0038], [0063]-[0065], FIGS. 2, 6; *see also id.*, [0034], claims 19-21; EX1006 at 7:39-65, 13:20-56, FIGS. 2, 6; *see also id.*, 7:2-3.)

42. In fact, other than (1) the addition to the ’204 application (the ’633 application, which issued as the ’911 patent, being a continuation of the ’204 application) of the single paragraph purported to disclose video features (*see my discussion and explanation below in Section VIII*), and (2) differences in the respective priority claims in the “Cross Reference to Related Applications” sections, it is my opinion that the specifications of the ’911 patent, Steuer750, and Steuer389 are substantively identical. (EX1001 at 1:6-18, 14:17-29; EX1005 at [0001]; EX1006 at 1:7-10; *see also* EX1008.)

43. In addition, based on my review, claims 1-17 of Steuer389 are nearly identical to claims 1-17 of the ’911 patent, except the word “audio” in claims 1-17 of Steuer389 is substituted with “video” in claims 1-17 of the ’911 patent. (*Compare* EX1001 at claims 1-17 *with* EX1006 at claims 1-17; *see also* EX1009 (showing a comparison of the claims).)

44. I note that based on my review, certain clauses in limitations 1.d and 11.c.2 in the challenged claims are moved around in the corresponding limitations in claims 1 and 11 of Steuer389. (See, e.g., EX1009 at 1, 4.) For example, while limitation 1.d of the '911 patent recites “wherein the time length window of the next sample is changed relative to a time length window of the first prior sample based on the attempt to identify the first prior sample indicating a result that is different than a result of an attempt to identify a second prior sample” (EX1001 at 15:23-28), the corresponding portion of claim 1 of Steuer389 similarly recites “wherein, if the attempt to identify the first prior sample indicates a result that is different than a result of an attempt to identify a second prior sample, then the time length window of the next sample is changed relative to a time window of the first prior sample” (EX1006 at 14:51-56). And while limitation 11.c.2 of the '911 patent recites “wherein determining the next sample comprises selecting a relative time location of the next sample within the video data such that the relative time location of the next sample is selected to be between a relative time location of the first prior sample and a relative time location of the second prior sample based on a result of attempting to identify a first prior sample against the computerized database of known content being different from a result of attempting to identify a second prior sample against the computerized database of known content” (EX1001 at 16:16-26), the

corresponding portion of claim 11 of Steuer389 similarly recites “wherein determining the next sample comprises selecting a relative time location of the next sample within the audio data such that if a result of attempting to identify a first prior sample against the computerized database of known content differs from a result of attempting to identify a second prior sample against the computerized database of known content, then the relative time location of the next sample within the audio data is selected to be between a relative time location of the first prior sample and a relative time location of the second prior sample” (EX1006 at 15:36-16:6).

45. Moreover, limitations 1.e-f and 11.d-e of the '911 patent regarding the submission of the determined next sample for attempted content identification and generation of a play stream of content identifications (EX1001 at 15:29-34, 16:27-32) are not included in the Steuer389 claims. (*See, e.g.*, EX1009 at 1-2, 4-5.)

B. Overview of Eldering (EX1006)

46. Eldering is directed to methods and systems used to identify media content (e.g., advertisements and programs) being consumed by users by matching samples of the consumed media content with a content database. (*See, e.g.*, EX1007 at Abstract, [0008]-[0010]; *see also id.*, [0001]-[0007], [0072].) Eldering discloses that the incoming content data is divisible into samples. (EX1007 at [0008] (describing the use of “evenly [] subsampled representations of the original video”

for the content identification process), [0070]-[0076] (describing content sampling examples), [0101], FIGS. 7-9.)

There is a need for a system and method that allows for the insertion of advertisements in video streams. There is also a need for a system which allows advertisements to be better targeted to audiences and for the ability for operators of commercial premises to cross-market services and products to the audience. Additionally, there is a need for a system which enables the operators of commercial premises to eliminate and substitute advertising of competitors' products and services included in broadcasts shown to guests on their premises.

In the absence of cue tones, such as broadcaster supplied cue tones, indicating the boundaries of advertisement breaks another means of detecting the display of an advertisement is required. One method includes calculating features about an incoming video stream. These features may include color histograms, color coherence vectors (CCVs), **and evenly or randomly**

highly subsampled representations of the original video (all known as fingerprints). The fingerprints of the incoming video stream are compared to a database of fingerprints for known advertisements, video sequences known to precede commercial breaks (ad intros), and/or sequences known to proceed commercial breaks (ad outros). When a match is found between the incoming video stream and a known advertisement or ad intro, the incoming video stream is associated with the known advertisement and/or ad intro and a targeted advertisement may be substituted.

The fingerprint of the incoming video stream (calculated fingerprint) may be compared to a plurality of fingerprints for known entities (e.g., ads, intros) within the database (known fingerprints). The comparison may be done based on small segments of a video stream at a time. A determination is made as to whether the calculated fingerprint and the known fingerprints within the database exceed some threshold level of dissimilarity. If the

comparison exceeds the threshold for certain known fingerprints within the database, the comparison of the calculated fingerprint to those known fingerprints stops for the time being. For those known fingerprints that the comparison was below the threshold level of dissimilarity the comparison continues. At each step of the comparison those known fingerprints exceeding the threshold level of dissimilarity cease. The process continues until one of the known fingerprints has a comparison that exceeds a threshold level of similarity (indicating a match) or the comparison of all of the known fingerprints within the database exceed the dissimilarity threshold at which point the video stream is not associated with any of the known fingerprints.

Calculating fingerprints for the incoming video stream and comparing to a database of fingerprints of known entities can also be used to detect certain programs and/or scenes and to record, bookmark or stop recording the programs and/or scenes.

(EX1007 at [0007]-[0010].)

47. Eldering discloses identifying content using video data or audio data. (EX1007 at [0008]-[0010] (describing content identification using video data, as shown above), [0050], [0064], [0086].)

Automatic techniques for detecting advertisements (or advertisement blocks) may include detecting aspects (features) of the video stream that indicate an advertisement is about to be displayed or is being displayed (feature based detection). **For example, advertisements are often played at a higher volume then programming so a sudden volume increase (without commands from a user) may indicate an advertisement.** Many times several dark monochrome (black) frames of video are presented prior to the start of an advertisement so the detection of these types of frames may indicate an advertisement. The above noted techniques may be used individually or in combination with one another. These techniques may be utilized along with temporal measurements, since commercial breaks

often begin within a certain known time range. However, these techniques may miss advertisements if the volume increases or if the display of black frames is missing or does not meet a detection threshold. Moreover, these techniques may result in false positives (detection of an advertisement when one is not present) as the programming may include volume increases or sequences of black frames.

(EX1007 at [0050] (disclosing audio data like volume level being indicative of content to be identified).)

In some countries, commercial break intros are utilized to indicate to the viewers that the subsequent material being presented is not programming but rather sponsored advertising. These commercial break intros vary in nature but may include certain logos, characters, or other specific video and audio messages to indicate that the subsequent material is not programming but rather advertising. The return to programming may in some instances also be preceded by a commercial break outro which is a short

video segment that indicates the return to programming. In some cases the intros and the outros may be the same with an identical programming segment being used for both the intro and the outro. Detecting the potential presence of the commercial break intros or outros may indicate that an advertisement (or advertisement block) is about to begin or end respectively. If the intros and/or outros were always the same, detection could be done **by detecting the existence of specific video or audio**, or specific logos or characters in the video stream, or by detecting specific features about the video stream (e.g., CCVs). However, the intros and/or outros need not be the same. The intros/outros may vary based on at least some subset of day, time, channel (network), program, and advertisement (or advertisement break).

(EX1007 at [0064] (disclosing detection of the start/end of commercial breaks using audio or video data).)

Advertisements only make up a portion of an incoming video stream so that continually calculating features for

the incoming video stream 1020 and comparing the features to known advertisement fingerprints 1030 may not be efficient. **According to one embodiment, the feature based techniques described above (e.g., volume increases, increase scene changes, monochrome images) may be used to detect the start of a potential advertisement (or advertisement block) and the calculating of features 1020 and comparing to known fingerprints 1030 may only be performed once a possible advertisement break has been detected.** It should be noted that some methods of detecting the possibility of an advertisement break in the video stream such as an increase in scene changes, where scene changes may be detected by comparing successive CCVs, may in fact be calculating features of the video stream 1020 so the advertisement detection process may begin with the comparison 1030.

(EX1007 at [0086] (disclosing detection of the start of potential advertisements using audio data like volume increases or video data like scene changes).)

48. Eldering further discloses performing post-content identification steps such as substituting one advertisement for another, bookmarking desired content, or starting/stopping recording of content based on the content identification. (*Id.*, [0008], [0010], [0043]-[0044], [0082], [0158]-[0163], [0167], claims 3, 5, FIGS. 1, 2, 10.)

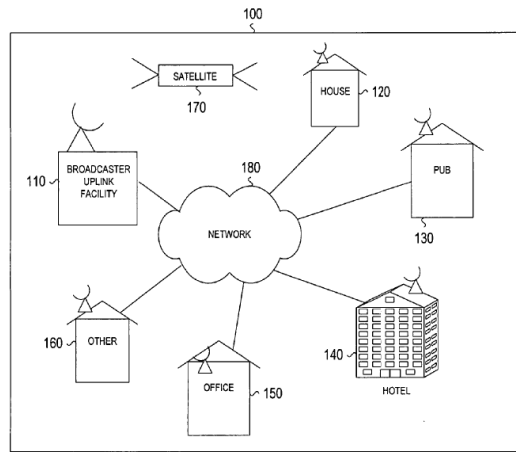


FIG. 1

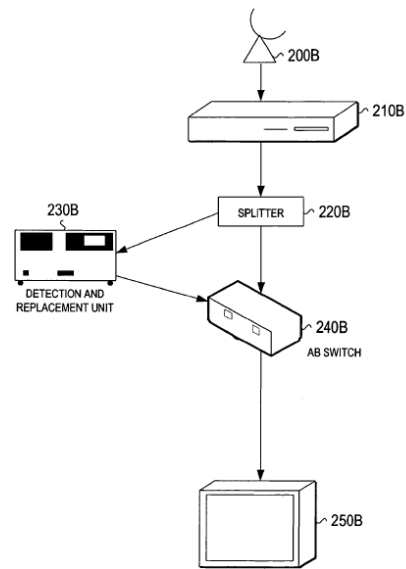


FIG. 2

(EX1007 at FIGS. 1-2; *see also id.*, [0039]-[0043], [0046]-[0048].)

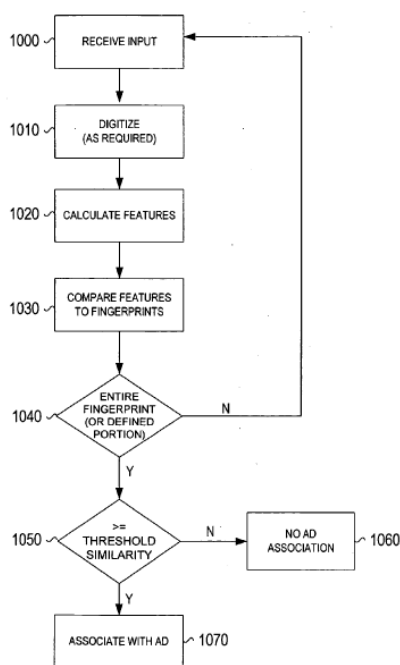


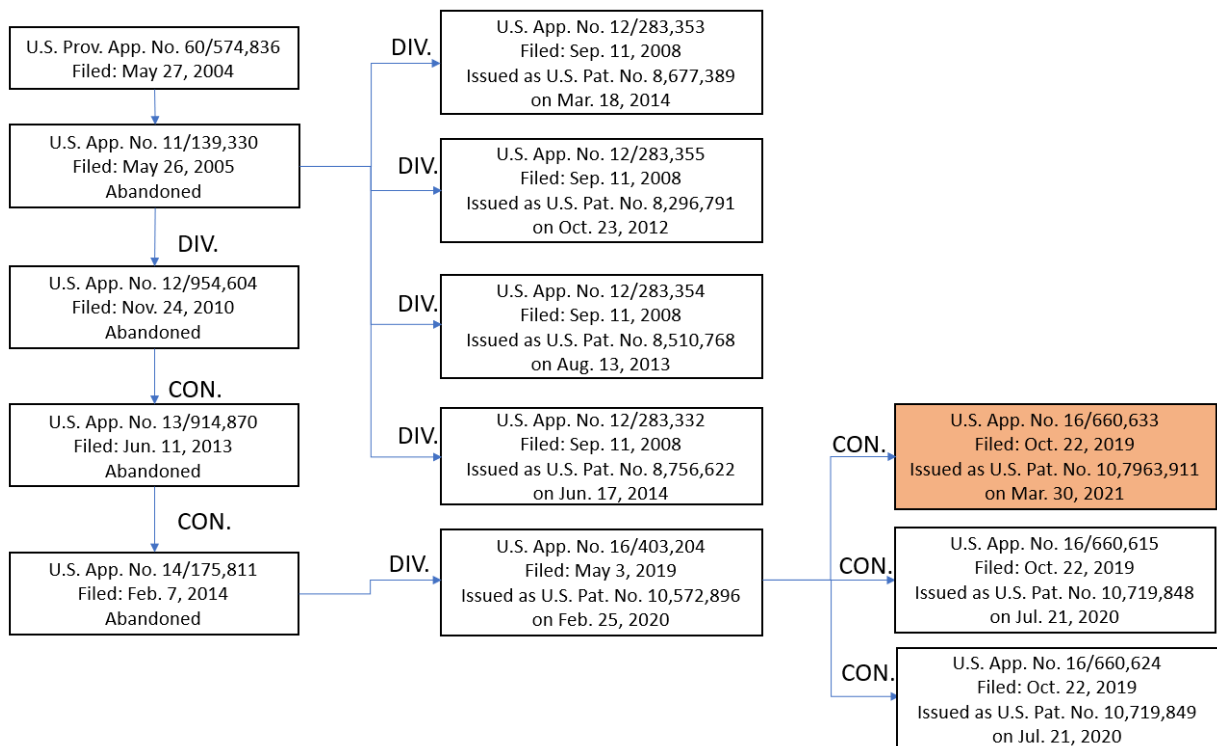
FIG. 10

(EX1007 at FIG. 10; *see also id.*, [0080]-[0082].)

VIII. DISCLOSURE OF CERTAIN CLAIM LIMITATIONS IN THE '911 PATENT CLAIMS AND ITS PARENT APPLICATIONS

49. In my opinion, a person of skill in the art would not have understood that the named inventors of the '911 patent possessed an invention with all of the features recited in the challenged claims (namely, claims 1-17) at the time of the alleged invention.

50. I understand that the '911 patent claims priority to several patents and/or patent applications, as shown in the alleged priority chain below in Demonstrative A (which I also discussed above in Section V):



Demonstrative A

51. However, as I discuss below, it is my opinion that the priority applications, namely U.S. Non-Provisional Application Nos. 11/139,330 (“the ’330 application”), 12/954,604 (“the ’604 application”), 13/914,870 (“the ’870 application”), 14/175,811 (“the ’811 application”), and 16/403,204 (“the ’204 application”) (collectively, “the priority non-provisional applications”) and U.S. Provisional Application No. 60/574,836 (“the ’836 provisional application”), do not convey to a person of ordinary skill in the art that the named inventors were in possession of all the features of the challenged claims at the time these priority applications were filed.

A. The Priority Applications Do Not Support The Video Data Features of The Challenged Claims

52. At least for the reasons I discuss below, it is my opinion that none of the priority applications support all the claim features of the challenged claims.

1. The Priority Applications Do Not Provide Support for the Claimed “Video Data Being Divisible Into Samples” and Related Video Features

53. Independent claim 1 of the ’911 patent recites:

A media measurement method comprising:

receiving, at one or more computers, video data from a monitoring device, the video data being divisible into samples; and

determining, by the one or more computers, a next sample of the video data for attempted identification of a media content item corresponding to the next sample by adjusting a length of a time window of the next sample based at least in part upon a result of an attempt by the one or more computers to identify a first prior sample as corresponding to a media content item stored in a computerized database of reference content;

wherein the time length window of the next sample is changed relative to a time length window of the first prior sample based on the attempt to identify the first prior sample indicating a result that is different than a result of an attempt to identify a second prior sample;

submitting, by the one or more computers, the determined next sample for attempted content identification; and

generating, by the one or more computers, a play stream of content identifications based on results of attempts to identify a series of samples comprising the determined next sample.

(EX1001 at 15:12-34.)

54. As shown above, independent claim 1 of the '911 patent recites features relating to the use of “**video data being divisible into samples**”, i.e., **video data samples**. (EX1001 at 15:12-34.) Specifically, the claim first recites receiving video data, which are divisible into samples, from a monitoring device. (*Id.*) The claim then recites “determining . . . **a next sample of the video data** for attempted identification of a media content item corresponding to the next sample **by adjusting a length of a time window of the next [video] sample** based at least in part upon a result of an attempt by the one or more computers to identify a first prior sample as corresponding to a media content item stored in a computerized database of reference content.” (*Id.*) “[T]he time length window of the next sample is changed relative to a time length window of the first prior sample based on the attempt to identify the first prior sample indicating a result that is different than a result of an attempt to identify a second prior sample.” (*Id.*) The next sample is submitted for attempted content identification. (*Id.*) A play stream of content identifications is generated based on results of attempts to identify a series of samples comprising the determined next sample.

55. Independent claim 11 of the '911 patent recites:

A media measurement method comprising:

receiving, at one or more computers, video data from a monitoring device, the video data being divisible into samples; and

determining, by the one or more computers, a next sample for attempted identification based at least in part upon a result of an attempt by the one or more computers to identify prior samples against a computerized database of known content

wherein determining the next sample comprises selecting a relative time location of the next sample within the video data such that the relative time location of the next sample is selected to be between a relative time location of the first prior sample and a relative time location of the second prior sample based on a result of attempting to identify a first prior sample against the computerized database of known content being different from a result of attempting to identify a second prior sample against the computerized database of known content;

submitting, by the one or more computers, the determined next sample for attempted content identification; and generating, by the one or more computers, a play stream of content identifications based on results of attempts to identify a series of samples comprising the determined next sample.

(EX1001 at 16:8-32.)

56. As shown above, independent claim 11 of the '911 patent similarly recites features relating to the use of “**video data being divisible into samples**”, i.e., **video data samples**. (EX1001 at 16:8-32.) The claim then recites “determining, by the one or more computers, a next sample for attempted identification based at least in part upon a result of an attempt . . . to identify prior samples against a computerized database of known content.” (*Id.*) “[D]etermining the next sample comprises selecting a relative time location of **the next sample within the video data** such that the relative time location of **the next [video] sample** is selected to be between a relative time location of the first prior sample and a relative time location of the second prior sample based on a result of attempting to identify a first prior sample against the computerized database of known content being different from a result of attempting to identify a second prior sample against the computerized

database of known content.” (*Id.*) The next sample is submitted for attempted content identification. (*Id.*) A play stream of content identifications is generated based on results of attempts to identify a series of samples comprising the determined next sample. (*Id.*)

57. As I discuss below, it is my opinion that the priority applications (the ’836 provisional application and the priority non-provisional applications) do not provide written description support for the above video data-related features.

a) The ’836 Provisional Application Does Not Support the Claimed Video Data Features

58. As shown above in Demonstrative A, the ’911 patent claims priority through each of the priority non-provisional applications to U.S. Provisional Application 60/574,836 (“the ’836 provisional application”). However, the ’836 provisional application does not convey to a person of ordinary skill in the art that the named inventors had possession of the video data features at the time the ’836 provisional application was filed (May 27, 2004). This is at least because it does not describe any specific features or embodiments that rely on video data (e.g., video data divisible into samples), let alone any disclosure corresponding to the claimed subject matter.

59. The ’836 provisional application discloses “measurement systems” that purport to involve recording and reporting on individuals’ media usage habits.

According to the disclosures of the '836 provisional application, this is accomplished by using monitoring or detection devices to collect and recognize media broadcast by a playback device. (EX1010 at 2-3.) But, as explained further below, the disclosures of the '836 provisional application fail to adequately describe embodiments that rely on the use of video data, let alone in the manner claimed (as I discussed above in Section VIII.A.1), and therefore the '836 provisional application does not support the claims of the '911 patent.

60. The '836 provisional application does not disclose any features relating to collecting and recognizing the visual aspects of video content. Instead, throughout the '836 provisional application, its disclosures focus on identifying media content by collecting and recognizing *audio* data associated with the content. For example, the '836 provisional application consistently discusses collecting specifically *audio* media content and using “*audio* fingerprinting” technology in order to recognize the content:

[A]n exemplary measurement system employs ***audio fingerprinting and recognition technology*** in order to permit the recognition of media content itself. Such an ***audio fingerprinting and recognition technology*** may, for example, include one or more previously developed technologies of the kind commercialized by Philips Research, Shazam Entertainment Ltd., and MusiKube LLC, among others, in the context of ***song identification***.

(*Id.* at 2 (emphasis added).)

[A]n exemplary method creates a record 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 audio signal portion of the media content*** itself . . .

(*Id.* at 3 (emphasis added).)

[A]n exemplary method includes ***analyzing digital fingerprints of audio energy collected from portable and stationary audio detection devices*** . . .

(*Id.* at 3 (emphasis added).)

[P]articular meta-data may be associated with an ***audio fingerprint*** . . .

(*Id.* at 4 (emphasis added).)

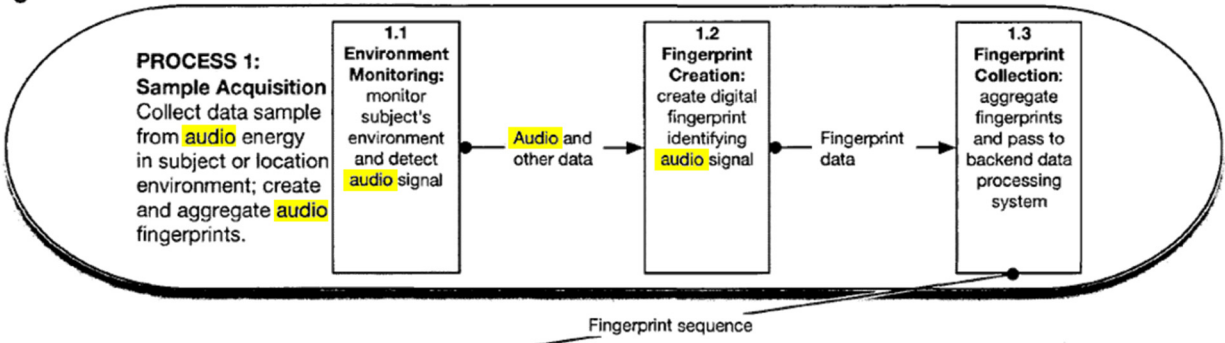
An electronic device is either impermanently affixed to the individual whose media usage is to be monitored, or placed in a stationary location to be monitored. ***This device samples audio energy*** either from the air by means of a microphone or directly from one or more sources of media content to be monitored by means of a direct electrical connection to such device or devices.

(*Id.* at 5 (emphasis added).)

Every few seconds, ***the sampler takes an audio snapshot*** of all detectable media activity within its range, ***extracts features of the audio snapshot, and creates a digital fingerprint*** identifying the signals present.

(*Id.* at 5 (emphasis added).)

Figure 3: Process Flow



(*Id.* at 9 (FIG. 3) (annotated).) Similarly, the '836 provisional application asserts that once the audio fingerprints have been generated, the media content is recognized, or matched, using those audio fingerprints:

[A]n exemplary method . . . *generate[s] digital fingerprints of the audio 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.

(*Id.* at 3 (emphasis added).)

[A] selected sample of *audio fingerprints* from each source *may be matched* . . .

(*Id.* at 4 (emphasis added).)

[A]dditional *audio fingerprints* from a source may be *selected for matching* . . .

(*Id.* at 4 (emphasis added).) The '836 provisional application also discloses methods “to enhance the speed and accuracy of matches of *audio fingerprints*.” (*Id.* at 3 (emphasis added).)

[M]ultiple *audio fingerprints* of differing depth, size, and complexity may be created during the monitoring process in order to enhance the speed and reduce the data processing power required *to match audio fingerprints* . . .

(*Id.* at 4 (emphasis added).)

[A]udio fingerprints representing multiple time windows may be created during the monitoring process in order to enhance the speed and reduce the data processing power required *to match audio fingerprints* . . .

(*Id.* at 4 (emphasis added).) Other purported methods of enhancing matching include “collaborative filtering technologies,” “comparison to known and prearranged sequences of media,” and using “time sequence data” to “correlat[e] successive fingerprint data samples.” (*Id.* at 3.) But in my opinion, a person of ordinary skill in the art would have understood that all of the various disclosed methods, techniques, and/or technologies rely on using audio fingerprints and the ’836 provisional application provides no equivalent or corresponding detailed disclosures that instead rely on using video data samples from the visual portion of monitored media content and identifying content based on analysis of such video data samples including adjusting a time window length of a next sample based on identification results associated with prior samples or performing a backtracking analysis to select a next sample based on identification results associated with prior samples (by, e.g., selecting a sample located between the prior two samples as the next sample), as recited in some variation in the challenged claims.

61. Even where the ’836 provisional application mentions video content, a person of ordinary skill in the art would have understood that those features are

limited to “content including at least some audio elements.” (EX1010 at 2.) Indeed, throughout the ’836 provisional application, it focuses on audio data or only the audio portion of video content:

[A]n exemplary measurement system may enable the monitoring of individuals’ media usage across multiple sources of *audio and/or audiovisual content* . . .

(*Id.* at 2 (emphasis added).)

[A] measurement system is designed to measure individual usage of *audio and audiovisual media* over long periods of time . . .

(*Id.* at 3 (emphasis added).)

Another exemplary method determines the source of transmission or storage of the media played by an individual based on *audio characteristics of that signal* (e.g. recognizes particular types of encoding or compression, or *hears screen refresh on [TV] or computer*, or recognizes *compression or frequency range of FM or AM radio, or CD or DVD*).

(*Id.* at 3 (emphasis added).)

A system is disclosed for measuring individuals' usage of *audio and audiovisual media* in order to understand how media content is used across a range of devices and transmission methods. The system uses portable and stationary devices to *detect audio signals* in the local environment and/or from the media devices to be measured. The system matches *digital fingerprints of the audio signal present in audio and audiovisual media* to the corresponding fingerprints of known media content stored in a database.

(*Id.* at 6.) Thus, in my opinion, the monitored media potentially including visual elements does not change my understanding of each of the exemplary methods described in the '836 provisional application because only the audio portion of video content is analyzed and processed.

62. The '836 provisional application does include a brief, vague, and conclusory portion that mentions video data, but it fails to provide adequate written description support for the claimed video data features of the claims of the '911 patent. (EX1010 at 4.) This generic disclosure states:

Another exemplary method creates a record 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).

These various exemplary methods may involve video rather than audio.

(EX1010 at 4.) In my opinion, this disclosure does not convey to a person of ordinary skill in the art that the named inventors had possession of the video data features at the time the '836 provisional application was filed, for the reasons explained below.

63. First, where the '836 provisional application states that “[t]hese various exemplary methods may involve video rather than audio,” it is unclear what methods the phrase “[t]hese various exemplary methods” refers to. For example, one disclosed exemplary method relates to “detect[ing] the simultaneous playback of more than one media source in a measured environment,” and illustrates such a scenario with an example where content is recognized when a “TV and radio are on at the same time,” or “if music is the background of an advertisement.” (EX1010 at

3.) It is unclear how these examples, which refer to content that only includes an audio portion—content played by radio and music content—would have instead “involve[d] video rather than audio.”

64. Further, the conclusory disclosure related to video appears only around halfway into '836 provisional application, at the end of a section with the heading “Detailed Specification,” and after which are additional disclosures. (EX1010 at 4.) It is therefore further unclear whether the phrase “[t]hese various exemplary methods” refers to the purported methods described prior to this disclosure, in the “Detailed Specification” section, to the later disclosed methods, or something else. Moreover, regardless of the examples to which the conclusory disclosure is directed, the '836 provisional application provides no details whatsoever to a person of ordinary skill in the art as to what changes would be made in order for the “various exemplary methods” to “involve video rather than audio.” (EX1010 at 4.)

65. I have been asked to assume an argument by Patent Owner that there are no substantive differences between audio and video in the context of substituting video data samples for audio data samples in the context of the '836 provisional application. But in my opinion, the '836 provisional application does not support such a conclusion. To the contrary, a person of ordinary skill in the art would have recognized the different challenges and obstacles faced with collection and

recognition of audio versus video (i.e., using audio data samples for content recognition as opposed to using video data samples for content recognition, and for performing claimed steps like “adjusting a length of a time window of the next sample” based on identification results associated with prior video samples (claim 1) and “selecting a relative time location of the next sample within the video data” based on identification results associated with prior video samples (claim 11)).

66. For example, a research competition funded by government organizations at the time, National Institute of Standards and Technology (NIST) and the Advanced Research and Development Activity (ARDA), demonstrates the early stages and challenges of video-data-based content recognition during the 2004-2005 timeframe. (EX1030 at 1-2 (Introduction); EX1031 at 1.) The 2004 research competition, Text REtrieval Conference (TREC) Video Retrieval Evaluation (VID), collectively TRECVID 2004 included participants from worldwide universities and research organizations. (EX1030 at 2 (Table 1).) TRECVID 2004 involved various tasks involving video content including shot boundary determination, story segmentation, feature-extraction, and search. (EX1030 at 1; *see also id.* at 3 (Section 2.1 describing the video content), EX1031 at 1 (Section 2), 5 (Section 3.2 describing the video test data).) The Guidelines for the TRECVID 2004 Evaluation (EX1031) set forth ten features for recognition including “Boat/ship,” “Madeline Albright,”

“Bill Clinton,” “Train,” Beach,” “Basket scored,” “Airplane takeoff,” “People walking/running,” “Physical Violence,” and “Road.” (EX1031 at 2-3 (Section 2.3); EX1030 at 10 (“The features to be detected were defined (briefly) as follows . . .”).) As shown in Table 2 of the TRECVID 2004 overview (EX1030), the search accuracy for identifying this content was less than 10% for eight out of ten features, and about 30% or less for the remaining two features. (EX1030 at 10 (Table 2, last column showing accuracy rate for the various features).) Indeed, another TRECVID presentation from 2005 shows that the average search accuracy between 2003-2005 for identifying content was less than 10%. (EX1032, 9.) The TRECVID documents confirm my understanding that in the 2004-2005 timeframe, a person of ordinary skill in the art would have understood that substituting audio data samples with video data samples for content recognition would not have been straightforward.

67. The '836 provisional application is silent on such issues and how the disclosed audio-based embodiments would have been implemented in the context of video in light of any unique challenges posed by video data.

68. Without any disclosures similarly describing ways of dealing with difficulties unique to the video context or explaining how the disclosures for dealing with difficulties in the audio context would have extended to the video context or could be adapted to the video context, it is my opinion that a person of ordinary skill

in the art would not have understood the named inventors of the '911 patent to have had possession of the claimed subject matter at the time the '836 provisional application was filed.

69. A detailed review of the claims confirms my understanding that the '836 provisional application does not provide adequate support for the claims. As I noted earlier in Section VIII.A.1, claim 1 of the '911 patent recites “receiving, at one or more computers, **video data from a monitoring device, the video data being divisible into samples.**” (EX1001 at 15:12-34.) The claim requires “determining . . . **a next sample of the video data** for attempted identification of a media content item corresponding to the next sample **by adjusting a length of a time window of the next [video] sample**” based on identification results associated with prior video samples. (*Id.*) Subsequently, a play stream of content identifications is generated based on results of attempts to identify a series of samples comprising the determined **next [video] sample.** (*Id.*) Thus, the claimed steps are tied to video data or premised on the use of video data samples.

70. As I also noted earlier in Section VIII.A.1, claim 11 recites a method that is similarly premised on the use of video data samples. (EX1001 at 16:8-32.) For example, independent claim 11 of the '911 patent similarly recites features relating to the use of “**video data being divisible into samples**”, i.e., **video data**

samples. (EX1001 at 16:8-32.) The claim then recites “determining, by the one or more computers, a next sample for attempted identification based at least in part upon a result of an attempt . . . to identify prior samples against a computerized database of known content.” (*Id.*) “[D]etermining the next sample comprises selecting a relative time location of **the next sample within the video data** such that the relative time location of **the next [video] sample** is selected to be between a relative time location of the first prior sample and a relative time location of the second prior sample based on a result of attempting to identify a first prior sample against the computerized database of known content being different from a result of attempting to identify a second prior sample against the computerized database of known content.” (*Id.*) The next sample is submitted for attempted content identification. (*Id.*) A play stream of content identifications is generated based on results of attempts to identify a series of samples comprising the determined next sample. (*Id.*)

71. But, as explained above, the ’836 provisional application does not disclose using data samples from the visual aspects of video content (e.g., collecting video data and/or generating video data samples from video content that the user is viewing), nor using such video data samples to identify media content. Thus, a person of ordinary skill in the art would not have understood that the ’836

provisional application discloses these limitations of claims 1 and 11 of the '911 patent.

72. Claims 2-10 depend from claim 1 and claims 12-17 depend from claim 11. (EX1001 at 15:35-16:7; 16:33-64.) These claims therefore lack support for at least the same reasons as claims 1 and 11.

b) The '330, '604, '870, and '811 Priority Non-Provisional Applications Do Not Provide Adequate Support For The Challenged Claims

73. The priority non-provisional applications, *i.e.*, the '330 application, '604 application, '870 application, and '811 application, share substantially the same as-filed specifications. (EX1011 at 577-599 ('330 application specification), 610-617 ('330 application abstract and drawings), EX1012 at 276-298 ('604 application specification), 311-318 ('604 application abstract and drawings), EX1013 at 21-43 ('870 application specification), 56-63 ('870 application abstract and drawings), EX1014 at 420-442 ('811 application specification), 455 ('811 application abstract), 460-466 ('811 application drawings).) Exhibits 1017, 1018, and 1019 are computer-generated redlines comparing the as-filed specifications of the '604 application, '870 application, and '811 application to the as-filed specification of the '330 application. As shown in exhibits 1017-1019, other than the changes to the "Cross Reference to Related Applications" sections in each of the priority applications, the specifications

of these priority non-provisional applications are substantively identical.⁴ Thus, while my opinions below primarily focus on the lack of support for the challenged '911 patent claims in the '330 application's disclosure, it is my opinion that a person of ordinary skill in the art would have understood that the disclosures of the '604 application, '870 application, and '811 application are deficient for the same reasons.

74. In particular, the disclosures of these priority non-provisional applications do not describe any features that rely on *video* data samples (e.g., receiving **video data** from a monitoring device, the **video data being divisible into samples**) and consequently, the subsequent related features set forth in the claims (e.g., determining **a next sample of the video data** for attempted identification of a media content item corresponding to the next video sample **by adjusting a length of a time window of the next video sample** as set forth in claim 1 and determining the next video sample comprises **selecting a relative time location of the next**

⁴ Because the specifications of the applications are substantively identical, I use the specification of the '330 application as representative of the specifications of the priority applications. Thus, my citations below refer only to the specification of the '330 application.

video sample within the video data such that the relative time location of the next sample is selected to be between a relative time location of the first prior sample and a relative time location of the second prior sample as set forth in claim 11). Instead, as I demonstrate below, the disclosures describe only embodiments based on **audio** data samples.

75. The '330 application uses the term “audio” 37 times, including, for instance, in the specification (33 times), abstract (2 times), and figures (2 times). (*See generally* EX1011 at 577-599, 610-617.) It also is replete with terms that are related only to audio (to the exclusion of video): song (24 times, excluding figures), radio (11 times, excluding figures), microphone (10 times, excluding figures), headphones (10 times, excluding figures), album (5 times, excluding figures), sound (2 times), music (2 times). By contrast, the specification of the '330 application uses the term “video” only 4 times. (*See generally* EX1011 at 577-599, 610-617.) The particular claim term “video data sample” appears only once outside of the claims in the '330 application specification, in the “BACKGROUND” section:

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.

(EX1011 at 578 ([0005]).) 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” (e.g., by sampling pixels in a video frame) so in my opinion, a person of ordinary skill in the art would not have understood this portion of the ’330 application to sufficiently describe such an embodiment.

76. The ’330 application specification also broadly refers to a “content of a media sample” in its “SUMMARY” section: “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.” (EX1011 at 578 ([0006]).) In

my opinion, this reference to a “media sample” is also insufficient to support the claims of the ’911 patent.

77. First, the statement refers to “enhanc[ing] recognition . . . of the content of a media sample by analyzing information exogenous to the sample.” (*Id.*) That is, a person of ordinary skill in the art would have understood that this statement is not directed at content recognition using the media sample but rather enhancing the content recognition by using information “exogenous to” (e.g., external to) the sample. A person of ordinary skill in the art would not have understood from this statement that the ’330 application specification is describing a content recognition process using video data samples, let alone the detailed claim steps using video data samples.

78. Second, the remaining portion of the specification of the ’330 patent is exclusively focused on content recognition methods using audio data samples (obtained from audio content or video content, as I discuss above and below). Thus, a person of ordinary skill in the art would have understood a “media sample” in the context of the overall disclosure of the ’330 application specification to be an audio sample obtained from audio content or video content.

79. Moreover, I have been asked to consider a scenario where the use of “video data sample” in the background section or the use of a “media sample” in the

summary section could be interpreted to vaguely and generally refer to a content recognition process using video data samples. In my opinion, however, even under this interpretation, there is simply no associated detailed disclosure of a content recognition process based on the use of video data samples, as required by the claims. For example, there is no support in the '330 application for performing the claimed steps like “adjusting a length of a time window of the next [video] sample” based on identification results associated with prior video samples (claim 1) and “selecting a relative time location of the next [video] sample within the video data” based on identification results associated with prior video samples (claim 11). (*See generally* EX1011.)

80. And while the '330 application specification uses certain video-related terms (*e.g.*, television, TV, movie, etc.), in my opinion, it is clear in view of the specification as a whole that these disclosures refer to analyzing and processing the audio data portion of captured video content. In particular, as shown in the reproduced portions of the '330 application specification below, while this portion mentions media played by a DVD player or television (which may include video content) and mentions media content such as “movie, commercial, television show, video game etc.”), the specification only describes capturing the audio sample from this media content:

When the audience member is consuming media from media player 170 *a* without the use of headphones 160 *a* (i.e. if media player 170 *a* has a speaker), monitor 101 captures audio energy acoustically through microphone 105. However, *when the audience member desires to use headphones 160 a to receive media content from media player 170 a* (e.g. an MP3 player, iPod, CD player, **DVD player, television**, radio, computer, or other media player) he or she can plug headphones 160 *a* into headphone port 103 and plug media player 170 *a* into media player port 102. In that case, *monitor 101 a captures audio energy through media player port 102.*

(EX1011 at 580-581 ([0017]).)

As another example, **headphones** 160b are adapted to receive a content signal 171 via wireless transmission 171 from media player 170b. Monitor 101 b is adapted to receive the same content signal 171 from media player 170b. In some embodiments, a syncing process allows monitor IO I b to adapt along with headphones 160b and media player 170b to changes in signal 171 **so that monitor 101b continues to receive the same signal as headphones** 160b. In some embodiments, syncing may be accomplished via a wireless and/or automatic process, such as providing a monitor that responsively changes

tuning based on tuning changes between the media player and the **headphones**.

(EX1011 at 581 ([0018]).)

Audio data captured by monitors 101, along with associated monitor data (e.g. device ID) (collectively, “audio data/monitor data” 144) may be uploaded to system 100 through upload port 104, directly, via memory device transfer (e.g. flash memory card, floppy disk, CD, etc.) or through a network such as network 150. **The audio data uploaded from a monitor 101 may be raw audio data, or may be audio data that has undergone one of a variety of levels of processing prior to upload.** For example, the uploaded audio data may include parameters useable to calculate **audio signatures (sometimes referred to as audio fingerprints) and landmarks or other values useful in the content recognition process.** Alternatively, the audio data may comprise pre-calculated signatures, landmarks, or other data useful for content recognition. On the other hand, in some embodiments, **the uploaded audio data may simply represent a raw signal corresponding to audio energy received at monitor 101.**

(EX1011 at 581 ([0019]).)

Content 141 may include media files associated with a known piece of media content (e.g. a song, *movie*, *commercial*, *television show*, *video game*, etc.). However, it will be understood by those skilled in the art that “content” in the sense of the known content 141 received by system 100 may, in some embodiments, simply refer to data about signals representing the content and does not necessarily refer to a stored version of the media content itself; e.g. “content” may be stored signatures and landmarks *derivable from a song’s audio*, but not necessarily an audio file useable to play the song itself.

(EX1011 at 582-583 ([0022]); *see also id.* at 583 ([0023]) (describing module 121 “divid[ing] **audio data** into samples and determin[ing] which samples should be submitted for further content recognition testing”), 586-587 ([0029]) (describing processing of audio data for content identification).)

At step 210, *characteristics of the audio data are analyzed* to attempt to identify the source of transmission or storage of the media played by an individual based on audio characteristics of that signal. For example, the process attempts to recognize particular types of encoding or compression, or *identifies sound associated with screen refresh on a television or computer monitor*, or recognizes compression or frequency range of FM or AM radio, or CD or DVD.

(EX1011 at 587 ([0030]); *see also id.* at 587 ([0031]) (describing content identification based on “**audio data** in the tested sample”), 587-588 [0032] (describing adjusting sample control and selection parameters for content identification including increasing resolution “when recent identification results suggest an **extended period of silence or absence of known content**”).)

Play List 560 contains a row for each internally continuous media consumption event. The start log time and end log time for each such event are listed in each row. *If an entire piece of content (e.g. a TV program segment, or an entire TV ad) is listened to without interruption, that would correspond to a row of Play List 560.*

(EX1011 at 594 ([0049]).) Thus, in my opinion, a person of ordinary skill in the art would have understood that none of the embodiments in the '330 application describe, in any way, the video data sample features featured in the claims of the '911 patent.

81. Instead, in my opinion, a person of ordinary skill in the art would have understood that the embodiments disclosed in the '330 application (and the remaining priority non-provisional applications) are solely focused on recognizing content based on audio data (e.g., the use of audio data samples to identify content) and subsequent processing of the content recognition results. (*See generally* EX1011

at 577-599, 610-17.) The '330 application does not disclose any embodiments related to, and thus does not provide any support for, the claimed feature of generating a play stream of content identifications based on video data samples, let alone the other related claim features.

82. Further, in my opinion, a person of ordinary skill in the art 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. In my opinion, a person of ordinary skill in the art would have understood that a system that relies on capturing audio would operate in different environments and be subject to different obstructions, challenges, and limitations to a system that relies on capturing video. As I explained above in Section VIII.A.1.a, government-funded efforts at the time demonstrated around a 10-30% accuracy rate for video content recognition (citing EX1030 at 1-3, 10; EX1031 at 1-3, 5; EX1032 at 9). Moreover, multiple disclosures in the '330 application describe peculiarities related to audio content that have no clear analog in the video context.

[A]n 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.

(EX1011 at 577 ([0002]).)

[O]ther 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.

(EX1011 at 578 ([0006]).)

When the audience member is consuming media from media player 170 *a* without the use of headphones 160 *a* (i.e. if media player 170 *a* has a speaker), *monitor 101 captures audio energy acoustically through microphone 105*. However, when the audience member desires to use headphones 160 *a* to receive media content from media player 170 *a* (e.g. an MP3 player, iPod, CD player, DVD player, television, radio, computer, or other media player) he or she can plug headphones 160 *a* into headphone port 103 and plug media player 170 *a* into media player port 102.

(EX1011 at 580 ([0017]).)

Other differences corresponding to *content types (e.g. speaking audio versus music audio)* might also be exploited based on prior content identification results to select a most efficient recognition algorithm from a plurality of algorithms.

(EX1011 at 584-585 ([0025]).)

83. The '330 application suffers from the same deficiency described above with respect to the '836 provisional application, namely, it provides no corresponding disclosures describing how difficulties unique to the video context are addressed. (*See* Section VIII.A.1.a, above.) Likewise, nowhere in the '330 application is an explanation for how its disclosures for dealing with difficulties in the audio context would have extended to or could be adapted to the video context.

(*Id.*)

c) The Combination of The '836 Provisional Application's Disclosure with the Priority Non-Provisional Applications Does Not Support the Video Data Features of the Challenged Claims

84. I have been asked to consider whether a person of ordinary skill in the art would consider the combination of the disclosure of the '836 provisional application with the disclosure of the priority non-provisional applications (i.e., the '330, '604, '870, and '811 applications I discussed above) as providing adequate support for the video-related claims.

85. In my opinion, a person of ordinary skill in the art would have understood that the disclosures of the priority non-provisional applications, when considered in combination with the disclosures of the '836 provisional application (including its vague video-data-related paragraph), do not provide support for the video data features and related features of the challenged claims of the '911 patent.

86. As an initial matter, I note that the '330, '604, '870, and '811 priority non-provisional applications do not reference the above-noted vague paragraph from the '836 provisional application concerning video, such that a person of ordinary skill in the art would have consulted that particular paragraph in reviewing and understanding the audio-based disclosures of the priority non-provisional

applications.⁵ (*See generally* EX1011 at 577-599 ('330 application specification); EX1012 at 276-298 ('604 application specification); EX1013 at 21-43 ('870 application specification); EX1014 at 420-442 ('811 application specification).) Moreover, in my opinion, a person of ordinary skill in the art would not have understood that the disclosed audio-based embodiments in the priority non-provisional applications could have been implemented in the video context during the relevant timeframe for all the reasons I explained above. (*See* Section VIII.A.1.a-b, above (citing EX1030 at 1-3, 10; EX1031 at 1-3, 5; EX1032 at 9).) There is simply no guidance provided to a person of ordinary skill in the art in the disclosures of any of the priority non-provisional applications that would have conveyed to such a person that the named inventors were in possession of the claimed invention

⁵ As I discussed above in Section VIII.A.1.a, there is no other disclosure in the '836 provisional application that describes using video data.

including the video data features during the timeframe when the priority non-provisional applications were filed (*e.g.*, in the 2005-2014 timeframe).⁶

87. Thus, in my opinion, a person of ordinary skill in the art would have understood that like the '836 provisional application, the priority non-provisional applications (alone or in combination with the '836 provisional application) fail to provide support for the video data features set forth in the claims of the '911 patent.

2. Additional Materials Referenced in The Priority Non-Provisional Applications Do Not Support the Video Data Features of the Claims of the '911 Patent

88. The '330 application, and similarly the other priority non-provisional applications, purport to be “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.” (*See, e.g.*, EX1011 at 584-585 ([0025]).) The '330 application then mentions several

⁶ As shown in Demonstrative A in Section V and at the outset of this section (*i.e.*, Section VIII), the '330 application was filed in 2005, the '604 application was filed in 2010, the '870 application was filed in 2013, and the 811 application was filed in 2014.

patent applications (which I understand are being submitted as Exhibits 1020-1023) that allegedly describe content recognition methods without specifying any particular portions of the documents as being relevant.

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. One such example is described in **U.S. patent application Ser. No. 09/839,476 entitled “System and Methods for Recognizing Sound and Music Signals in High Noise and Distortion” by Wang et al. and published Jun. 27, 2002 with publication number US2002/0083060**. Aspects of other such examples are described in the following published PCT applications by Koninklijke Philips Electronics N.V.: **“Fingerprinting Multimedia Contents” (WO2004/044820, published May 27, 2004), “Fingerprint Extraction” (WO2004/030341 published Apr. 8, 2004), and “Improvements in and Relating to Fingerprint Searching” (WO2004/040475 published May 13, 2004).**

(EX1011 at 584-585 ([0025]).)

89. To the extent the above-noted documents disclose features related to using video data samples for content recognition, none of the priority non-provisional applications make any mention of specific video recognition-related features in these documents or how their use may be incorporated into the disclosed embodiments of the priority non-provisional applications. Instead, the priority non-provisional applications merely generically state that the above-noted documents are examples of content recognition methods and algorithms. (EX1011 at 584-585 ([0025]).) In my opinion, a person of ordinary skill in the art would not have understood from this limited, non-specific disclosure that specific video-data-based content recognition features are what was intended to be considered in these documents, especially since, as I discussed above in Section VIII.A.1.b, the priority non-provisional applications are limited to audio-based content recognition embodiments.

90. In other words, in my opinion, the priority non-provisional applications simply do not identify with any specificity what portions of the above-noted documents (EXS1020-1023) relate to using video data samples and where such features can be found within the documents. In this regard, I note that one of the references identified in the '330 application, namely U.S. Patent Publication No.

2002/0083060, relates to identification of audio and provides no disclosures related to identification of video content. (*See generally* EX1020.)

91. Indeed, none of the documents (EXS1020-1023) exclusively relate to content identification based on video data samples, so in my opinion, a person of ordinary skill in the art would have understood that only the features of the above-noted documents related to content recognition based on audio data were relevant to the '330 application (particularly because content identification based on audio data samples was the focus of the disclosures of the priority non-provisional applications). Thus, a person of ordinary skill in the art reviewing the documents mentioned in paragraph [0025] of the '330 application specification in light of the disclosure of the priority non-provisional applications would have had to speculate without any guidance as to what teachings should be combined, and how to accomplish such combinations, in order to adapt the disclosed embodiments to the use of video data.

3. New Disclosures Added to The '204 Application That Issued as The '896 Patent

92. In my opinion, the addition of new disclosures to the as-filed specification of the '204 application, which is another non-provisional application to which the '911 patent claims priority, further confirms my understanding regarding the inadequate disclosures in the priority non-provisional applications, *i.e.*,

the '330 application, '604 application, '870 application, and '811 application. In particular, I note that the '204 application added the following single generic paragraph regarding video:

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.

(EX1015 (file history of '204 application) at 232 (¶[0061]).)

93. While the new paragraph above appears similar to the short paragraph regarding video in the '836 provisional application (EX1010 at 4) I discussed above in Section VIII.A.1.a, at least the reference to the “**FIGS. 1-6**” embodiments and the last sentence stating that the “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**” is altogether new. (*See also* EX1028.) Figures 1-6 and their corresponding detailed descriptions in the ’204 application were not included in the ’836 provisional application, and so in my opinion, extrapolating the application of video data to those detailed embodiments (to the extent a person of ordinary skill in the art would have understood it as such given the unclear language) is not supported by the ’836 provisional application. (*See generally* EX1010 at 2-9; EX1011 at 611-616 (FIGS. 1-6).) The new paragraph above is also not included in any of the priority non-provisional applications. (*See generally* EX1011 at 577-599 (’330 application specification); EX1012 at 276-298 (’604 application specification); EX1013 at 21-43 (’870 application specification); EX1014 at 420-442 (’811 application specification).) Thus, in my opinion, there is no support in the ’836 provisional application or the other priority non-provisional applications for video data being relied upon in the detailed embodiments of “FIGS. 1-6,” to the extent this new paragraph can be understood in the ’204 application.

94. Thus, in my opinion, the '204 application contains new disclosures not present in any earlier priority application.⁷

⁷ To be clear, I take no position regarding whether the newly added disclosures to the specification of the '204 application are sufficient to provide support for the video-related claim features.

IX. THE PRIOR ART DISCLOSES AND/OR SUGGESTS ALL RECITED FEATURES OF CLAIMS 1-17 OF THE '911 PATENT

A. Grounds 1 and 2:

Ground 1: The Combination of Steuer750 and Eldering Discloses and/or Suggests Claims 1-17

Ground 2: The Combination of Steuer389 and Eldering Discloses and/or Suggests Claims 1-17

95. **Ground 1** relies on the combination of Steuer750 (EX1005) and Eldering (EX1007) whereas **Ground 2** relies on the combination of Steuer389 (EX1006) and Eldering. However, Steuer750 and Steuer389 have substantively identical specifications, as I discussed above in Section VII.A. As such, in the interest of efficiency and clarity, I discuss both grounds collectively, with citations provided to both Steuer750 and Steuer389. That is, I provide citations below to Steuer750 (to support **Ground 1**) and Steuer389 (to support **Ground 2**). In doing so, I am not relying on Steuer750 for **Ground 2** or Steuer389 for **Ground 1**.

96. As I noted above in Section VII.A, Steuer389 further includes claims 1-17 which are nearly identical to challenged claims 1-17 here, except the word “audio” in claims 1-17 of Steuer389 is substituted with “video” in the challenged

claims.⁸ (*Compare* EX1001 at claims 1-17 *with* EX1006 at claims 1-17; *see also* EX1009 (showing a comparison of the claims); *see also* my discussion above in Section VII.A regarding some of the differences in the claims.) Thus, under **Ground 2**, I rely on all the same disclosures in Steuer389 corresponding to the relied-upon disclosures in Steuer750 under **Ground 1** (and as noted above, provides citations to both Steuer750 and Steuer389), but further rely on claims 1-17 of Steuer389 as they almost verbatim disclose challenged claims 1-17 except the “video” feature. The citations to the claims of Steuer389 are applicable only to **Ground 2**. The reliance on Eldering is the same in both **Ground 1 and 2**, as is my opinion regarding the rationale for combining Eldering with Steuer750 or Steuer389.

1. Claim 1

a) A media measurement method comprising:

97. I have been asked to assume that the preamble of claim 1 is limiting. Under that assumption, it is my opinion that Steuer750 and Steuer389 both independently disclose the limitation therein. (EX1005 at Title, Abstract, FIGS. 1-

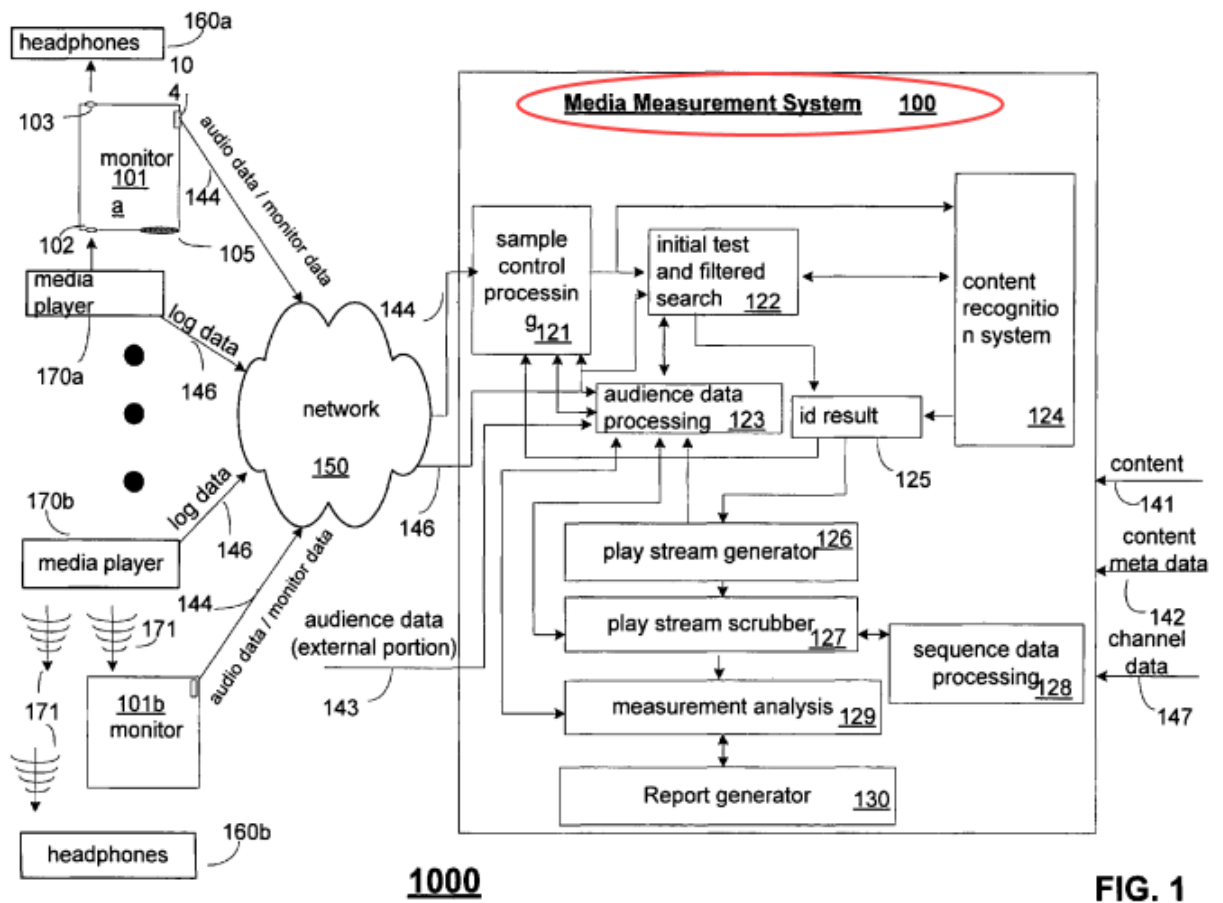
⁸ Additionally, limitations 1.e, 6.e, and 10.e of the challenged claims are not included in the Steuer389 claims.

2, claims 1, 19, 35, 40, 44, 48, 61-62; EX1006 at Title, Abstract, FIGS. 1-2, claims 1-11.) As an initial matter, both Steuer750 and Steuer389 claim “[a] media measurement method[.]” (EX1005 at claims 1, 19, 35, 40, 44, 48, 61-62; EX1006 at claims 1, 11.) Steuer750 and Steuer389 also refer to a media measurement method in their disclosures.

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 . . . 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.

(EX1005 at [0006]; *see also id.* at Abstract, [0002]; EX1006 at 1:60-2:10; *see also id.* at Abstract, 1:14-32.)

98. Steuer750 and Steuer389 further disclose a “Media Measurement System 100” (Figure 1) for carrying out its methods and “illustrates a media measurement method” in Figure 2 (EX1005 at [0009]; EX1006 at 2:26-28).



(EX1005 at FIG. 1 (annotated); *see also id.*, [0008], [0016]-[0017]; EX1006 at FIG. 1; *see also id.*, 2:23-25, 2:66-3:25.)

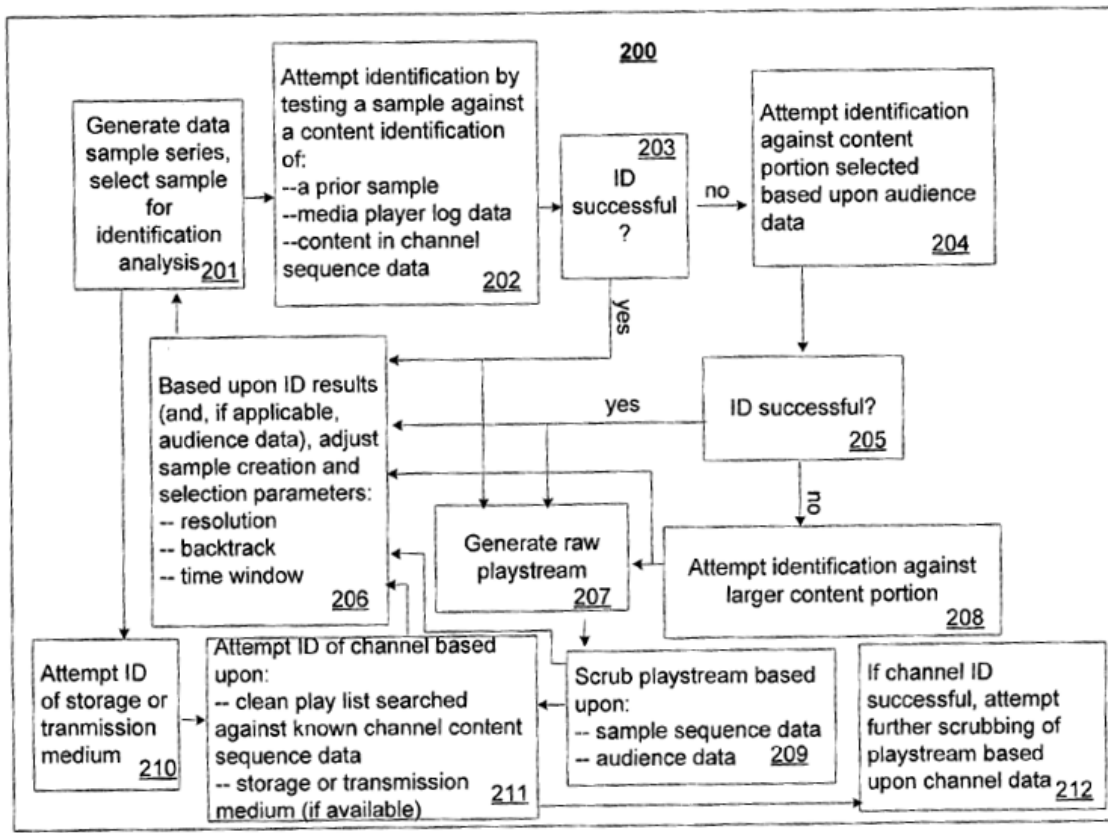


FIG. 2

(EX1005 at FIG. 2; *see also id.*, [0009], [0034]; EX1006 at FIG. 2; *see also id.*, 2:26-28, 6:58-7:3.)

99. Thus, Steuer750 and Steuer389 both disclose “[a] media measurement method,” as claimed. (*See also* Sections IX.A.1.b–f, below.)

- b) receiving, at one or more computers, video data from a monitoring device, the video data being divisible into samples; and**

100. In my opinion, Steuer750 in combination with Eldering and Steuer389 in combination with Eldering disclose or suggest this limitation.

101. Steuer750 and Steuer389 disclose receiving, at one or more computers (e.g., the media measurement system 100 implemented by computer system 700), audio data from monitors 101 (“a monitoring device”), the audio data being divisible into samples.

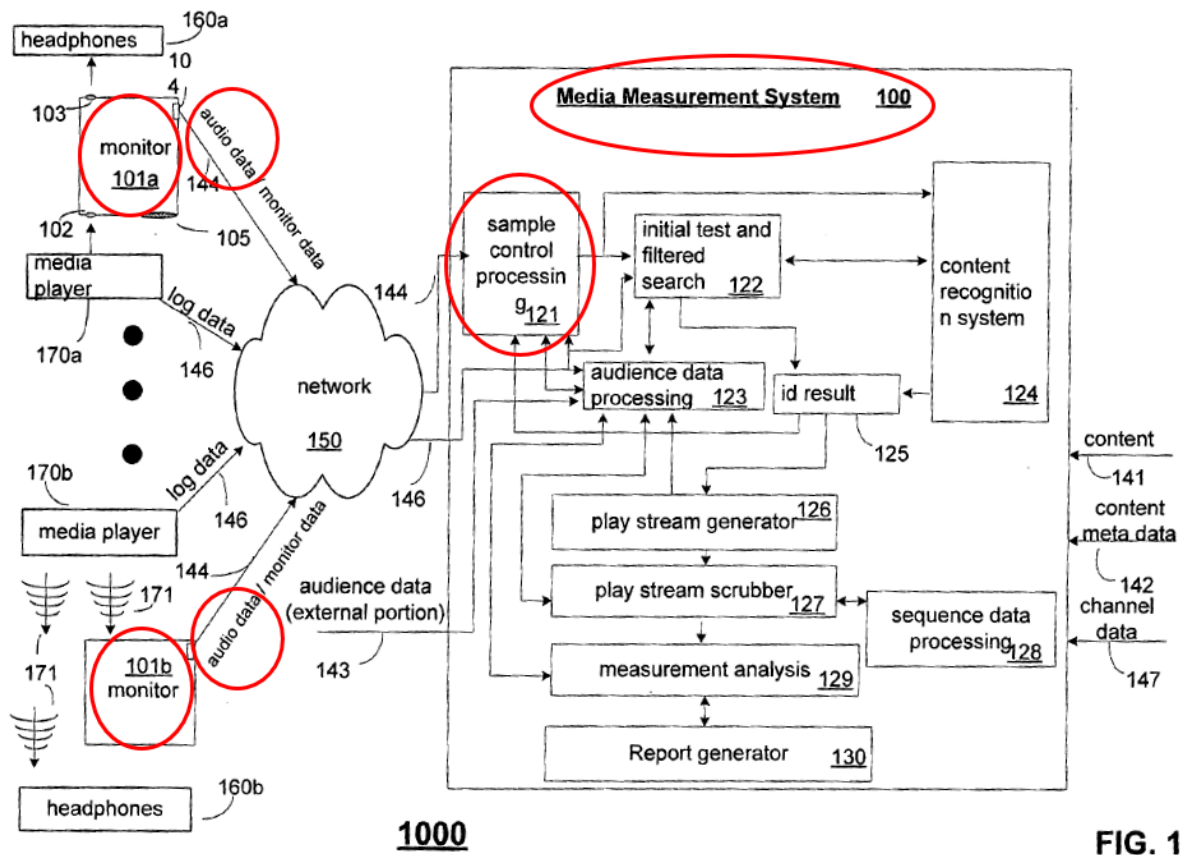
Turning to the details of system 100, audio data and monitor data (e.g. device ID) is received at sample control processing module 121. Module 121 provides monitor data, portions of which might be audience data, to processing module 123. **Module 121 also divides the audio data into samples and determines which samples should be submitted for further content recognition testing. In alternative embodiments, the audio data may already have been divided into samples prior to receipt by system 100 (e.g. at monitor 101).** In such alternatives, a module such as module 121 might either use the provided divisions or combine the data and then re-divide it into samples as determined by the system. Module 121 may be used either to submit a sample to initial test and filtered search module 122 or to content recognition system 124.

(EX1005 at [0023], [0034], [0066], [0016]-[0022], [0024]-[0031], FIGS. 1, 2, 4, 7; claims 1, 6, 19, 40, 44; EX1006 at 4:49-62, 6:58-7:3, 13:58-14:15, 2:66-4:48, 4:63-6:27, FIGS. 1, 2, 4, 7.)

102. For example, Steuer750 discloses and claims, albeit in the “audio data” context, “receiving audio data from a monitoring device, the audio data being divisible into a series of samples.” (EX1005 at claim 19.)

103. Steuer389’s claim 1 similarly recites almost identical language as claim limitation 1.b except it substitutes the claimed video data with audio data: “receiving, at one or more computers, audio data from a monitoring device, the audio data being divisible into samples.” (EX1006 at claim 1.)

104. Steuer750 and Steuer389 disclose that “audio data and monitor data (e.g. device ID) is received [from monitoring device 101] at sample control processing module 121 [of system 100],” where “[m]odule 121 [] divides the audio data into samples” or “the audio data may already have been divided into samples prior to receipt by system 100 (e.g. at monitor 101).” (EX1005 at [0023]; EX1006 at 4:49-62.)



(EX1005 at FIG. 1 (annotated); EX1006 at FIG. 1.)

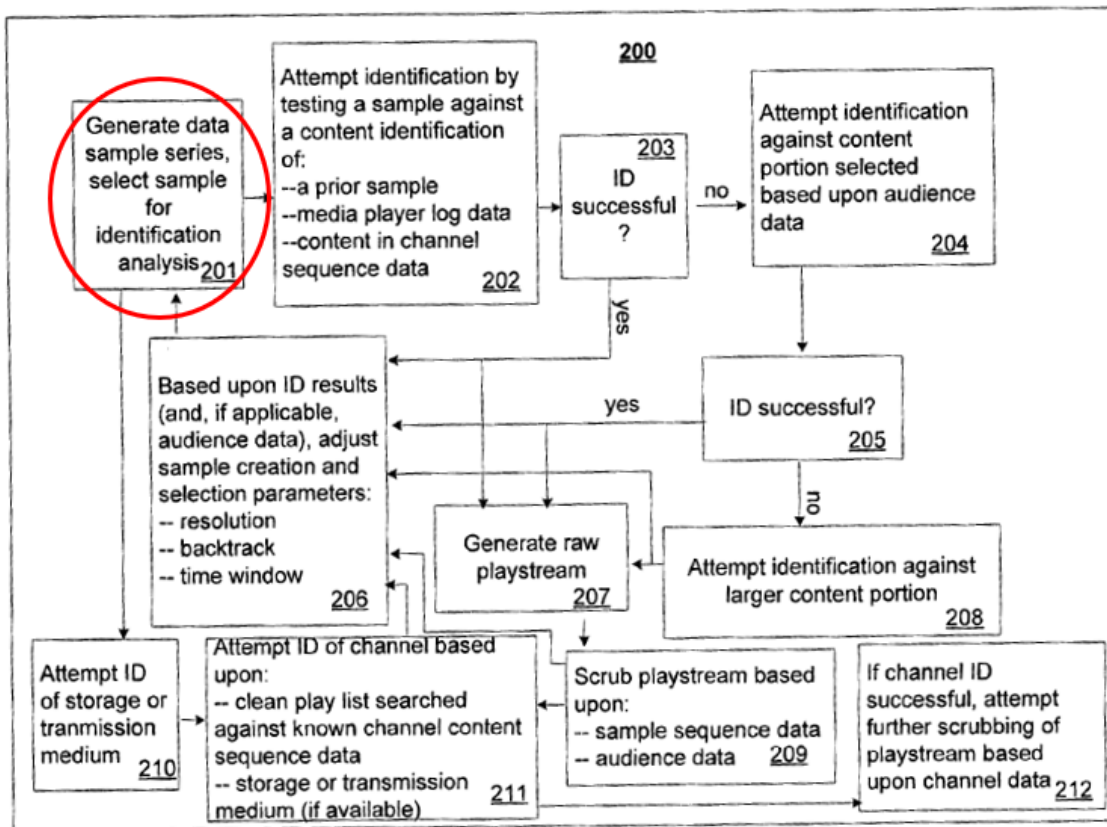


FIG. 2

(EX1005 at FIG. 2 (annotated); EX1006 at FIG. 2.)

105. The disclosures of Steuer750 and Steuer389 are limited to media measurement methods/systems for receiving **audio** data from a monitoring device for content identification and performing subsequent data processing steps based on the content identification results. (EX1005 at Abstract, [0006], [0017]-[0019], [0022]-[0023], [0034]-[0035]; EX1006 at Abstract, 1:60-2:10, 3:4-54, 4:13-62,

6:58-7:16.)⁹ Steuer750 and Steuer389 do not disclose that the received data is “video” data divisible into samples. However, as discussed below, it would have been obvious to modify Steuer750’s and Steuer389’s disclosed media measurement methods such that the received data is video data divisible into samples for the purposes of video content recognition (instead of, or in addition to, receiving audio data) based on the teachings of Eldering.

106. Similar to Steuer750 and Steuer389, Eldering is directed to methods and systems used to identify media content (e.g., advertisements and programs) being consumed by users by matching samples of the consumed media content with a content database.

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**

⁹ Indeed, the ’330 application published as Steuer750 and thus Steuer750 is limited to audio-based embodiments for the reasons I discussed in Section VIII.A.1.b. The same is true for Steuer389 as its disclosure is nearly identical to that of Steuer750. (EX1008.)

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.

(EX1007 at Abstract.)

In the absence of cue tones, such as broadcaster supplied cue tones, indicating the boundaries of advertisement breaks another means of detecting the display of an advertisement is required. One method includes calculating features about an incoming video stream. These features may include color histograms, color coherence vectors (CCVs), and **evenly or randomly highly subsampled representations of the original video (all known as fingerprints).** The fingerprints of the incoming video stream are compared to a database of fingerprints for known advertisements, video sequences known to precede commercial breaks (ad intros), and/or sequences known to proceed commercial breaks (ad outros). **When a match is found between the incoming video stream and a known advertisement or ad intro, the incoming video stream is associated with the known advertisement and/or ad intro and a targeted advertisement may be substituted.**

The fingerprint of the incoming video stream (calculated fingerprint) may be compared to a plurality of fingerprints for known entities (e.g., ads, intros) within the database (known fingerprints). The comparison may be done based on small segments of a video stream at a time. A determination is made as to whether the calculated fingerprint and the known fingerprints within the database exceed some threshold level of dissimilarity. If the comparison exceeds the threshold for certain known fingerprints within the database, the comparison of the calculated fingerprint to those known fingerprints stops for the time being. For those known fingerprints that the comparison was below the threshold level of dissimilarity the comparison continues. At each step of the comparison those known fingerprints exceeding the threshold level of dissimilarity cease. The process continues until one of the known fingerprints has a comparison that exceeds a threshold level of similarity (indicating a match) or the comparison of all of the known fingerprints within the database exceed the dissimilarity threshold at which point the video stream is not associated with any of the known fingerprints.

Calculating fingerprints for the incoming video stream and comparing to a database of fingerprints of known entities can also be used to detect certain programs and/or scenes and to record, bookmark or stop recording the programs and/or scenes.

(EX1007 at [0008]-[0010]; *see also id.*, [0001]-[0007], [0072]; EX1005 at [0002] (similarly describing “track[ing] usage of media” including “tracking consumption of advertisements”), [0049] (describing detection of “ad49”), [0053] (describing detection of advertisements “ad23” and “ad25”); EX006, 1:14-32, 10:5-27, 11:10-26.)

107. As noted above, Eldering discloses that the incoming content data is divisible into samples. (EX1007 at [0008] (describing the use of “evenly [] subsampled representations of the original video” for the content identification process), [0070]-[0076] (describing content sampling examples), [0101], FIGS. 7-9.)

108. Eldering discloses identifying content using video data or audio data. (EX1007 at [0008]-[0010] (describing content identification using video data, as shown above), [0050], [0064], [0086].)

Automatic techniques for detecting advertisements (or advertisement blocks) may include detecting aspects (features) of the video stream that indicate an advertisement is about to be displayed or is being displayed (feature based detection). **For example, advertisements are often played at a higher volume than programming so a sudden volume increase (without commands from a user) may indicate an advertisement.** Many times several dark monochrome (black) frames of video

are presented prior to the start of an advertisement so the detection of these types of frames may indicate an advertisement. The above noted techniques may be used individually or in combination with one another. These techniques may be utilized along with temporal measurements, since commercial breaks often begin within a certain known time range. However, these techniques may miss advertisements if the volume increases or if the display of black frames is missing or does not meet a detection threshold. Moreover, these techniques may result in false positives (detection of an advertisement when one is not present) as the programming may include volume increases or sequences of black frames.

(EX1007 at [0050] (disclosing audio data like volume level being indicative of content to be identified).)

In some countries, commercial break intros are utilized to indicate to the viewers that the subsequent material being presented is not programming but rather sponsored advertising. These commercial break intros vary in nature but may include certain logos, characters, or other specific video and audio messages to indicate that the subsequent material is not programming but rather advertising. The return to programming may in some instances also be preceded by a commercial break outro which is a short video segment that indicates the return to programming. In some cases the intros and the outros may be the

same with an identical programming segment being used for both the intro and the outro. Detecting the potential presence of the commercial break intros or outros may indicate that an advertisement (or advertisement block) is about to begin or end respectively. If the intros and/or outros were always the same, detection could be done **by detecting the existence of specific video or audio**, or specific logos or characters in the video stream, or by detecting specific features about the video stream (e.g., CCVs). However, the intros and/or outros need not be the same. The intros/outros may vary based on at least some subset of day, time, channel (network), program, and advertisement (or advertisement break).

(EX1007 at [0064] (disclosing detection of the start/end of commercial breaks using audio or video data).)

Advertisements only make up a portion of an incoming video stream so that continually calculating features for the incoming video stream 1020 and comparing the features to known advertisement fingerprints 1030 may not be efficient. **According to one embodiment, the feature based techniques described above (e.g., volume increases, increase scene changes, monochrome images) may be used to detect the start of a potential advertisement (or advertisement block) and the calculating of features 1020 and comparing to known fingerprints 1030 may only be performed once a possible**

advertisement break has been detected. It should be noted that some methods of detecting the possibility of an advertisement break in the video stream such as an increase in scene changes, where scene changes may be detected by comparing successive CCVs, may in fact be calculating features of the video stream 1020 so the advertisement detection process may begin with the comparison 1030.

(EX1007 at [0086] (disclosing detection of the start of potential advertisements using audio data like volume increases or video data like scene changes).)

109. Eldering further discloses performing post-content identification steps such as substituting one advertisement for another, bookmarking desired content, or starting/stopping recording of content based on the content identification. (*Id.*, [0008], [0010], [0043]-[0044], [0082], [0158]-[0163], [0167], claims 3, 5, FIGS. 1, 2, 10 *see also* my analysis and discussion above in Section VII.B (Eldering summary).)

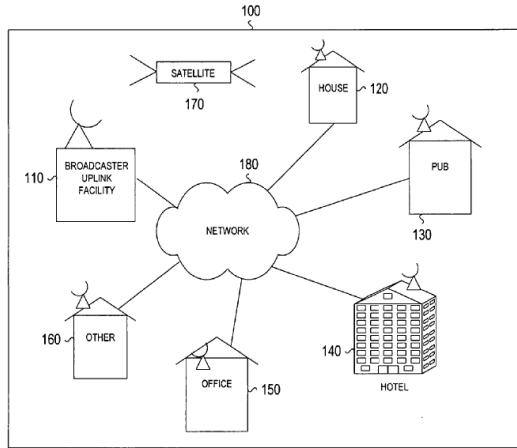


FIG. 1

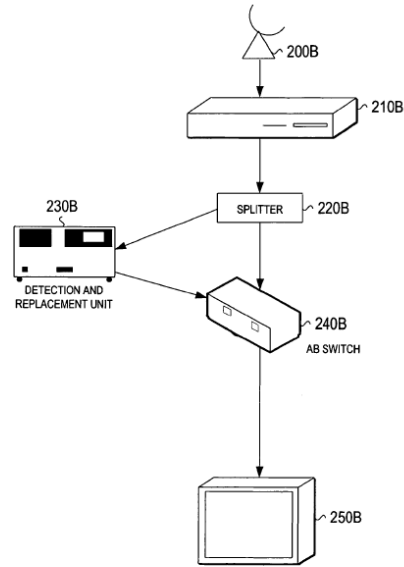


FIG. 2

(EX1007 at FIGS. 1-2; *see also id.*, [0039]-[0043], [0046]-[0048].)

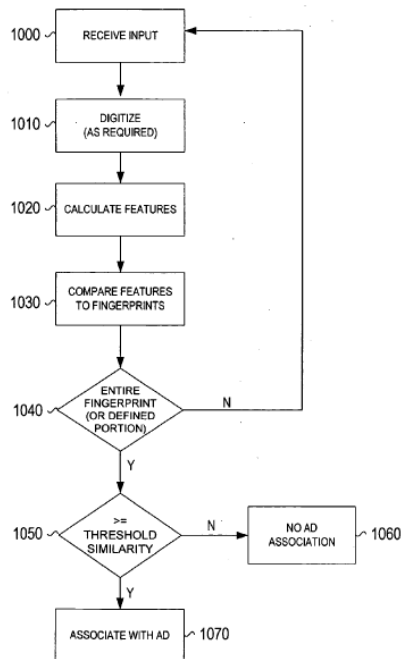


FIG. 10

(EX1007 at FIG. 10; *see also id.*, [0080]-[0082].)

110. Thus, Eldering discloses content identification systems and methods in the same technical field as Steuer750 and Steuer389, including content identification based on data divisible into samples. Accordingly, a person of ordinary skill in the art would have had reason to consider the teachings of Eldering when implementing Steuer750 or Steuer389.

111. Similar to limitation 1.b, Eldering discloses receiving, at a network interface device 200 (like the computer of Steuer750/389), an incoming video stream (“video data”). (EX1007 at [0046]-[0047], FIG. 2; *see also id.*, [0039], [0080], FIGS. 1, 10.) As discussed above, the received video stream is divisible into samples. (*Id.*, [0008], [0070]-[0076], [0101], FIGS. 7-9.)

FIG. 2 illustrates an exemplary configuration for manual local **detection of advertisements within a video programming stream. The incoming video stream is received by a network interface device (NID) 200. The type of network interface device will be dependent on how the incoming video stream is being delivered to the location. For example, if the content is being delivered via satellite (e.g., 170 of FIG. 1) the NID 200 will be a satellite dish (illustrated as such) for receiving the incoming video stream. The incoming video stream is provided to a STB 210 (a tuner) that tunes to a desired**

channel, and possibly decodes the channel if encrypted or compressed. It should be noted that the STB 210 may also be capable of recording programming as is the case with a DVR or video cassette recorder VCR.

The STB 210 forwards the desired channel (video stream) to a splitter 220 that provides the video stream to a detection/replacement device 230 and a selector (e.g., A/B switch) 240. The detection/replacement device 230 detects and replaces advertisements by creating a presentation stream consisting of programming with targeted advertisements. The selector 240 can select which signal (video stream or presentation stream) to output to an output device 250 (e.g., television). The selector 240 may be controlled manually by an operator, may be controlled by a signal/message (e.g., ad break beginning message, ad break ending message) that was generated and transmitted from an upstream detection location, and/or may be controlled by the detection/replacement device 230. The splitter 220 and the selector 240 may be used as a bypass circuit in case of an operations issue or problem in the detection/replacement device 230. The default mode for the selector 240 may be to pass-through the incoming video stream.

(EX1007 at [0046]-[0047].)

Alternatively, the detection may be based on calculating features (statistical parameters) about the incoming video stream. The

features calculated may include, for example, color histograms or CCVs as discussed above. **The features may be calculated for an entire video frame, as discussed above, number of frames, or may be calculated for evenly/randomly highly subsampled representations of the video frame. For example, the video frame could be sampled at a number (e.g., 64) of random locations or regions in the video frame and parameters such as average color) may be computed for each of these regions. The subsampling can also be performed in the temporal domain.** The collection of features including CCVs for a plurality of images/frames, color histograms for a plurality of regions, may be referred to as a fingerprint.

FIG. 7 illustrates an exemplary pixel grid 700 for a video frame. For ease of understanding, we limit the pixel grid to 12×12 (144 pixels), limit the color variations for each color (RGB) to the two most significant bits (4 color variations), and have each pixel have the same number associated with each of the colors (RGB) so that a single number represents all colors. **A plurality of regions 710, 720, 730, 740, 750, 760, 770, 780, 785, 790, 795 of the pixel grid 700 are sampled and an average color for each of the regions 710, 720, 730, 740, 750, 760, 770, 780, 785, 790, 795 is calculated.** For example, the region 710 has an average color of 1.5, the region 790 has an average color of 0.5 and the region 795 has an average color of 2.5.

One advantage of the sampling of regions of a frame instead of an entire frame is that the entire frame would not need to be copied in order to calculate the features (if copying was even needed to calculate the features). Rather, certain regions of the image may be copied in order to calculate the features for those regions. As the regions of the frame would provide only a partial image and could not be used to recreate the image, there would be less potential copyright issues. As will be discussed in more detail later, the generation of fingerprints for known entities (e.g., advertisements, intros) that are stored in a database for comparison could be done for regions as well and therefore create less potential copyright issues.

FIG. 8 illustrates two exemplary pixel grids 800 and 810. Each of the pixel grids is 11×11 (121 pixels) and is limited to a single bit (0 or 1) for each of the colors. The top view of each pixel grid 800, 810 has a plurality of regions identified 815-850 and 855-890 respectively. The lower view of each pixel grids 800, 810 has the coherent and incoherent pixels identified, where the threshold level is greater than 5.

FIG. 9 illustrates exemplary comparisons of the pixel grids 800, 810 of FIG. 8. Color histograms 900, 910 are for the entire frame 800, 810 respectively and the difference in the color histograms 920 is 0. CCVs 930, 940 are for the entire frame 800, 810 respectively and the difference in the CCVs 950 is 0. Average colors 960, 970 capture the average colors for the various

identified regions in frames 800, 810. The difference is the average color of the regions 980 is 3.5 (using the sum of absolute values).

FIGS. 7-9 focused on determining the average color for each of the regions but the techniques illustrated therein are not limited to average color determinations. For example, a color histogram or CCV could be generated for each of these regions. For CCVs to provide useful benefits the regions would have to be big enough or all of the colors will be incoherent. All of the colors will be coherent if the coherent threshold is made too low.

The calculated features/fingerprints (e.g., CCVs, evenly/randomly highly subsampled representations) are compared to corresponding features/fingerprints for known intros and/or outros. The fingerprints for the known intros and outros could be calculated and stored in advance. The comparison of calculated features of the incoming video stream (statistical parameterized representations) to the stored fingerprints for known intros/outros will be discussed in more detail later.

(EX1007 at [0070]-[0076].)

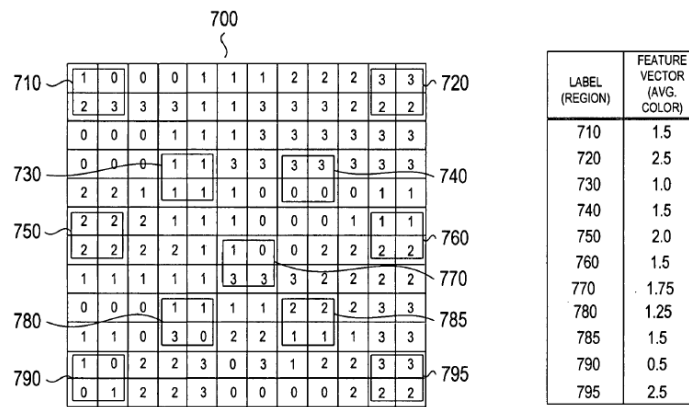


FIG. 7

(EX1007 at FIG. 7.)

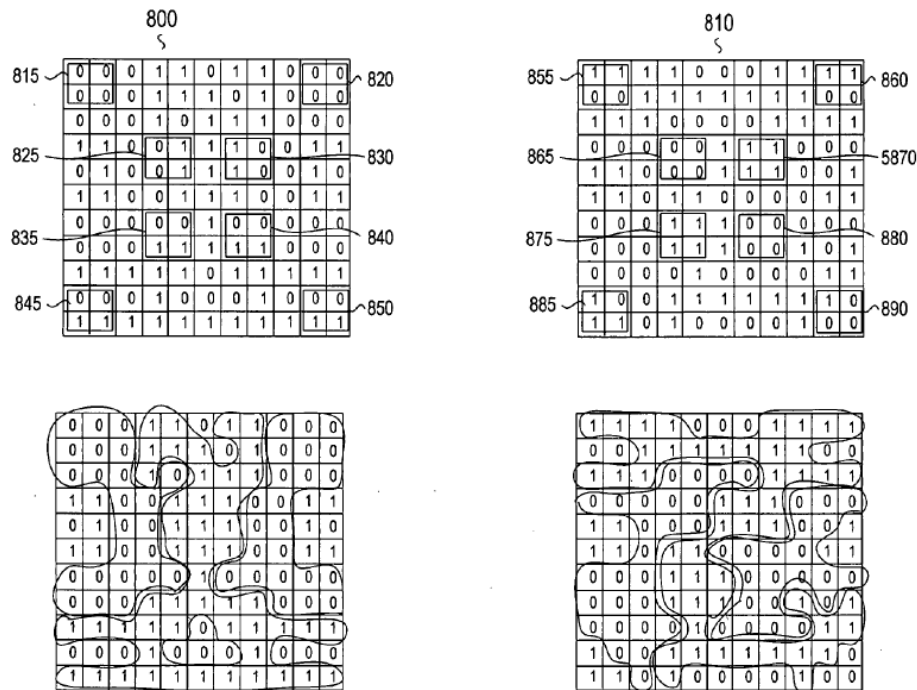


FIG. 8

(EX1007 at FIG. 8.)

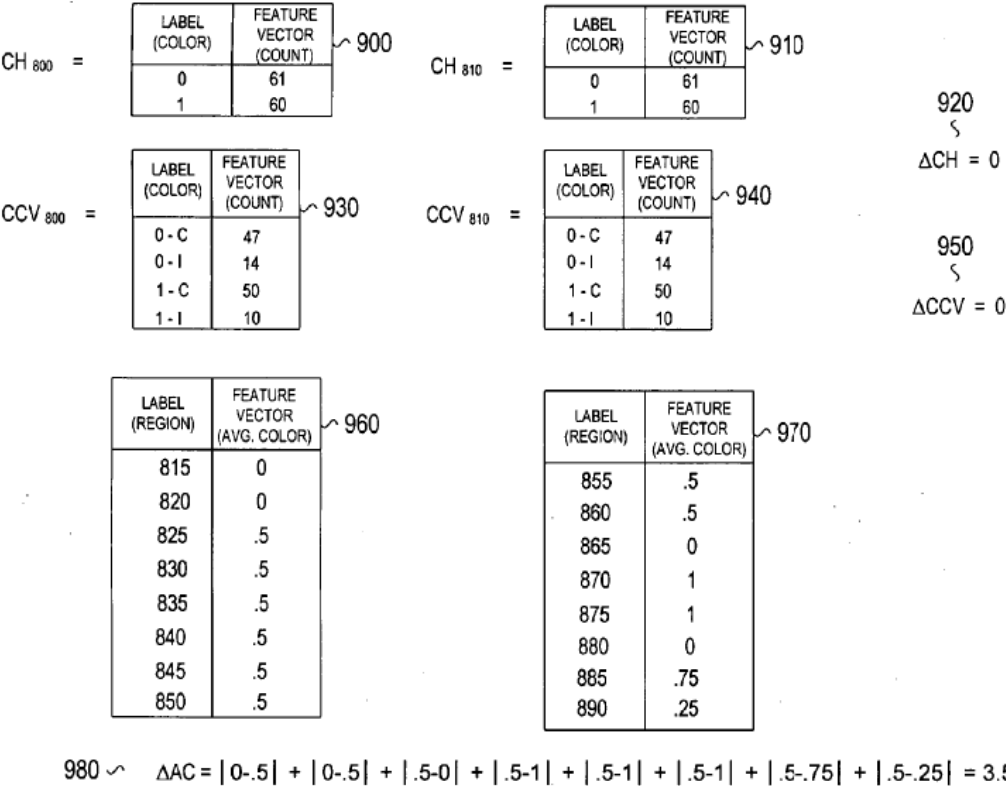


FIG. 9

(EX1007 at FIG. 9.)

112. Eldering further discloses additional process steps related to content identification, similar to those disclosed in Steuer750 and Steuer389, such as adjusting the parameters for content identification based on results associated with identifying prior samples. (EX1007 at [0090]-[0093], [0097], FIGS. 11-14; EX1005 at [0032], [0034], [0037]-[0038]; EX1006 at 6:28-46, 6:58-7:3, 7:39-65.)

113. In view of Eldering’s teachings, a person of ordinary skill in the art would have been motivated to consider and draw from such teachings when

implementing Steuer750's or Steuer389's disclosed methods and systems such that, in addition to, or as an alternative to, using audio data samples, video data samples would have been used to generate content identification results and perform subsequent data processing steps. As explained above, Steuer750, Steuer389, and Eldering all concern content identification processes that require collecting data samples, making content identifications based on those data samples, and comparing sequences of content identifications against known sequences.

114. A person of ordinary skill in the art would have been motivated and would have known how to modify Steuer750 and Steuer389 to incorporate Eldering's teachings as proposed, because as discussed above, Eldering discloses a way of identifying content based on matching video data samples with video data stored in a database. Thus, a person of ordinary skill in the art would have recognized that the application of using video data instead of the audio data-based approach disclosed in Steuer750 or Steuer389 would have been based on obvious modifications using known technology, such as sampling video data instead of (or in addition to) sampling audio data, and using the sampled video data to identify content as taught by Eldering.

115. In view of the foregoing, a person of ordinary skill in the art would have had the capability and knowledge to implement, and would have been motivated to

try, the above modification to include the use of video data instead of (or in addition to) audio data as required by the claim, while ensuring proper operation of the Steuer750/Steuer389 systems and methods. Therefore, a person of ordinary skill in the art would have had a reasonable expectation of success that the combined Steuer750-Eldering method and combined Steuer389-Eldering method would have operated as expected.

116. The above modification would have been a mere combination of known technologies (e.g., the concept of gathering video data samples with Steuer750's/Steuer389's disclosed systems and methods for making content identifications), according to known methods (e.g., sampling the video stream), to achieve predictable results (e.g., generating content identification results by comparing video data samples with known content).

117. Therefore, the Steuer750-Eldering system and the Steuer389-Eldering system discloses or suggests claim limitation 1.b.

- c) determining, by the one or more computers, a next sample of the video data for attempted identification of a media content item corresponding to the next sample by adjusting a length of a time window of the next sample based at least in part upon a result of an attempt by the one or more computers to identify a first prior sample as corresponding to a media content item stored in a computerized database of reference content;**

118. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination discloses or suggests this limitation.

119. First, Steuer389's claim 1 recites almost identical language as claim limitation 1.c, albeit in the context of "audio data" as noted with respect to limitation 1.b: "determining, using the one or more computers, a next sample of the audio data for attempted identification of a media content item corresponding to the next sample by adjusting a length of a time window of the next sample based at least in part upon a result of an attempt using the one or more computers to identify a first prior sample as corresponding to a media content item stored in a computerized database of reference content." (EX1006 at claim 1.)

120. Further, with reference to Figure 2, Steuer750 and Steuer389 disclose that "[a]t step 206, an identification result is used to adjust sampling control and selection parameters for one or more subsequent samples. Parameters that might be adjusted include resolution, backtrack selection, **sample time window**, or recognition algorithm selected." (EX1005 at [0037], [0063]-[0065] (similarly describing the adjustment of time window length of a next sample based on identification results of prior samples), FIGS. 2, 6; *see also id.*, [0034] (disclosing that the "time window" of a sample corresponds to the sample's "length"), claims 19-21; EX1006 at 7:39-42, 13:20-56, FIGS. 2, 6; *see also id.*, 7:2-3.) Steuer750 and

Steuer389 describe with respect to step 206 that “once media content for a limited series of samples has been identified, the **time window size for each sample** might be decreased from, for example, 5 seconds to 3 seconds” which “would mean that samples submitted for attempted identification would be shorter in length.” (EX1005 at [0038], [0063]-[0065]; EX1006 at 7:57-65, 13:20-56.)

FIG. 2 illustrates a media measurement method 200 in accordance with aspects of an embodiment of the present invention. One or more elements of method 200 may be carried out, for example, by system 100 of FIG. 1 or other similar systems. At step 201, method 200 receives audio data from a monitoring device. Step 201 caches a series of samples and determines the sample resolution with which samples are submitted for content identification. In particular, as described further below (see FIG. 4 and accompanying text), sample resolution refers to how many samples in a given sequence of samples are processed for content identification. **Step 201 also determines length, or “time window” of each sample.**

(EX1005 at [0034]; EX1006 at 7:2-3.)

At step 206, an identification result is used to adjust sampling control and selection parameters for one or more subsequent samples. Parameters that might be adjusted include resolution, backtrack selection, sample time window, or recognition algorithm selected. For example, if a sample has

been identified successfully, and the identification indicates a match with the prior sample, the sampling resolution might be decreased from $\frac{1}{5}$ to $\frac{1}{10}$, meaning that, in a sequence of samples, nine samples rather than four are skipped before selecting the next sample to analyze for identification. As another example, if an identification result does not match the result of the prior sample identified, then step 206 might initiate a back track analysis and select an intervening sample that has not yet been analyzed. Step 201 would then pass the sample selected from backtracking to step 202 for attempted identification. As another example, resolution might be increased when recent identification results suggest an extended period of silence or absence of known content.

As yet another example, once media content for a limited series of samples has been identified, the time window size for each sample might be decreased from, for example, 5 seconds to 3 seconds. This would mean that samples submitted for attempted identification would be shorter in length. This can create efficiencies because once the content is identified, the process needs only to detect a change in content, and this can, in some cases, be accomplished utilizing a smaller amount of audio data, thereby preserving system resources.

(EX1005 at [0037]-[0038]; EX1006 at 7:39-65.)

FIG. 6 illustrates timeline structure for data samples generated by system and methods such as module 121 of FIG. 1 **and steps 201 and 206 of FIG. 2.**

As illustrated, sample set 600 reflects a division of audio data into sixteen samples, three of which are selected for analysis. The samples selected for recognition analysis are referenced generally as the “ n_0^{th} ,” n_1^{th} ,” and “ n_2^{th} ” samples. Samples between the selected samples are not separately numbered, but are marked off between vertical lines along the horizontal length of timelines structure 600 as illustrated. **Each sample has a sample time window length “i.”** Furthermore, the sample set has varying “resolutions” including $r_0 = \frac{1}{5}$ between the n_0^{th} and n_1^{th} samples and $r_1 = \frac{1}{10}$ between the n_1^{th} and the n_2^{th} samples. In other words, “resolution” here refers to the portion of samples being selected during a period of time. In the present example, a process portion such as steps 201 and 206 of process 200 of FIG. 2 has, **based on the identification results of the n_0^{th} and n_1^{th} samples,** adjusted the resolution from $\frac{1}{5}$ to $\frac{1}{10}$. This could occur, for example, as previously described, because the n_0^{th} and n_1^{th} samples match the same content ID and process 200 determines that the likelihood that the next selected sample using a resolution of $\frac{1}{5}$ would also match justifies decreasing the resolution to $\frac{1}{10}$ so that the next sample selected for analysis is the n_2^{th} . **However, if the n_2^{th} sample does not match the n_1^{th} sample,** then process 200 can make a backtrack decision and go

back in the series of samples to select, for example, the $(n_1+6)^{\text{th}}$ sample for content recognition analysis in order to pinpoint more closely where in the play stream the content change occurred. The order in which samples are selected for analysis in this example is illustrated by arrows 1, 2, and 3 in FIG. 6.

Although not illustrated in FIG. 6, as previously described, steps 201 and 206 of method 200 can also determine based on identification results the time window length can be adjusted. In the context of FIG. 6, this would correspond to changing the value of “i,” the time window length of a particular sample.

(EX1005 at [0063]-[0065]; EX1006 at 13:20-56.)

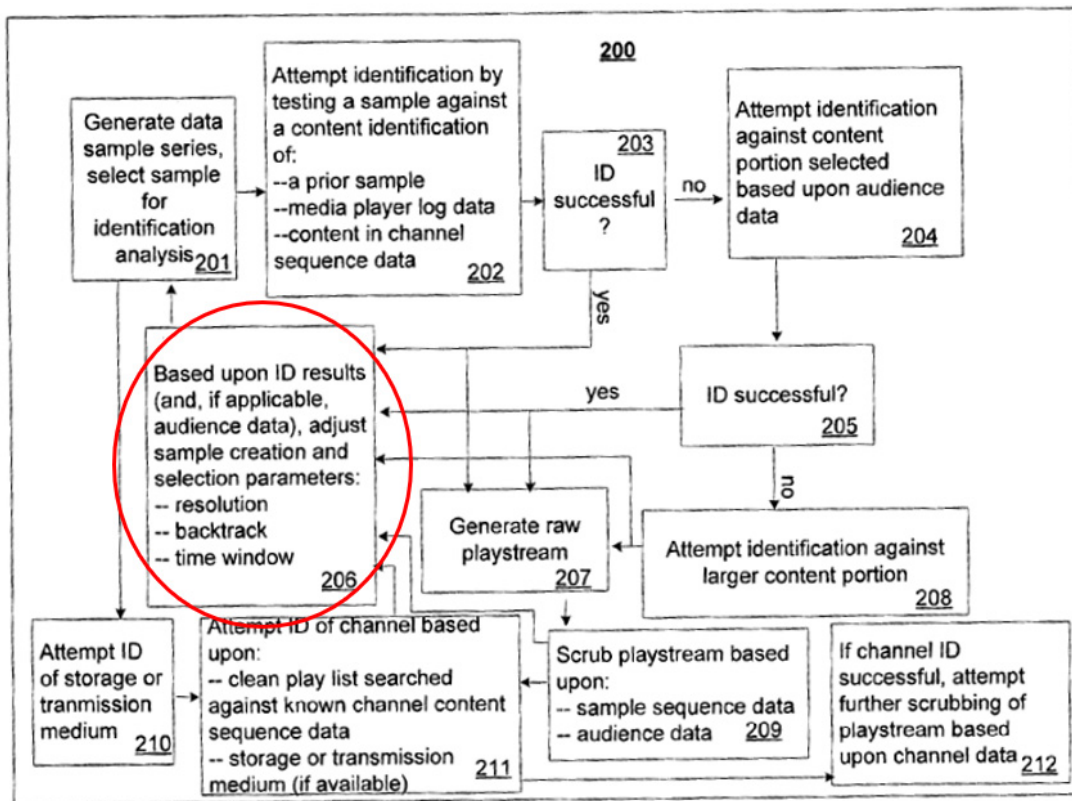
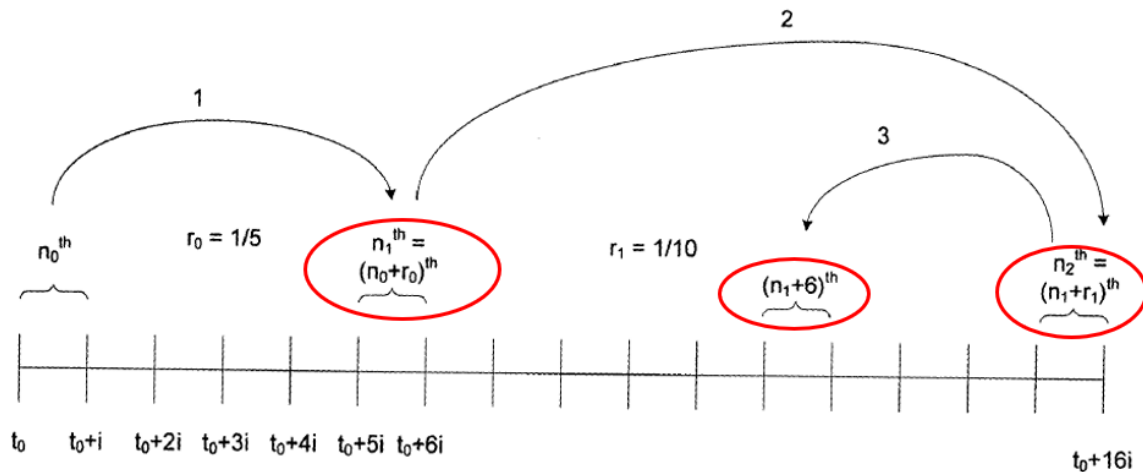


FIG. 2

(EX1005 at FIG. 2 (annotated); EX1006 at FIG. 2.)



600

FIG. 6

(EX1006 at FIG. 6 (annotated); EX1006 at FIG. 6.)

121. The above processes are performed using one or more computers (e.g., the media measurement system 100 implemented by computer system 700), as discussed above. (See Section IX.A.1.b, above.)

122. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitation 1.b, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content

identification systems using video data instead of the disclosed audio-based approach. (See Section IX.A.1.b, above.) Moreover, as I discussed above, Eldering discloses adjusting the parameters for content identification based on results associated with identifying prior samples, similar to Steuer750's and Steuer389's disclosures above. (EX1007 at [0090]-[0093], [0097], FIGS. 11-14.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination discloses or suggests claim limitation 1.c.

- d) wherein the time length window of the next sample is changed relative to a time length window of the first prior sample based on the attempt to identify the first prior sample indicating a result that is different than a result of an attempt to identify a second prior sample;**

123. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination discloses or suggests this limitation for reasons similar to those given above for limitation 1.c.

124. First, Steuer389's claim 1 recites almost identical language as claim limitation 1.d, albeit in the context of "audio data" as noted with respect to limitation 1.b: "wherein, if the attempt to identify the first prior sample indicates a result that is different than a result of an attempt to identify a second prior sample, then the time length window of the next sample is changed relative to a time window of the first prior sample." (EX1006 at claim 1.)

125. Further, both Steuer750 and Steuer389 disclose adjusting selection parameters like the time length window for subsequent samples (“next sample”) based on prior identification results. (*See* Section IX.A.1.c, above; EX1005 at [0037]-[0038], [0063]-[0065], FIGS. 2, 6; EX1006 at 7:39-42, 7:57-61, 13:20-56, FIGS. 2, 6.) The adjustment of selection parameters includes adjustment when an identification result of a current sample does not match the identification result of a prior sample. (EX1005 at [0037] (“As another example, if an identification result does not match the result of the prior sample identified, then step 206 might initiate a back track analysis and select an intervening sample that has not yet been analyzed.”), [0038] (describing adjusting the time window of the next sample based on prior identification results) [0064]-[0065] (similarly describing scenarios where the time window length for a next sample being adjusted based on identification results for prior samples not matching each other); EX1006 at 7:38-65, 13:23-56.)

126. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750’s and Steuer389’s disclosed content identification systems using video data instead of the disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering

combination and the Steuer389-Eldering combination discloses or suggests claim limitation 1.d.

- e) **submitting, by the one or more computers, the determined next sample for attempted content identification; and**

127. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for reasons similar to those given above for limitation 1.c.

128. For example, Steuer750 and Steuer389 both disclose that the next sample determined in step 206 (“determined next sample”) is provided to step 202 (via step 201) for attempting content identification based on that next sample, as claimed. (See Section IX.A.1.c, above; EX1005 at [0037]-[0038] (“As another example, if an identification result does not match the result of the prior sample identified, then step 206 might initiate a back track analysis and select an intervening sample that has not yet been analyzed. **Step 201 would then pass the sample selected from backtracking to step 202 for attempted identification . . .** As yet another example, once media content for a limited series of samples has been identified, the time window size for each sample might be decreased from, for example, 5 seconds to 3 seconds. **This would mean that samples submitted for attempted identification** would be shorter in length.”), [0063]-[0065] FIGS. 2, 6,

claims 19-21; EX1006 at 7:39-65, 13:20-56, FIGS. 2, 6.) The above processes are performed using one or more computers (e.g., the media measurement system 100 implemented by computer system 700), as discussed above. (See Section IX.A.1.b, above.)

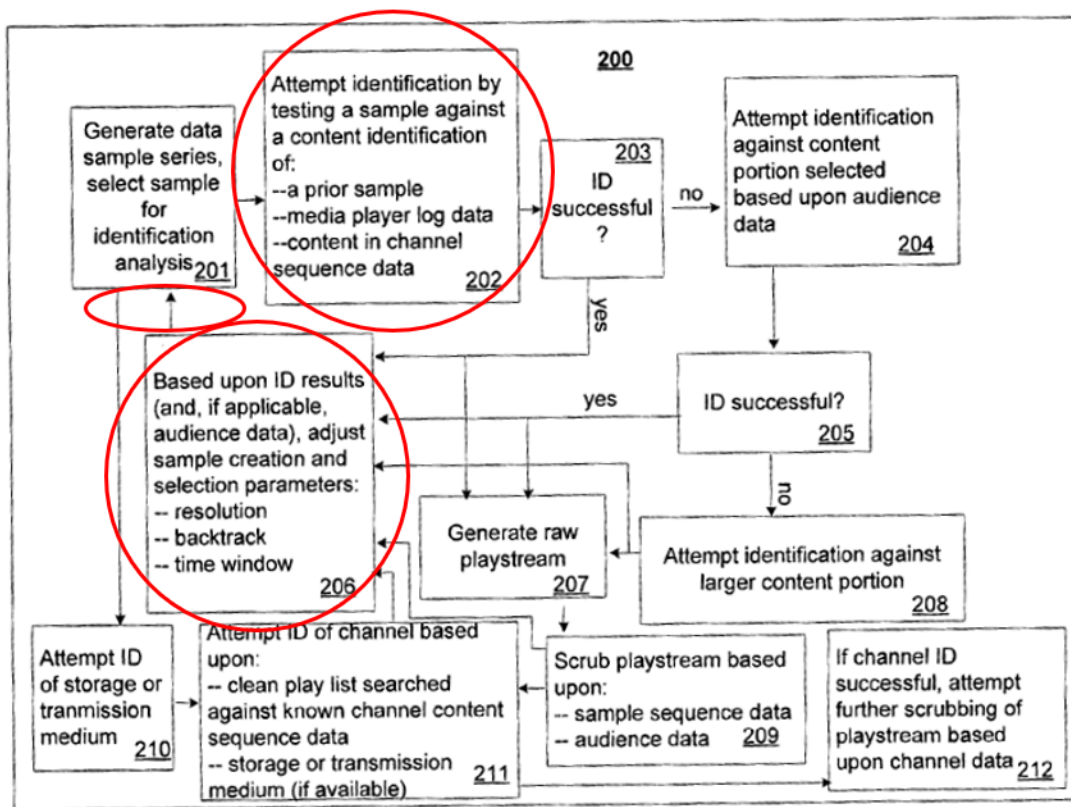


FIG. 2

(EX1005 at FIG. 2 (annotated); EX1006 at FIG. 2.)

129. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would

have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data instead of the disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination discloses or suggests claim limitation 1.e.

- f) **generating, by the one or more computers, a play stream of content identifications based on results of attempts to identify a series of samples comprising the determined next sample.**

130. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

131. For example, Steuer750 and Steuer389 both disclose that the content identification process discussed above proceeds to generating a raw play stream 410 ("play stream") in step 207 (and shown in Figure 4) of content identifications ("content ID" column in raw play stream 410) based on results of attempts to identify a series of samples comprising the determined sample as discussed above with respect to at least steps 201, 202, and 206 of Figure 2. (EX1005 at [0040], [0045], FIGS. 2, 4; *see also id.*, [0036], [0041]-[0042], FIG. 3; EX1006 at 8:1-22, 9:12-31, FIGS. 2, 4; *see also id.*, 7:16-38, 8:23-49, FIG. 3.) These processes are performed using one or more computers (e.g., the media measurement system 100 implemented by computer system 700), as discussed above. (*See* Section IX.A.1.b, above.)

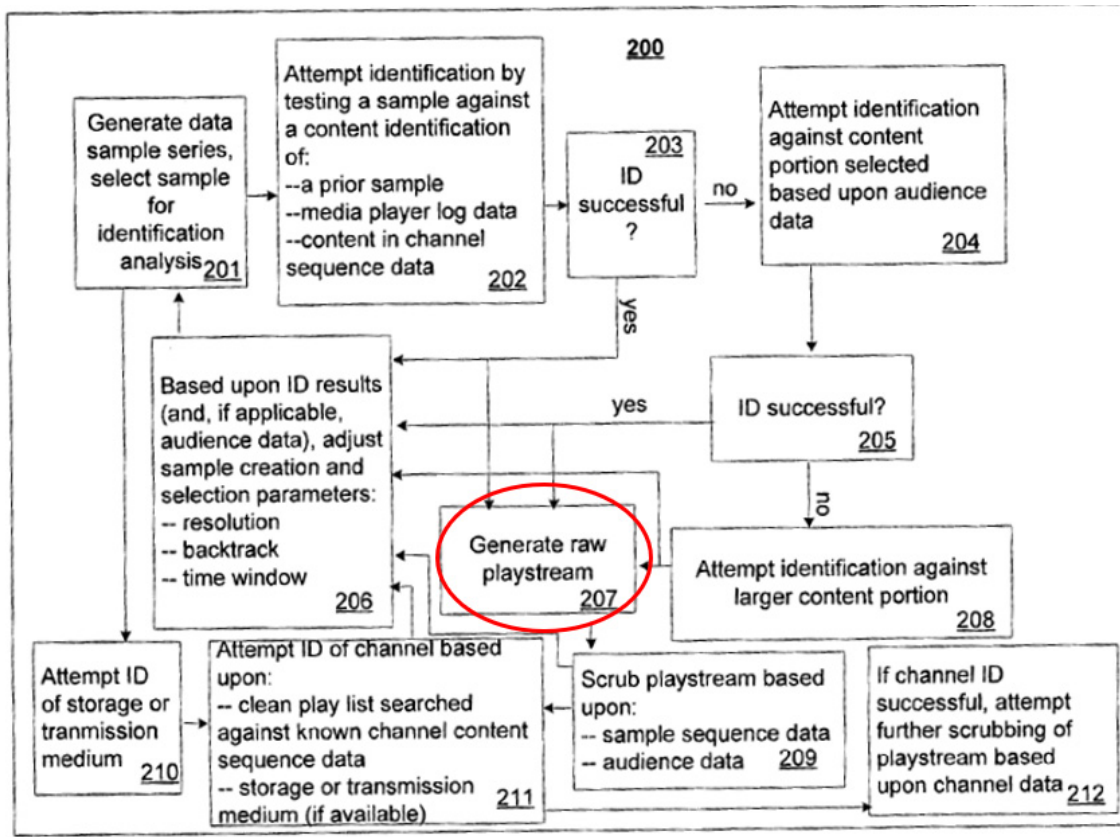


FIG. 2

(EX1005 at FIG. 2 (annotated); EX1006 at FIG. 2.)

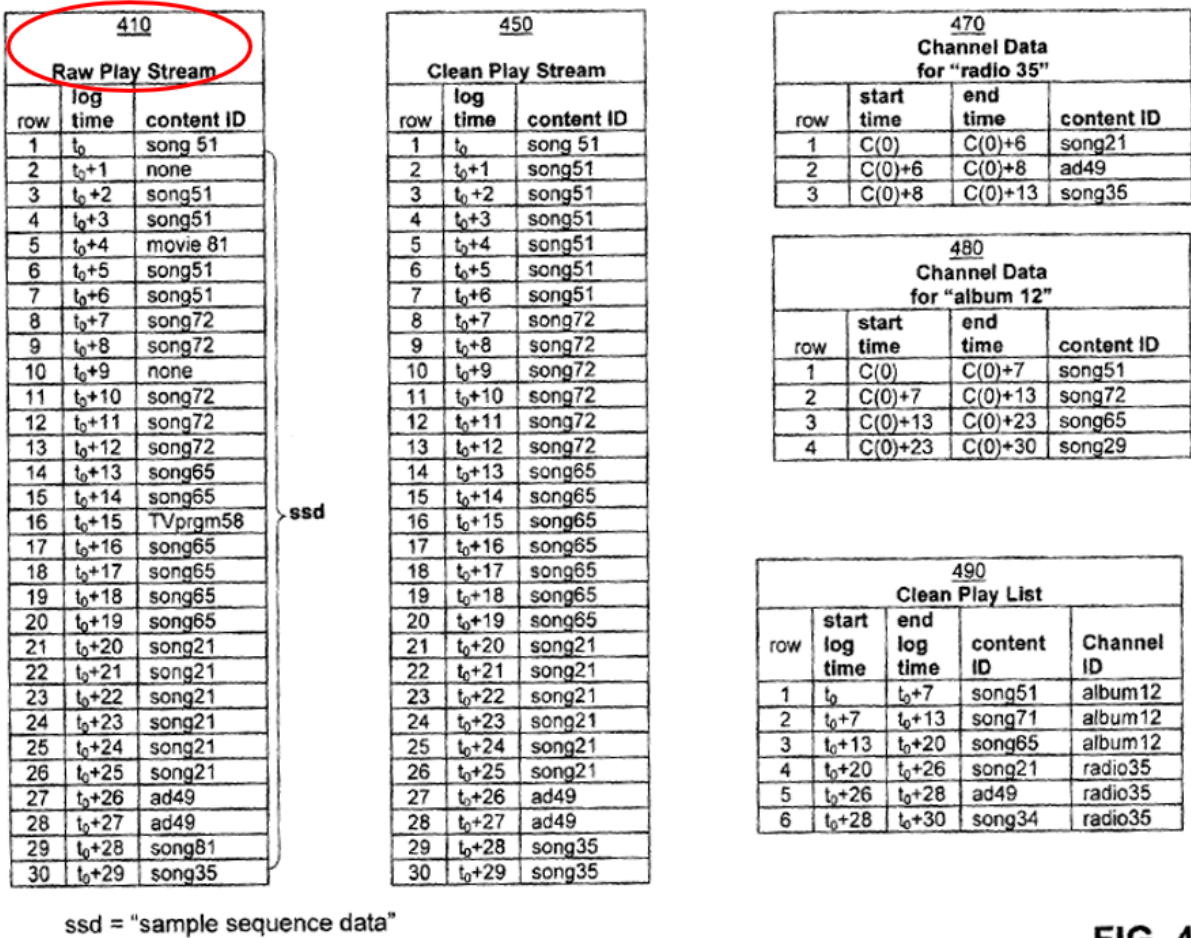


FIG. 4

(EX1005 at FIG. 4 (annotated); EX1006 at FIG. 4.)

132. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of the disclosed audio-based

approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination discloses or suggests claim limitation 1.e.

2. Claim 2

a) The method of claim 1 wherein determining comprises selecting a format of a sample.

133. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

134. For example, Steuer389's claim 2 recites almost identical language as claim limitation 2.a, albeit in the context of "audio data" as noted with respect to limitation 1.b: "wherein determining comprises selecting a format of a sample." (EX1006 at claim 2.)

135. Steuer750 similarly discloses that the step of determining a next sample (limitation 1.c) comprises selecting a format of a sample, in the context of its audio-based approach. (EX1005 at claims 19-20 (disclosing "determining [a sample for attempted identification based at least in part upon a result of an attempt to identify a prior sample against known content] comprises selecting a format of a sample"), claim 21 (disclosing that the "format includes length of a sample time window").)

136. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above

with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest claim 2.

3. Claim 3

- a) The method of claim 2 wherein: selecting a format includes selecting video parameters corresponding to requirements of a content recognition algorithm; and**

137. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

138. For example, Steuer389's claim 3 recites almost identical language as claim limitation 3.a, albeit in the context of "audio parameters" as opposed to "video parameters": "selecting a format includes selecting audio parameters corresponding to requirements of a content recognition algorithm." (EX1006 at claim 3.)

139. Steuer750 similarly discloses that the step of selecting a format (limitation 2.a) includes selecting parameters corresponding to requirements of a content recognition algorithm, in the context of its audio-based approach. (EX1005 at claims 19-20 (disclosing "determining [a sample for attempted identification based at least in part upon a result of an attempt to identify a prior sample against

known content] comprises selecting a format of a sample”), claim 22 (disclosing that the “format includes audio parameters corresponding to requirements of a content recognition algorithm”).)

140. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750’s and Steuer389’s disclosed content identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 3.a.

- b) determining comprises, for a sample subsequent to an n’t h sample, selecting video parameters corresponding to one of a plurality of available content recognition algorithms based at least in part upon an identification result corresponding to the n’t h sample.**

141. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

142. For example, Steuer389’s claim 3 recites almost identical language as claim limitation 3.b, albeit in the context of “audio parameters” as opposed to “video parameters”: “determining comprises, for a sample subsequent to an n’t h sample,

selecting audio parameters corresponding to one of a plurality of available content recognition algorithms based at least in part upon an identification result corresponding to the n'th sample.” (EX1006 at claim 3.)

143. Steuer750 similarly discloses that the step of determining a next sample (limitation 1.c) comprises, for a sample subsequent to an n'th sample, selecting parameters corresponding to one of a plurality of available content recognition algorithms based at least in part upon an identification result corresponding to the n'th sample, in the context of its audio-based approach. (EX1005 at claims 19-20 (disclosing “determining [a sample for attempted identification based at least in part upon a result of an attempt to identify a prior sample against known content] comprises selecting a format of a sample”), claim 22 (disclosing that the “determining comprises, for a sample subsequent to an n'th sample, selecting audio parameters corresponding to one of a plurality of available content recognition algorithms based at least in part upon an identification result corresponding to the n'th sample”).)

144. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content

identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 3.b.

4. Claim 4

- a) The method of claim 2 wherein: selecting a format includes selecting video parameters corresponding to settings of a content recognition algorithm; and**

145. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

146. For example, Steuer389's claim 4 recites almost identical language as claim limitation 4.a, albeit in the context of "audio parameters" as opposed to "video parameters": "selecting a format includes selecting audio parameters corresponding to settings of a content recognition algorithm." (EX1006 at claim 4.)

147. Steuer750 similarly discloses that the step of selecting a format (limitation 2.a) includes selecting parameters corresponding to settings of a content recognition algorithm, in the context of its audio-based approach. (EX1005 at claims 19-20 (disclosing "determining [a sample for attempted identification based at least in part upon a result of an attempt to identify a prior sample against known content] comprises selecting a format of a sample"), claim 23 (disclosing that the "format

includes audio parameters corresponding to settings of a content recognition algorithm”)..)

148. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750’s and Steuer389’s disclosed content identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 4.a.

- b) determining comprises, for a sample subsequent to an n’t h sample, selecting a setting from a plurality of settings of a content recognition algorithm based at least in part upon an identification result corresponding to the n’t h sample.**

149. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

150. For example, Steuer389’s claim 4 recites almost identical language as claim limitation 4.b, albeit in the context of “audio parameters” as opposed to “video parameters”: “determining comprises, for a sample subsequent to an n’t h sample, selecting a setting from a plurality of settings of a content recognition algorithm

based at least in part upon an identification result corresponding to the n'th sample.”
(EX1006 at claim 4.)

151. Steuer750 similarly discloses that the step of determining a next sample (limitation 1.c) comprises, for a sample subsequent to an n'th sample, selecting a setting from a plurality of settings of a content recognition algorithm based at least in part upon an identification result corresponding to the n'th sample. (EX1005 at claims 19-20 (disclosing “determining [a sample for attempted identification based at least in part upon a result of an attempt to identify a prior sample against known content] comprises selecting a format of a sample”), claim 23 (disclosing that the “determining comprises, for a sample subsequent to an n'th sample, selecting a setting from a plurality of settings of a content recognition algorithm based at least in part upon an identification result corresponding to the n'th sample”).)

152. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering

combination and the Steuer389-Eldering combination disclose or suggest limitation 4.b.

5. Claim 5

- a) The method of claim 1 wherein channel data is used in determining the next sample, the channel data relating to recently identified samples.**

153. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

154. For example, Steuer389's claim 5 recites almost identical language as claim limitation 5.a, albeit in the context of "audio data" as noted with respect to limitation 1.b: "wherein channel data is used in determining the next sample, the channel data relating to recently identified samples." (EX1006 at claim 5.)

155. Steuer750 similarly discloses in the context of "audio data" that "channel data is used in determining the sample, the channel data relating to recently identified samples." (EX1005 at claim 29; *see also id.*, claim 19 ("determining a sample for attempted identification based at least in part upon a result of an attempt to identify a prior sample against known content").)

156. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would

have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 5.a.

6. Claim 6

- a) The method of claim 1 wherein content data is used in determining the next sample, the content data relating one or more recently identified samples.**

157. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

158. For example, Steuer389's claim 6 recites almost identical language as claim limitation 6.a, albeit in the context of "audio data" as noted with respect to limitation 1.b: "wherein content data is used in determining the next sample, the content data relating one or more recently identified samples." (EX1006 at claim 6.)

159. Steuer750 similarly discloses in the context of "audio data" that "content data is used in determining the sample, the content data relating one or more recently identified samples." (EX1005 at claim 30; *see also id.*, claim 19

(“determining a sample for attempted identification based at least in part upon a result of an attempt to identify a prior sample against known content”).)

160. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750’s and Steuer389’s disclosed content identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 6.a.

7. Claim 7

a) The method of claim 6 wherein content data includes at least one of content type and content length.

161. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

162. For example, Steuer389’s claim 7 recites almost identical language as claim limitation 7.a, albeit in the context of “audio data” as noted with respect to limitation 1.b: “wherein content data includes at least one of content type and content length.” (EX1006 at claim 7.)

163. Steuer750 similarly discloses in the context of “audio data” that “content data includes at least one of content type and content length.” (EX1005 at claim 31; *see also id.*, claims 19, 30.)

164. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750’s and Steuer389’s disclosed content identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 7.a.

8. Claim 8

a) The method of claim 1 wherein sample sequence data is used in determining the next sample, the sample sequence data relating to recently identified samples.

165. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

166. For example, Steuer389’s claim 8 recites almost identical language as claim limitation 8.a, albeit in the context of “audio data” as noted with respect to limitation 1.b: “wherein sample sequence data is used in determining the next

sample, the sample sequence data relating to recently identified samples.” (EX1006 at claim 8.)

167. Steuer750 similarly discloses in the context of “audio data” that “sample sequence data is used in determining the sample, the sample sequence data relating to recently identified samples.” (EX1005 at claim 32; *see also id.*, claim 19.)

168. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750’s and Steuer389’s disclosed content identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 8.a.

9. Claim 9

- a) The method of claim 1 wherein log data is used in determining the next sample, the log data relating to a media player of an audience member using the monitoring device.**

169. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

170. For example, Steuer389's claim 9 recites almost identical language as claim limitation 9.a, albeit in the context of "audio data" as noted with respect to limitation 1.b: "wherein log data is used in determining the next sample, the log data relating to a media player of an audience member using the monitoring device." (EX1006 at claim 9.)

171. Steuer750 similarly discloses in the context of "audio data" that "log data is used in determining the sample, the log data relating to a media player of an audience member using the monitoring device." (EX1005 at claim 33; *see also id.*, claim 19.)

172. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of its disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 9.a.

10. Claim 10

- a) The method of claim 1 wherein location data is used in determining the next sample, the location data relating to a media player of an audience member using the monitoring device.**

173. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

174. For example, Steuer389's claim 10 recites almost identical language as claim limitation 10.a, albeit in the context of "audio data" as noted with respect to limitation 1.b: "wherein location data is used in determining the next sample, the location data relating to a media player of an audience member using the monitoring device." (EX1006 at claim 10.)

175. Steuer750 similarly discloses in the context of "audio data" that "location data is used in determining the sample, the location data relating to a media player of an audience member using the monitoring device." (EX1005 at claim 34; *see also id.*, claim 19.)

176. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of its disclosed audio-based

approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 10.a.

11. Claim 11

a) A media measurement method comprising:

177. I have been asked to assume the preamble of claim 9 is limiting. Under that assumption, Steuer750 and Steuer389 both independently disclose the limitation therein for the same reasons as I discussed above for the preamble of claim 1. (*See* Section IX.A.1.a, above.)

178. Moreover, Steuer389's claim 11 includes an identical preamble: A media measurement method comprising" (EX1006 at claim 11).

179. Thus, Steuer750 and Steuer389 both disclose the claim 11 preamble. (*See also* Sections IX.A.11.b–e, below.)

b) receiving, at one or more computers, video data from a monitoring device, the video data being divisible into samples; and

180. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for the same reasons as I discussed above for limitation 1.b. (*See* Section IX.A.1.b, above.)

181. Moreover, Steuer389's claim 11 recites almost identical language as claim limitation 11.b except it substitutes the claimed video data with audio data: "receiving, at one or more computers, audio data from a monitoring device, the audio data being divisible into samples." (EX1006 at claim 11.)

182. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of the disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitation 11.b.

c)

- (1) determining, by the one or more computers, a next sample for attempted identification based at least in part upon a result of an attempt by the one or more computers to identify prior samples against a computerized database of known content**
- (2) wherein determining the next sample comprises selecting a relative time location of the next sample within the video data such that the relative time location of the next sample is selected to be between a relative time location of the first prior sample and a relative time location of the second prior sample based on a result of attempting to identify a first prior sample against the computerized database of known content being different from a result of attempting to identify a second prior sample against the computerized database of known content;**

183. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation.

184. First, Steuer389's claim 11 recites almost identical language as claim limitation 11.c.1 and 11.c.2, albeit in the context of "audio data" as noted with respect to limitation 11.b: "determining, using the one or more computers, a next sample for attempted identification based at least in part upon a result of an attempt using the one or more computers to identify prior samples against a computerized database of known content wherein determining the next sample comprises selecting

a relative time location of the next sample within the audio data such that if a result of attempting to identify a first prior sample against the computerized database of known content differs from a result of attempting to identify a second prior sample against the computerized database of known content, then the relative time location of the next sample within the audio data is selected to be between a relative time location of the first prior sample and a relative time location of the second prior sample.” (EX1006 at claim 1.)

185. Further, with respect to claim limitation 11.c.1, Steuer750 and Steuer389 both disclose this limitation with reference to Figure 2 as they state, e.g.: “At step 206, **an identification result is used to adjust sampling control and selection parameters for one or more subsequent samples.** Parameters that might be adjusted include resolution, backtrack selection, sample time window, or recognition algorithm selected.” (EX1005 at [0037], [0063]-[0065] (similarly describing the adjustment of time window length of a next sample based on identification results of prior samples with respect to a computerized database of known contents), FIGS. 2, 6, claims 19, 24-28; EX1006 at 7:39-56, 13:20-56, FIGS. 2, 6.)

186. Further, with respect to claim limitation 11.c.2, Steuer750 and Steuer389 both disclose with reference to Figure 2 that “if an identification result [of

a sample] **does not match** the result of the prior sample identified, then step 206 might initiate a **back track analysis** and **select an intervening sample that has not yet** Steuer750 and Steuer389 both disclose **been analyzed.**” (EX1005 at [0037]); EX1006 at 7:39-56, FIG. 2.) Similarly, Steuer750 and Steuer389 both disclose with reference to Figures 2 and 6 that “if the n_2^{th} sample **does not match** the n_1^{th} sample, then process 200 can make a **backtrack decision and go back in the series of samples to select, for example, the $(n_1+6)^{\text{th}}$ sample for content recognition analysis** in order to pinpoint more closely where in the play stream the content change occurred,” where “[t]he order in which samples are selected for analysis in this example is illustrated by arrows 1, 2, and 3 in FIG. 6.” (EX1005 at [0064], FIGS. 2, 6; EX1006 at 13:23-56, FIGS. 2, 6.)

At step 206, an identification result is used to adjust sampling control and selection parameters for one or more subsequent samples. Parameters that might be adjusted include resolution, backtrack selection, sample time window, or recognition algorithm selected. For example, if a sample has been identified successfully, and the identification indicates a match with the prior sample, the sampling resolution might be decreased from $\frac{1}{5}$ to $\frac{1}{10}$, meaning that, in a sequence of samples, nine samples rather than four are skipped before selecting the next sample to analyze for identification. **As**

another example, if an identification result does not match the result of the prior sample identified, then step 206 might initiate a back track analysis and select an intervening sample that has not yet been analyzed. Step 201 would then pass the sample selected from backtracking to step 202 for attempted identification. As another example, resolution might be increased when recent identification results suggest an extended period of silence or absence of known content.

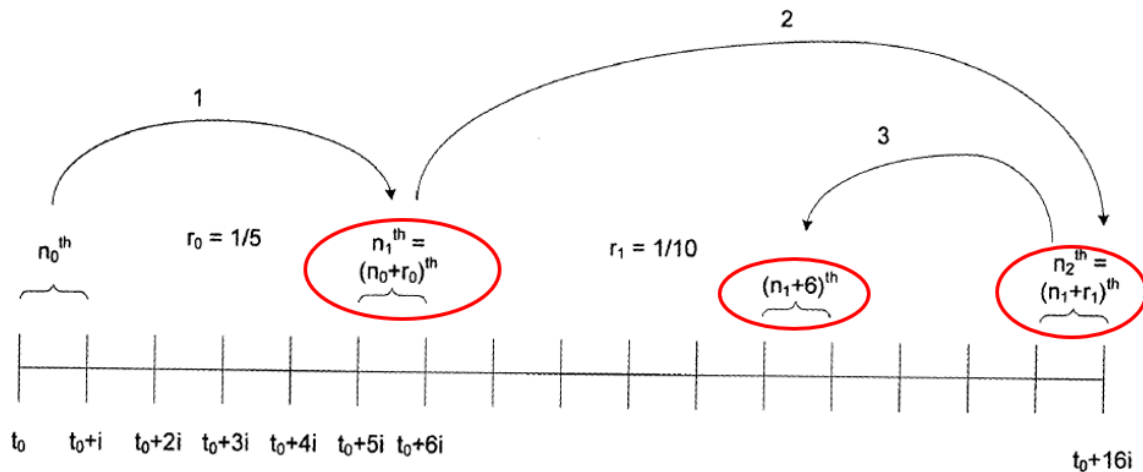
(EX1005 at [0037]; EX1006 at 7:39-56.)

FIG. 6 illustrates timeline structure for data samples generated by system and methods such as module 121 of FIG. 1 **and steps 201 and 206 of FIG. 2.**

As illustrated, sample set 600 reflects a division of audio data into sixteen samples, three of which are selected for analysis. The samples selected for recognition analysis are referenced generally as the “ n_0^{th} ,” “ n_1^{th} ,” and “ n_2^{th} ” samples. Samples between the selected samples are not separately numbered, but are marked off between vertical lines along the horizontal length of timelines structure 600 as illustrated. Each sample has a sample time window length “ i .” Furthermore, the sample set has varying “resolutions” including $r_0=1/5$ between the n_0^{th} and n_1^{th} samples and $r_1=1/10$ between the n_1^{th} and the n_2^{th} samples. In other words, “resolution” here refers to the portion of samples being selected during a period of time. In the present example, a

process portion such as steps 201 and 206 of process 200 of FIG. 2 has, based on the identification results of the n_0^{th} and n_1^{th} samples, adjusted the resolution from $\frac{1}{5}$ to $\frac{1}{10}$. This could occur, for example, as previously described, because the n_0^{th} and n_1^{th} samples match the same content ID and process 200 determines that the likelihood that the next selected sample using a resolution of $\frac{1}{5}$ would also match justifies decreasing the resolution to $\frac{1}{10}$ so that the next sample selected for analysis is the n_2^{th} . **However, if the n_2^{th} sample does not match the n_1^{th} sample, then process 200 can make a backtrack decision and go back in the series of samples to select, for example, the $(n_1+6)^{\text{th}}$ sample for content recognition analysis in order to pinpoint more closely where in the play stream the content change occurred. The order in which samples are selected for analysis in this example is illustrated by arrows 1, 2, and 3 in FIG. 6.**

(EX1005 at [0063]-[0064]; EX1006 at 13:20-50.)



600

FIG. 6

(EX1006 at FIG. 6 (annotated); EX1006 at FIG. 6.)

187. Thus, Steuer750 and Steuer389 both disclose determining the next sample comprises selecting a relative time location of the next sample (e.g., an intervening sample such as the $(n_1+6)^{th}$ sample) within the audio data such that the relative time location (e.g., (n_1+6)) of the next sample is selected to be between a relative time location of the first prior sample (e.g., time location of the n_1^{th} sample) and a relative time location of the second prior sample (e.g., time location of the n_2^{th} sample) based on a result of attempting to identify a first prior sample (e.g., the n_1^{th}

sample) against the computerized database of known content being different from a result of attempting to identify a second prior sample (e.g., the n_2^{th} sample) against the computerized database of known content, as recited in limitation 1.c.2.

188. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of the disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Indeed, as discussed above, Eldering discloses adjusting the parameters for content identification based on results associated with identifying prior samples, similar to Steuer750's and Steuer389's disclosures above. (EX1007 at [0090]-[0093], [0097], FIGS. 11-14.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitations 11.c.1-2.

- d) submitting, by the one or more computers, the determined next sample for attempted content identification; and**

189. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for the same reasons as discussed above for limitation 1.e. (*See* Section IX.A.1.e, above.) I note that for claim 1, I cited step 206 of Figure 2 for its adjustment of the time length window of the next sample based on prior content identification results, for example, while for claim 11, step 206 is cited for selection of an intervening sample as the next sample based on prior content identification results. (*See* Sections IX.A.1.c-d, IX.A.11.c.1-2, above.) A person of ordinary skill in the art would have understood that the subsequent processes for submitting the determined next sample for attempting content identification using steps 201 and 202 operate the same regardless of the selection parameter being adjusted for the next sample in step 206.

190. Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest claim limitation 11.d.

- e) **generating, by the one or more computers, a play stream of content identifications based on results of attempts to identify a series of samples comprising the determined next sample.**

191. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for the same reasons as discussed above for limitation 1.f. (*See* Section IX.A.1.e, above.) I note that for claim 1, I cited step 206 of Figure 2 for its adjustment of the time length window of the next sample based on prior content identification results, for example, while for claim 11, step 206 is cited for selection of an intervening sample as the next sample based on prior content identification results. (*See* Sections IX.A.1.c-d, IX.A.11.c.1-2, above.) A person of ordinary skill in the art would have understood that the subsequent processes for generating a play stream (e.g., by way of steps 207 and 209) operate the same regardless of the selection parameter being adjusted for the next sample in step 206.

192. Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest claim limitation 11.e.

12. Claim 12

- a) The method of claim 11 further comprising, for a series of samples of the video data: determining a sample by selecting an $(n \pm r)$ 'th sample for to be submitted for content recognition wherein r is determined at least in part based upon an identification result corresponding to the n 'th sample.**

193. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for at least reasons similar to those given above with respect to limitations 11.c.1-2. (*See* Section IX.A.11.c, above.)

194. First, Steuer389's claim 12 recites almost identical language as claim limitation 12.a, albeit in the context of "audio data" as noted with respect to limitation 11.b: "The method of claim 11 further comprising, for a series of samples of the audio data: determining a sample by selecting an $(n \pm r)$ 'th sample for to be submitted for content recognition wherein r is determined at least in part based upon an identification result corresponding to the n 'th sample." (EX1006 at claim 12.)

195. Further, as discussed above with respect to limitations 11.c.2, Steuer750 and Steuer389 both disclose that for a series of samples of the audio data as described with reference to step 206 of Figure 2 and related Figure 6, the content recognition process includes determining a sample by selecting an $(n \pm r)$ 'th sample (e.g., an intervening sample such as the $(n_1 + 6)^{\text{th}}$ sample) to be submitted for content

recognition wherein r (e.g., “6” in this example) is determined at least in part based upon an identification result corresponding to the n 'th sample (e.g., the n_1 th sample or the n_2 th sample). (See Section IX.A.11.c, above; EX1005 at [0064], FIGS. 2, 6, claims 19, 24-28 (e.g., claim 24 disclosing “determining a sample comprises, for an $(n+r)$ 'th sample submitted for content recognition, determining r at least in part based upon an identification result corresponding to the n 'th sample”); EX1006 at 13:23-56, FIGS. 2, 6.)

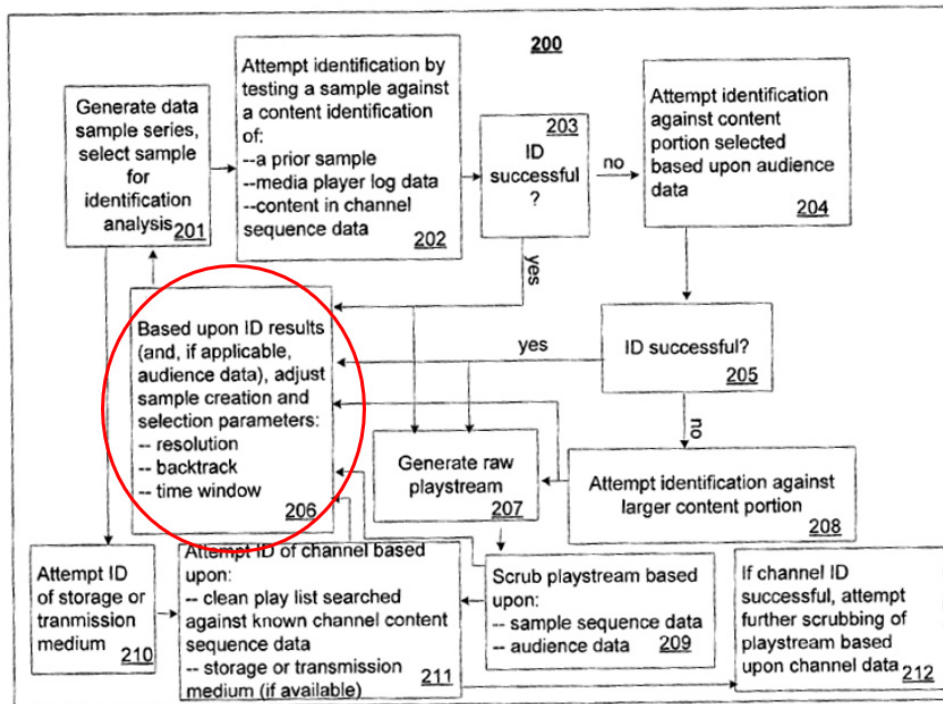
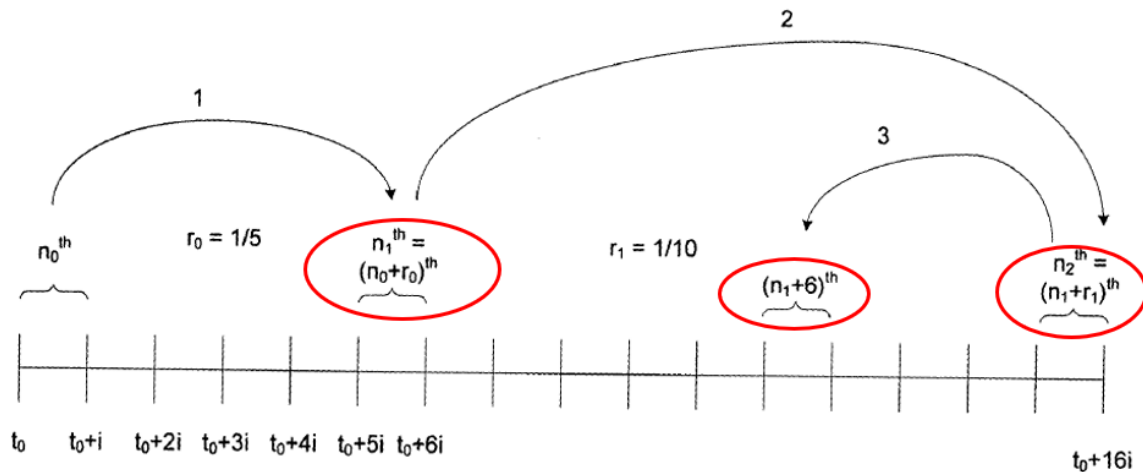


FIG. 2

(EX1005 at FIG. 2 (annotated); EX1006 at FIG. 2.)



600

FIG. 6

(EX1006 at FIG. 6 (annotated); EX1006 at FIG. 6.)

196. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of the disclosed audio-based approach. (See Sections IX.A.1.b-c, above.) Indeed, as discussed above, Eldering discloses adjusting the parameters for content identification based on results

associated with identifying prior samples, similar to Steuer750's and Steuer389's disclosures above. (EX1007 at [0090]-[0093], [0097], FIGS. 11-14.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitations 12.a.

13. Claim 13

- a) The method of claim 12 wherein the magnitude of r is increased from a prior value if the identification result corresponding to the n 'th sample does not indicate an identification.**

197. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for at least reasons similar to those given above with respect to limitations 11.c.1-2 and 12.a. (*See* Sections IX.A.11.c, IX.A.12.a, above.)

198. For example, Steuer389's claim 13 recites almost identical language as claim limitation 13.a, albeit in the context of "audio data" as noted with respect to limitation 11.b: "wherein the magnitude of r is increased from a prior value if the identification result corresponding to the n 'th sample does not indicate an identification." (EX1006 at claim 13; *see also id.*, 7:39-56, 13:23-56, FIGS. 2, 6.)

199. Steuer750 similarly discloses in the context of "audio data" that "wherein the value of r is increased from a prior value if the identification result

corresponding to the n'th sample does not indicate an identification.” (EX1005 at claim 25; *see also id.*, claims 19, 24, 26-28, [0037], [0064], FIGS. 2, 6.)

200. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of the disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Indeed, as discussed above, Eldering discloses adjusting the parameters for content identification based on results associated with identifying prior samples, similar to Steuer750's and Steuer389's disclosures above. (EX1007 at [0090]-[0093], [0097], FIGS. 11-14.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitations 13.a.

14. Claim 14

- a) The method of claim 12 wherein the magnitude of r is increased from a prior value if the identification result corresponding to the n'th sample indicates absence of any media content.**

201. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for at least reasons similar

to those given above with respect to limitations 11.c.1-2 and 12.a. (*See* Sections IX.A.11.c, IX.A.12.a, above.)

202. For example, Steuer389's claim 14 recites almost identical language as claim limitation 13.a, albeit in the context of "audio data" as noted with respect to limitation 11.b: "wherein the magnitude of r is increased from a prior value if the identification result corresponding to the n 'th sample indicates absence of any media content." (EX1006 at claim 14; *see also id.*, 7:39-56, 13:23-56, FIGS. 2, 6.)

203. Steuer750 similarly discloses in the context of "audio data" that "wherein the value of r is increased from a prior value if the identification result corresponding to the n 'th sample indicates absence of any media content." (EX1005 at claim 26; *see also id.*, claims 19, 24-25, 27-28, [0037], [0064], FIGS.2, 6.)

204. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of the disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Indeed, as discussed above, Eldering discloses adjusting the parameters for content identification based on results associated with identifying prior samples, similar to Steuer750's and Steuer389's

disclosures above. (EX1007 at [0090]-[0093], [0097], FIGS. 11-14.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitations 14.a.

15. Claim 15

- a) **The method of claim 14 further comprising: depending upon identification results corresponding to an n 'th and an $(n \pm r)$ 'th sample, determining whether any samples captured between the n 'th and the $(n \pm r)$ 'th sample should be submitted for a content recognition attempt.**

205. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for at least reasons similar to those given above with respect to limitations 11.c.1-2 and 12.a. (*See* Sections IX.A.11.c, IX.A.12.a, above.)

206. For example, Steuer389's claim 15 recites almost identical language as claim limitation 13.a, albeit in the context of "audio data" as noted with respect to limitation 11.b: "depending upon identification results corresponding to an n 'th and an $(n \pm r)$ 'th sample, determining whether any samples captured between the n 'th and the $(n \pm r)$ 'th sample should be submitted for a content recognition attempt." (EX1006 at claim 15; *see also id.*, 7:39-56, 13:23-56, FIGS. 2, 6.)

207. Steuer750 similarly discloses in the context of "audio data" that "wherein determining comprises, depending upon identification results

corresponding to an n 'th and an $(n+r)$ 'th sample, determining whether any samples captured between the n 'th and the $(n+r)$ 'th sample should be submitted for a content recognition attempt.” (EX1005 at claim 27; *see also id.*, claims 19, 24-26, 28, [0037], [0064], FIGS. 2, 6.)

208. Moreover, as discussed above, Steuer389 and Steuer750 both disclose depending upon identification results corresponding to the n_1^{th} sample (“ n 'th [sample]”) and the n_2^{th} sample (“ $(n\pm r)$ 'th sample”), determining whether any samples captured between the n 'th and the $(n\pm r)$ 'th sample (e.g., an intervening sample such as the $(n_1+6)^{\text{th}}$ sample shown in Figure 6) should be submitted for a content recognition attempt, as required by limitation 15.a. (EX1005 at [0064], FIGS. 2, 6; EX1006 at 13:23-56, FIGS. 2, 6.)

209. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of the disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Indeed, as discussed above, Eldering discloses adjusting the parameters for content identification based on results associated with identifying prior samples, similar to Steuer750's and Steuer389's

disclosures above. (EX1007 at [0090]-[0093], [0097], FIGS. 11-14.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitations 15.a.

16. Claim 16

- a) The method of claim 15 further comprising: submitting one or more samples captured between the n'th and the (n±r)'th sample for a content identification attempt if the identification results corresponding to the n'th and (n±r)'th sample do not match.**

210. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for at least reasons similar to those given above with respect to limitations 11.c.1-2, 12.a, and 15.a. (*See* Sections IX.A.11.c, IX.A.12.a, IX.A.15.a, above.)

211. For example, Steuer389's claim 16 recites almost identical language as claim limitation 16.a, albeit in the context of "audio data" as noted with respect to limitation 11.b: "submitting one or more samples captured between the n'th and the (n±r)'th sample for a content identification attempt if the identification results corresponding to the n'th and (n±r)'th sample do not match." (EX1006 at claim 16; *see also id.*, 7:39-56, 13:23-56, FIGS. 2, 6.)

212. Steuer750 similarly discloses in the context of "audio data" that "submitting one or more samples captured between the n'th and the (n+r)'th sample

for a content identification attempt if the identification results corresponding to the n 'th and $(n+r)$ 'th sample do not match.” (EX1005 at claim 18; *see also id.*, claims 19, 24-27, [0037], [0064], FIGS. 2, 6.)

213. Moreover, as discussed above, Steuer389 and Steuer750 both disclose submitting an intervening sample such as the (n_1+6) th sample shown in Figure 6 (“one or more samples captured between the n 'th and the $(n\pm r)$ 'th sample”) captured between the n_1 th sample (“ n 'th [sample]”) and the n_2 th sample (“ $(n\pm r)$ 'th sample”) for a content identification attempt if the identification results corresponding to the n 'th and $(n\pm r)$ 'th sample do not match. (EX1005 at [0064], FIGS. 2, 6; EX1006 at 13:23-56, FIGS. 2, 6; *see also* Section IX.A.11.d, above.)

214. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of the disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Indeed, as discussed above, Eldering discloses adjusting the parameters for content identification based on results associated with identifying prior samples, similar to Steuer750's and Steuer389's disclosures above. (EX1007 at [0090]-[0093], [0097], FIGS. 11-14.) Thus, the

Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitations 16.a.

17. Claim 17

- a) **The method of claim 11 wherein: if the result of attempting to identify the first prior sample against the computerized database of known content does not differ from the result of attempting to identify the second prior sample against the computerized database of known content, then the relative time location of the next sample within the video data is selected to be after relative time locations of the first and second prior samples.**

215. In my opinion, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest this limitation for at least reasons similar to those given above with respect to limitations 11.c.1-2 and 12.a. (*See* Sections IX.A.11.c, IX.A.12.a, above.)

216. First, Steuer389's claim 17 recites almost identical language as claim limitation 16.a, albeit in the context of "audio data" as noted with respect to limitation 11.b: "wherein: if the result of attempting to identify the first prior sample against the computerized database of known content does not differ from the result of attempting to identify the second prior sample against the computerized database of known content, then the relative time location of the next sample within the audio

data is selected to be after relative time locations of the first and second prior samples.” (EX1006 at claim 16; *see also id.*, 7:39-56, 13:23-56, FIGS. 2, 6.)

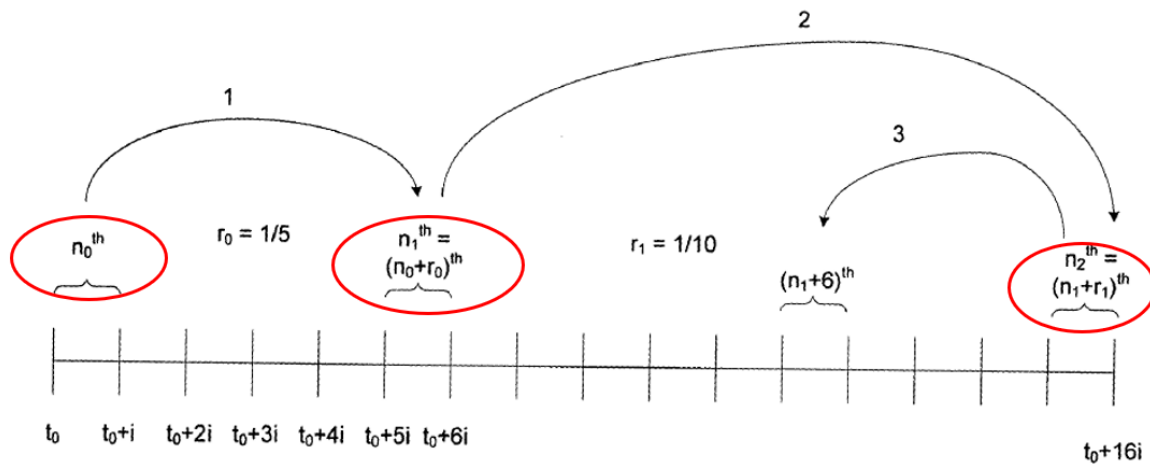
217. Moreover, Steuer389 and Steuer750 both disclose:

FIG. 6 illustrates timeline structure for data samples generated by system and methods such as module 121 of FIG. 1 and steps 201 and 206 of FIG. 2.

As illustrated, sample set 600 reflects a division of audio data into sixteen samples, three of which are selected for analysis. The samples selected for recognition analysis are referenced generally as the “ n_0^{th} ,” “ n_1^{th} ,” and “ n_2^{th} ” samples. Samples between the selected samples are not separately numbered, but are marked off between vertical lines along the horizontal length of timelines structure 600 as illustrated. Each sample has a sample time window length “i.” **Furthermore, the sample set has varying “resolutions” including $r_0=1/5$ between the n_0^{th} and n_1^{th} samples and $r_1=1/10$ between the n_1^{th} and the n_2^{th} samples.** In other words, “resolution” here refers to the portion of samples being selected during a period of time. **In the present example, a process portion such as steps 201 and 206 of process 200 of FIG. 2 has, based on the identification results of the n_0^{th} and n_1^{th} samples, adjusted the resolution from $1/5$ to $1/10$. This could occur, for example, as previously described, because the n_0^{th} and n_1^{th} samples match the same content ID and process 200 determines that the likelihood that the next selected sample using a resolution of $1/5$ would also match justifies decreasing the resolution to $1/10$ so**

that the next sample selected for analysis is the n_2^{th} . However, if the n_2^{th} sample does not match the n_1^{th} sample, then process 200 can make a backtrack decision and go back in the series of samples to select, for example, the $(n_1+6)^{\text{th}}$ sample for content recognition analysis in order to pinpoint more closely where in the play stream the content change occurred. The order in which samples are selected for analysis in this example is illustrated by arrows 1, 2, and 3 in FIG. 6.

(EX1005 at [0064], FIGS. 2, 6; *see also id.*, [0037] (“For example, if a sample has been identified successfully, and the identification indicates a match with the prior sample, the sampling resolution might be decreased from $\frac{1}{5}$ to $\frac{1}{10}$, meaning that, in a sequence of samples, nine samples rather than four are skipped before selecting the next sample to analyze for identification. As another example, if an identification result does not match the result of the prior sample identified, then step 206 might initiate a back track analysis and select an intervening sample that has not yet been analyzed.”); EX1006 at 13:23-50, FIGS. 2, 6; *see also id.*, 7:39-56.)



600

FIG. 6

(EX1006 at FIG. 6 (annotated); EX1006 at FIG. 6.)

218. Thus, Steuer389 and Steuer750 both disclose that if the result of attempting to identify the n_0^{th} sample (“first prior sample”) against the computerized database of known content does not differ from the result of attempting to identify the n_1^{th} sample (“second prior sample”) against the computerized database of known content, then the relative time location of the n_2^{th} sample (“next sample”) within the data is selected to be after relative time locations of the first and second prior samples, as required by claim 17.

219. Steuer750 and Steuer389 disclose these features in the context of its audio-based content identification systems and methods, but as I discussed above with reference to claim limitations 1.b-c, a person of ordinary skill in the art would have been motivated to implement Steuer750's and Steuer389's disclosed content identification systems using video data samples instead of the disclosed audio-based approach. (*See* Sections IX.A.1.b-c, above.) Indeed, as discussed above, Eldering discloses adjusting the parameters for content identification based on results associated with identifying prior samples, similar to Steuer750's and Steuer389's disclosures above. (EX1007 at [0090]-[0093], [0097], FIGS. 11-14.) Thus, the Steuer750-Eldering combination and the Steuer389-Eldering combination disclose or suggest limitations 17.a.

X. CONCLUSION

220. 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.

Dated: January 31, 2025

By: 
Timothy Wahlers