UNITED STATES PATENT AND TRADEMARK OFFICE

_____

BEFORE THE PATENT TRIAL AND APPEAL BOARD

_____

SAMSUNG ELECTRONICS CO., LTD.
Petitioner

v.

ANONYMOUS MEDIA RESEARCH HOLDINGS, LLC.
Patent Owner

_____

Case IPR2024-01348
U.S. Patent 10,572,896

_____

PATENT OWNER'S PRELIMINARY RESPONSE

TABLE OF CONTENTS

I. Introduction 6

II. Level of Ordinary Skill 10

III. Petitioner Fails to Show that the Challenged Claims Lack Written Description Support in the Prior Non-Provisional Applications. 11

A. Overview of Written Description Law 13

B. The Common Written Description Demonstrates Possession of the Claimed Invention. 14

1. The "BACKGROUND OF THE INVENTION" 17

2. State of the Art 21

3. The "SUMMARY" 26

4. The "DETAILED DESCRIPTION" 29

5. The Challenged Claims 38

C. Petitioner's Contrary Position Lacks Merit and Evidentiary Support. ... 40

1. Petitioner's Attempt to Rewrite the "BACKGROUND" Should be Rejected 41

2. Petitioner Improperly Ignores the Teachings of the "SUMMARY" 46

3. Petitioner's Reliance on the Number of Times "Audio" and "Video" Are Used Is Misplaced. 49

4. Petitioner Mischaracterizes the Nature of the Invention and the State of the Art 50

5. Petitioner's Arguments Regarding Alleged "New Matter" Are Wrong and Irrelevant. 58

6. The Testimony of Petitioner's Expert on the Issue of Written Description Should Be Given No Weight. 60

IV. Institution Should Be Denied Because the Petition Fails to Demonstrate Obviousness 62

V. Conclusion 71

EXHIBIT LIST

Exhibit No.	Description
2001	Expert Declaration of W. Leo Hoarty
2002	*Curriculum Vitae* of W. Leo Hoarty
2003	Computer-generated Redline Between As-Filed Specification in U.S. Patent Application No. 11/139,330 (Ex. 1011) and Specification of U.S. Patent 10,572,896 (Application No. 16/403,204) (Ex. 1001)
2004	Joshua Grossnickle et al. THE HANDBOOK OF ONLINE MARKETING RESEARCH (2001)
2005	U.S. Patent No. 3,919,479
2006	U.S. Patent No. 4,230,990
2007	U.S. Patent No. 4,697,209
2008	U.S. Patent No. 4,677,466
2009	U.S. Patent No. 4,739,398
2010	U.S. Patent No. 5,019,899
2011	U.S. Patent No. 5,436,653
2012	U.S. Patent No. 5,581,658
2013	U.S. Patent No. 6,675,383
2014	U.S. Patent No. 6,633,651
2015	RESERVED
2016	U.S. Patent No. 6,990,453
2017	U.S. Patent No. 7,289,643
2018	U.S. Patent App. Pub. No. 2002/0026635

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

Exhibit No.	Description
2101	Dec. 17, 2024 Decision on Petitions under 37 C.F.R. § 1.78 and § 1.182 in U.S. Application No. 16/403,204

I. Introduction

Petitioner carries the burden of demonstrating a reasonable likelihood that the challenged claims are unpatentable. Petitioner fails to carry its burden for at least two independent reasons. Institution should be denied.

First, Petitioner fails to carry its burden of demonstrating that the asserted references—the inventors' related published application and patent (Steuer750 and Steuer791, respectively) and Neumeier—qualify as prior art. Petitioner alleges that the challenged claims are, at best, entitled to the benefit of a filing date of November 24, 2010. The sole basis for this assertion is that the earlier provisional and non-provisional applications allegedly do not provide written description support for the challenged claims.[1] In trying to persuade the Board of this, Petitioner focuses largely on whether the challenged claims are supported by the provisional application and

[1] The Petition does not contend that the challenged claims lack entitlement to an earlier priority date for any other reason, and Petitioner may not raise new theories that could have been presented in the Petition. *Netflix, Inc. v. DivX, LLC*, 84 F.4th 1371, 1377 (Fed. Cir. 2023) ("[W]e have rejected, many times, post-hoc attempts on appeal to include additional, new arguments not contained in the petition.") (citing *Yita LLC v. MacNeil IP LLC*, 69 F.4th 1356, 1366-67 (Fed. Cir. 2023) (affirming Board refusal to consider new theories not raised in the petition)).

whether the provisional can be relied on at all. This, however, is an irrelevant distraction intended to draw the Board's attention away from a dispositive fact—all of the prior non-provisional applications (including the original non-provisional application filed on May 27, 2005) share a common written description that expressly teaches that the invention applies to both audio- and **video**-sample-based content recognition systems. *Infra* Section III.

The common written description plainly teaches that the invention applies "robust" and "**existing**" "audio or **video** data sample" content recognition systems, which the invention improves upon by using "data exogenous to a particular" sample to deduce additional information about audience members' consumption habits (e.g., play-altering activity). Consistent with this disclosure, the written description further teaches that the invention applies to "media sample[s]" (i.e., "audio or **video** data sample[s]"). Indeed, the disclosed invention is agnostic as to whether the content identification results are generated based on an audio or video data sample, because the technological novelty lies in the use of additional data to discern useful information beyond that generated by existing content recognition systems. Because the common written description (a) expressly teaches that the invention applies existing video-sample-based content recognition systems, (b) such systems were clearly known and understood in the art, and (c) the type of content recognition system that the invention applies is merely incidental to the invention, a person

having ordinary skill in the art would readily recognize the inventors to have possessed the claimed invention based on the common written description. Because Petitioner fails to demonstrate that the challenged claims lack written description support in the common written description, Petitioner necessarily fails to demonstrate that any of the asserted references qualify as prior art. *Infra* Section III.B.

Petitioner cannot legitimately dispute any of this. Instead, it mischaracterizes the teachings from the written description and the state of the art. *Infra* Section III.C. As one of the most egregious examples, Petitioner and its expert resort to an implausible interpretation of what the common written description plainly teaches, arguing that when it expressly states that the invention applies "existing" content recognition systems based on "audio or **video** data sample[s]," this somehow only refers to "audio sample" content recognition systems. Petitioner also mischaracterizes the state of the art when it alleges—without any explanation—that the disclosed invention somehow operates or applies differently depending on whether the generated content results are based on audio or video data samples. Even a cursory review of the '896 Patent and challenged claims reveals that the system used to obtain a content identification result is incidental to the claimed invention—the novelty lies in what the claimed invention does with the results obtained using

existing, well-known content recognition systems.[2]

Second, institution should be denied regardless of the written description issue, because the Petition fails to demonstrate obviousness even assuming the asserted references qualify as prior art. *Infra* Section IV. As an initial matter, Petitioner cannot have it both ways. Given Petitioner's repeated representations to the Board that a POSITA would not have understood the inventors to possess the invention because of the so-called "obstacles" and "challenges" in applying Steuer750/Steuer791 to video data samples, Petitioner must then also address those so-called obstacles and challenges to demonstrate that it would have been obvious to combine Steuer750/Steuer791 with Neumeier. Petitioner instead only offers a superficial analysis that fails to address or explain how *any* of these so-called obstacles would be addressed when combining Neumeier's video-sample-based content recognition system with Steuer750 or Steuer791.

Beyond the glaring inconsistency between Petitioner's written description and obviousness positions, Petitioner does not provide any articulated rationale

[2] Roku also filed an IPR petition against the challenged claims. Notably, however, Roku did not allege that the challenged claims lack written description support in the May 2005 non-provisional application. *See Roku, Inc. v. Anonymous Media Research Holdings, LLC*, IPR2024-01056, Paper 1 (July 15, 2024).

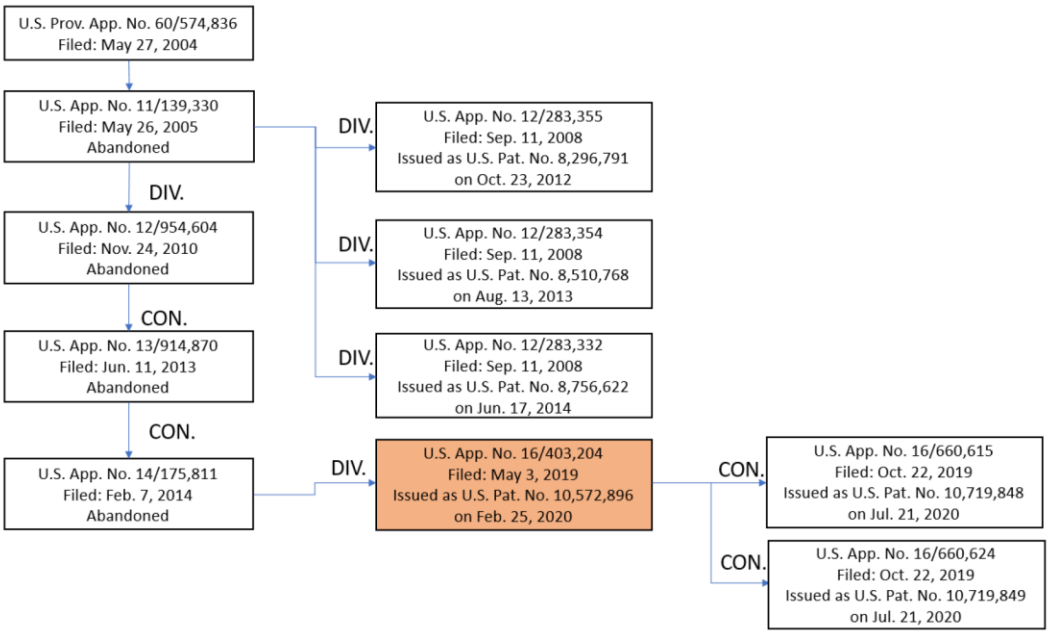
explaining *why* a POSITA would have modified Steuer750/Steuer791 to apply to the video-sample-based content recognition technology of Neumeier. Petitioner's obviousness arguments consist largely of generic statements that audio- and video-sample-based content recognition systems were known in the art and "could have" been combined—but the fact that claim features were independently known in the art and could hypothetically be combined does not constitute a rationale to combine those references. And although Petitioner includes two single-sentence obviousness rationales, these conclusory statements do not address the *actual* legal requirements for showing obviousness under those rationales (e.g., market pressure or recognized need in the art).

II. Level of Ordinary Skill

Patent Owner applies Petitioner's proposed definition of a POSITA for purposes of this proceeding only. Patent Owner reserves the right to propose its own level of ordinary skill, and application of Petitioner's proposal is not a concession or admission that the proposal is correct. Patent Owner's technical expert, Mr. Hoarty, has two years of college education in the field of electrical engineering, more than a decade of industry experience designing and implementing automatic content recognition ("ACR") systems, and numerous patents to his name covering both audio- and video-sample-based ACR technologies. Ex. 2001 ¶¶2-18, 53; Ex. 2002. He thus qualifies as a POSITA. Ex. 2001 ¶53.

III. Petitioner Fails to Show that the Challenged Claims Lack Written Description Support in the Prior Non-Provisional Applications.

A central focus of the Petition is convincing the Board that the ’836 provisional application does not provide written description support for the claims of the ’896 Patent, and further that the ’896 Patent is not entitled to rely on the ’836 provisional for written description support. *See generally*, Pet., 12-40. The Board need not resolve those issues, because the written description that is common to all priority non-provisional applications—including the ’330 application filed on May 26, 2005—clearly provides full written description support for the challenged claims. Specifically, the ’896 Patent claims priority to the ’836 provisional application through the ’811, ’870, ’604, and ’330 non-provisional applications (collectively the “Prior Non-Provisional Applications”) as illustrated below:



Pet. 3. The Prior Non-Provisional Applications all have identical written

descriptions, except for differences in the "CROSS REFERENCE TO RELATED APPLICATIONS." *See* Ex. 1017-1019 (redline comparisons of '330 specification to '604, '870, and '811 specification, respectively). The specification of the '896 Patent also contains this identical written description. *See* Ex. 2003 (redline comparison of '896 specification to '330 application specification). Thus, the entire written description appearing after the "CROSS REFERENCE TO RELATED APPLICTIONS" in the Prior Non-Provisional Applications also exists within the specification of the '896 Patent.[3]

Petitioner tellingly exerts little effort addressing this common written description. *See* Pet., 24-29. Even where it does, Petitioner and its expert offer implausibly strained interpretations of disclosures that plainly teach that the invention relies on existing audio- and video-sample-based content recognition systems. When the written description is viewed ***objectively*** from the perspective of a POSITA—not through the biased lens of Petitioner—it cannot reasonably be disputed that each Prior Non-Provisional Application, including the '330 application, demonstrates that the inventors possessed the invention as applied to video data

[3] In Section III.B below, Patent Owner cites to the specification of the '330 application (Ex. 1011) because it contains the disclosures common to all of the Prior Non-Provisional Applications and the '896 Patent, as demonstrated by the redlines.

samples. Because Petitioner fails to demonstrate that the challenged claims are not entitled to the benefit of the ’330 application due to a lack of written description support, Petitioner fails to demonstrate that the asserted references qualify as prior art.

A. Overview of Written Description Law

“Sufficiency of written description is a question of fact, reviewed for substantial evidence.” *General Hosp. Corp. v. Sienna Biopharmaceuticals, Inc.*, 888 F.3d 1368, 1371 (Fed. Cir. 2018). “[W]ritten description is about whether the skilled reader of the patent disclosure can recognize that what was claimed corresponds to what was described; it is not about whether the patentee has proven to the skilled reader that the invention works, or how to make it work, which is an enablement issue.” *Alcon Research Ltd. v. Barr Labs., Inc.*, 745 F.3d 1180, 1191 (Fed. Cir. 2014). “Such description need not recite the claimed invention in haec verba but must do more than merely disclose that which would render the claimed invention obvious.” *ICU Med., Inc. v. Alaris Med. Sys.*, 558 F.3d 1368, 1377 (Fed. Cir. 2009). The written description requirement is satisfied “when ‘the essence of the original disclosures’ conveys the necessary information—‘regardless of *how* it’ conveys such information.” *Inphi Corp. v. Netlist, Inc.*, 805 F.3d 1350, 1354 (Fed. Cir. 2015) (emphasis in original).

“The descriptive text needed to meet these requirements varies with the nature

and scope of the invention at issue, and with the scientific and technologic knowledge **already in existence**."[4] *Capon v. Eshar*, 418 F.3d 1349, 1357 (Fed. Cir. 2005). In other words, "[t]he 'written description' requirement must be applied **in the context of the particular invention** and **the state of the knowledge**." *Id.* at 1358; *See also LizardTech, Inc. v. Earth Res. Mapping, Inc.*, 424 F.3d 1336, 1345 (Fed. Cir. 2005) (explaining that "**only enough** must be included to convince a person of skill in the art that the inventor possessed the invention" with the understanding that "such a person comes to the patent **with the knowledge of what has come before**").

B. The Common Written Description Demonstrates Possession of the Claimed Invention.

The common written description demonstrates that the inventors possessed the disclosed invention as applied to video data samples. As explained in greater detail in the following subsections, the "BACKGROUND" expressly teaches that "robust content recognition technologies" had already been developed and that the disclosed invention "**appl[ies] <u>existing</u> content recognition** technologies," improving upon them "by utiliz[ing] data exogenous to a particular audio **<u>or video data sample</u>**" being analyzed. *Infra* Section III.B.1. In other words, audio- and

[4] All emphasis herein has been added, unless otherwise specifically stated.

video-sample-based content recognition technologies were well-known and understood in the art, and the invention simply relies on existing content recognition technologies for the generation of content identification results that the invention utilizes. *Infra* Section III.B.2. The invention improves upon those existing systems by analyzing the identification results in conjunction with additional data to generate additional useful information about the media consumption habits of audience members (e.g., play-altering actions of an audience member). The principles of operation of the disclosed invention are thus agnostic as to whether an audio- or video-sample-based content recognition system is being applied—the invention takes the content identification results output from those existing systems (e.g., a set of content identifiers), and the novelty lies in the utilization of additional data to generate further useful information about audience media consumption habits.

Consistent with this understanding, the "SUMMARY" of the invention does not merely describe applying the invention to "audio samples" but instead refers to the invention as applying to "media sample[s]" (audio and/or **video** data samples). *Infra* Section III.B.3. Given that "media sample[s]" refers to audio and **video** data samples and that the "BACKGROUND" expressly refers to the invention being applied to content recognition systems that rely on audio **or video** data samples, there can be no legitimate dispute that the written description expressly teaches that the disclosed invention is applicable to video-sample-based content recognition

systems. This is especially true given that Petitioner does not dispute—nor can it—that audio- and video-data-based sampling systems were well-known and understood in the art or that the principles of operation of the disclosed invention apply equally regardless of whether audio or video data samples are being analyzed. Indeed, Petitioner points to nothing in the common written description suggesting or indicating that the inventors were inventing a new fingerprint matching technology—the written description instead plainly states that the disclosed features apply those existing technologies and improves upon them by generating additional useful information about audience member media consumption.

Although Petitioner makes much ado about the "DETAILED DESCRIPTION" focusing on audio samples, Petitioner inappropriately disregards that (a) the written description expressly teaches that the invention applies to both audio and video data samples, (b) that video-sample-based content recognition systems were well-known and understood, (c) that the principles of operation explained with respect to the Figures 1-6 embodiment apply equally to these alternative video-data-based embodiments, and (d) that **the written description says as much.** *Infra* Section III.B.4. In other words, there was no need for the inventors to provide a detailed example of the invention as applied to video data samples, because the principles of operation apply in the same manner to the content identification results of audio- and video-sample-based content recognition systems,

which again the written description teaches were "existing" and "robust."

In short, when the common written description is viewed objectively, it cannot be reasonably disputed that it expressly teaches that the disclosed invention applies to both audio- and video-sample-based content recognition systems. Such systems were well-known and understood and the invention applies to both in the same manner according to the same principles of operation. Patent Owner addresses each section of the common written description in detail below.

1. The "BACKGROUND OF THE INVENTION"

Turning first to the "BACKGROUND," it provides a lengthy discussion of then-existing "media measurement systems." Ex. 1011, 577-578 (¶¶[0002]-[0005]). POSITAs understood "media measurement" to be an industry term that refers to the monitoring of media consumption habits of audience members. Ex. 2001 ¶57 (citing Ex. 2004, 11). The data acquired by such systems is then used to help broadcasters and advertisers measure the level of consumption of particular content (e.g., advertisements, TV shows), which can help determine future content placement. *Id.*

The "BACKGROUND" first describes a drawback with "channel-centric" media measurement. Ex. 1011, 577 (¶[0002]). A POSITA would understand that "channel-centric media measurement" typically involves monitoring the channel (e.g., radio station or television channel) to which an audience member is tuned. Ex. 2001 ¶58. The only insight provided by such monitoring is which channel audience

members are listening to or watching—it cannot provide insights on whether the audience member takes play-altering actions (e.g., muting, skipping) or whether the audience member consumed media (e.g., audio or video) over a different delivery mechanism (e.g., radio broadcast rather than television broadcast). *Id*.

To address some of the drawbacks of channel-centric measurement, the "BACKGROUND" explains that "[s]ome media measurement systems have used codes to 'tag' and track particular content. However, such systems are limited in that they can only track content that has been properly encoded." Ex. 1011, 577 (¶[0003]). In other words, although "tagging" allows for the tracking of some content consumed by an audience member (not simply the identity of the channel being listened to and/or watched), a drawback of such systems is the fact that the content itself must be encoded with the tags. Ex. 2001 ¶59.

To overcome drawbacks with "tagging," the "BACKGROUND" explains:

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

Ex. 1011, 577-578 (¶[0004]). A POSITA would have understood that this passage is teaching that "robust content recognition technologies" had been developed and existed to identify content by comparing data samples to reference samples associated with known content (e.g., TV shows, songs, commercials). Ex. 2001 ¶60. Such content recognition eliminates the need for tags, because the content is being identified based on a sample of the audio or video data content itself, rather than an embedded code or tag identifying the content. *Id*. ¶¶60-61. While the passage provides examples of commercially marketed content identification systems, the passage makes clear that content recognition is applied in "media measurement systems" where a "large database of **media** content" is searched against, and a POSITA would understand "media content" to encompass both audio- and video-data-based content. *Id.* ¶62. As the passage goes on to explain, such systems "can be computationally intensive" given that a large database of media content may be required for matching purposes. *Id.* ¶63.

The "BACKGROUND" concludes with a key passage discussing the intended use and purpose of the invention:

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

improved media measurement system and method **is needed**.

Ex. 1011, 578 (¶[0005]). A POSITA would readily understand this to mean that the intent of the inventors is to "apply[] **existing** content recognition technologies"—including existing audio and video data sampling technologies—in conjunction with additional data beyond the content identification results generated by those systems. Ex. 2001 ¶¶64-65; *See also* Ex. 1011, 584 (¶[0025]) (a detailed embodiment "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**.").

A POSITA would understand that the reference to "audio … data sample" is referring to the application of these robust known content recognition systems to identify content by sampling an audio signal and comparing the sample to reference audio samples, while "video data sample" is referring to the application of these robust known content recognition systems to identify content by sampling a video signal and comparing the sample to reference video samples. Ex. 2001 ¶65. Stated another way, a POSITA would understand that the "BACKGROUND" teaches that a purpose of the invention is to apply existing content recognition technologies that identify content based on audio or video data samples, and then "utilize data exogenous to a particular audio or video data sample" to further improve those

systems. *Id.* ¶66. The novel features described are thus agnostic as to whether audio- or video-sample-based content recognition systems are being used to generate content identification results—the improvement lies in the use of additional data to improve the content identification results generated based on either of these existing technologies. *Id.*

2. State of the Art

"The 'written description' requirement must be applied in the context of the particular invention and **the state of the knowledge**." *Capon v. Eshar*, 418 F.3d 1349, 1357-58 (Fed. Cir. 2005) (determining that the Board erred by "refusing to consider the state of the scientific knowledge" and by finding a lack of written description because the specification did not describe structure known in the art). It cannot be disputed that the state of a POSITA's knowledge was such that audio- and video-sample-based content recognition systems were well-known and understood in the art as of May 26, 2005. Ex. 2001 ¶¶67-68. Indeed, the "BACKGROUND" explicitly states as much, and the detailed embodiment teaches that the disclosed system is "adapted to implement one or more **known** content recognition methods that identify content by extracting parameters from a **media** signal and applying an algorithm to those parameters to search for a content match. **Those skilled in the art will recognize that many such methods and algorithms exist**." Ex. 1011, 584 (¶[0025]); Ex. 2001 ¶69.

In fact, the common written description discusses a variety of these "known" content recognition systems as applied to *both* audio and video data samples. Ex. 1011, 584-85 (¶[0025]). For example, WO 2004/044820 describes content recognition as applied to audio and video data samples:

> For audio signals, the perceptual features are energies of the audio contents in selected sub-bands. For image signals, the perceptual features are average luminances of blocks into which the image is divided. The conversion into a binary sequence is performed by thresholding, for example, by comparing each feature sample with its neighbors.

Ex. 2026, 2:16-20.[5] The reference then proceeds to discuss various considerations applicable to audio-sample-based content recognition and explains that "[s]imilar considerations apply to image and video material and to other kinds of perceptual features being used for fingerprint extraction." *Id.* at 3:17-18. Likewise, WO 2004/040475 teaches that video-sample-based content recognition was known, and that "[w]hilst the above embodiments of the present invention have been described with reference to audio information streams, it will be appreciated that the invention can be applied to other information signals, particularly multi-media signals, **including video signals**." Ex. 2024, 13:19-22 (discussing and citing video-sample-

[5] Citations to patent application publications is to [PDF page number]:[line].

based content recognition systems). Thus, not only was audio- and video-sample-based content recognition well-known and understood, it was also well-known and understood that their principles of operation were similar. Ex. 2001 ¶¶69-71.

Even beyond the common written description stating that these content recognition systems were well-known and understood, countless other references evidence as much. One of the earliest provided by Patent Owner—U.S. Patent No. 4,230,990 ("Lert")—was filed in 1979 and explains that sample-based automated content identification was understood in the art and involves sampling an audio or video signal (e.g., radio station signal for audio, television broadcast signal for video) to extract features ("signatures") for comparison to reference signatures, which allows the content of the signal to be identified. Ex. 2006, 3:47-4:16; Ex. 2001 ¶73. Lert thus demonstrates that, ***more than 25 years before the '330 application was filed***, the operation and implementation of audio- and video-sample-based content recognition was well-known. Ex. 2001 ¶73.

A POSITA would have understood that for both audio- and video-sample-based automatic content recognition, a sample is taken and then features (commonly referred to as "fingerprints" or "signatures") of the sample are extracted and compared to features of reference samples to identify content. *Id.* ¶74. Specifically, content recognition using audio samples works by analyzing the audio signal to create a unique fingerprint which is then compared against a database of known

audio fingerprints to identify the source of the audio, like a song or commercial, by matching it to the closest match in the database. *Id.* ¶75. Essentially, audio-sample-based content recognition extracts distinctive features from the audio signal like frequency patterns, tempo, and amplitude to create a unique fingerprint that can be recognized, even with background noise or slight variations in quality. *Id*.

Content recognition using video samples works in a very similar manner. It extracts distinctive features from the video signal like color patterns, pixel arrangement, motion, and scene changes within the video frame to create a unique fingerprint that can be compared against a database of known video fingerprints to identify the source of the video, like a TV show or commercial. Ex. 2001 ¶76. A POSITA would have understood that when the final passage of the "BACKGROUND" refers to the application of existing content recognition systems to "audio or video data sample[s]," it is referring to these types of known content recognition systems. *Id.* ¶77.

Countless other references spanning the decades preceding the '330 application only further demonstrate how well-known and understood audio- and video-sample-based content recognition was. *See, e.g.*:

Date	**Exhibit**	**Exemplary Disclosures**	
		Video Recognition	**Audio Recognition**
1975	2005	Abstract, 2:57-3:17	*Id.*
1980	2006	3:45-4:16, 9:53-62	*Id.*
1984	2007	Abstract, 3:1-17, 4:27-5:12	*Id.*, 10:29-35

1985	2008	Abstract, 3:37-5:19.	*Id.*
1986	2009	3:35-55.	*Id.*
1988	2010	Abstract, 4:14-55.	N/A
1992	2011	Abstract, 10:12-31.	*Id.*
1993	2012	5:9-6:3	*Id.*
1997	2013	10:27-43, 10:61-67	*Id.*
1998	2014	1:38-48, 1:58-60	N/A
2001	2016	5:36-46	Abstract
2001	2017	3:59-63, 6:14-19, 11:14-19	*Id.*
2001	2018	[0027]	[0027], [0049]
2002	2019	10:22-33	4:2-13, 10:22-33
2002	2020	2:9-11, 6:31-33	2:9-11, 4:24-28, 4:29-32
2002	2021	[0007], [0035]	*Id.*
2003	2022	[0019], [0020]	[0020]
2003	2023	[0002], [0027]	[0002], [0010]
2003	2024	Abstract, 1:14-22	Abstract, 1:14-22, 11:19-22
2003	2025	5:38	5:38, 4:25-29
2003	2026	6:18-21	2:21-24
2003	2027	Abstract, [0010], [0011]	*Id.*
2004	2028	Abstract, 2:52-62	4:61-64
2004	2029	Abstract, 3:45-55, 3:64-67	N/A
2004	2035	[0003], [0011] [0012], [0042]	[0043]
2005	2030	Abstract, [0012]	Abstract, [0026]
2005	2031	Abstract, [0018], [0019]	*Id.*
2005	2032	Abstract 4:26-40	4:50-53
2005	2033	Abstract, [0039]	[0041]

Ex. 2001 ¶78 (discussing table and providing version with text of references in Appendix A).

In addition to these references demonstrating that audio- and video-sample-based content recognition systems were well-understood, they also demonstrate that

POSITAs understood that systems built upon existing audio-sample-based content recognition could easily be applied to video-sample-based content recognition, and *vice versa*. Ex. 2001 ¶¶79-83 (discussing *e.g.*, Ex. 2023 ¶¶[0009], [0027]; Ex. 2022 ¶[0019]; Ex. 2030 ¶[0050]; Ex. 2016, Abstract, 5:36-46; Ex. 2031 ¶[0031]). These references also evidence that both audio and video fingerprinting operate in fundamentally the same way, and they are typically described together without distinction. *Id.* ¶84 (citing *e.g.*, Ex. 2022 ¶[0020]; Ex. 2006, 3:45-4:17).

In short, the "BACKGROUND" clearly conveys that the invention relies on existing audio- and video-sample-based content recognition systems, that such systems were well-known and understood in the art, and that the novel improvement lies not in the use of those existing content recognition technologies, but in the use of additional data beyond the content identification results generated by such systems to deduce additional useful information about audience consumption (e.g., play-altering actions). *Id.* ¶85.

3. The "SUMMARY"

The "SUMMARY" further conveys to a POSITA that the inventors possessed the invention as applied to both audio and video data samples. Immediately following the final passage of the "BACKGROUND," the "SUMMARY" recites:

> **Some embodiments of the present invention provide a media measurement system and method that <u>enhances recognition</u> (e.g. in terms of accuracy or efficiency) of the content of a <u>media sample</u>**

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

Ex. 1011, 578 (¶[0006]). A POSITA would understand that the first sentence of this passage—which, again, immediately follows the conclusion of the "BACKGROUND"—teaches that the disclosed invention relies on existing audio- and video-sample-based content recognition systems and "enhances" their recognition using information external to the sample itself. Ex. 2001 ¶86. This is not only because this statement immediately follows, and is consistent with, the

"BACKGROUND" statement that "the **media** measurement context provides opportunities to utilize data **exogenous** to a particular **audio or video data sample**," but also because the first sentence teaches recognizing the content of a "media sample," which refers to both audio and/or video data samples. *Id.* (citing Ex. 2034, 3:66-4:1 (explaining that the "media sample[s] may be audio or video samples")).

A POSITA would understand that: (1) "media" refers to both audio and video and "media sample" refers to both audio and video data samples; (2) both audio and video data samples are **expressly referenced** in the immediately preceding "BACKGROUND" passage as being "media"; and (3) as explained in greater detail below, the disclosed invention is agnostic as to whether a media measurement system relies on audio or video data samples—the invention improves both existing audio- and video-sample-based content recognition systems in precisely the same way by using additional data beyond the results generated by those systems. *Id.* ¶87.

The remainder of the "SUMMARY" is consistent with this understanding. It repeatedly refers to a "media" measurement system, "sample sequence data" and "sequence data," which encompasses both audio- and video-sample-based data. *Id.* ¶88. Although the final sentence does mention an "audio signal" and not "video signal," a POSITA would understand that is merely one embodiment because the sentence begins with "[y]et **other** embodiments." *Id.* ¶89.

As further explained below, a POSITA would understand that the disclosed

invention is agnostic as to *how* the content identification results are generated—it simply takes those results (whether based on audio and/or video data sample matching) and deduces additional information about audience member consumption habits (e.g., play-skipping) using additional data. A POSITA would thus understand that, in addition to the "BACKGROUND," the "SUMMARY" demonstrates that the inventors possessed the invention as applied to video data samples. *Id.* ¶90.

4. The "DETAILED DESCRIPTION"

The novel features disclosed—including those disclosed in the detailed description—rely on content identification results generated using existing content recognition systems. *Id.* ¶91. In reference to the example provided by Figure 1 below, media monitoring devices 101 capture audio data from media players 170 and transmit the data to media measurement system 100 over network 150. Ex. 1011, 580 (¶[0017]), 581 (¶[0019]).

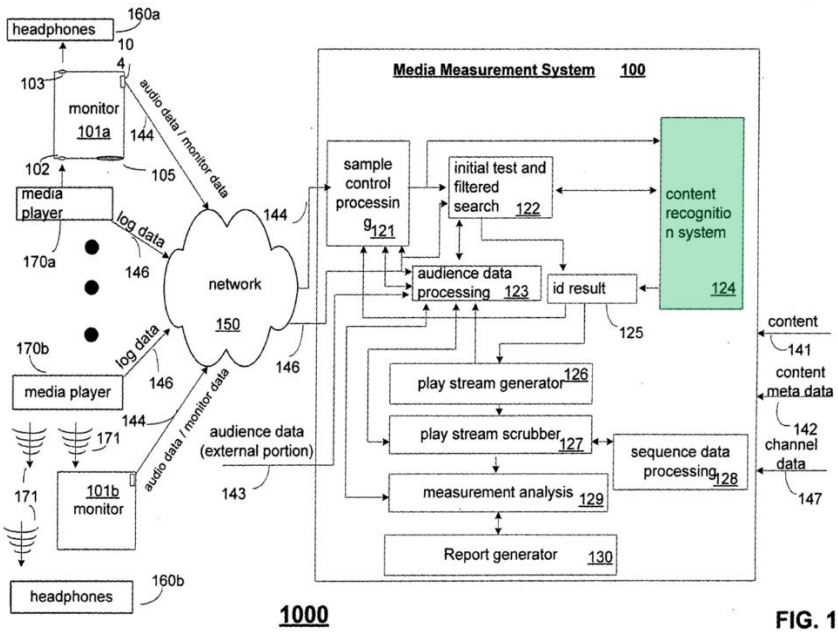


FIG. 1

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

The common written description does not provide details on how the content recognition system identifies the content of the captured samples. Ex. 2001 ¶93. Instead, the written description states "Content recognition system 124 is adapted to implement **one or more known** content recognition methods that identify content by extracting parameters from a **media** signal and applying an algorithm to those parameters to search for a content match. Those skilled in the art will recognize that **many such methods and algorithms exist**." Ex. 1011 at 584 (¶[0025]). A POSITA would understand that a "media signal" includes both audio and video signals and

thus that the specification teaches that content recognition systems that extract parameters from a media signal (i.e., an audio *or* video signal) to identify content were well-known in the art and could be used with the disclosed invention. Ex. 2001 ¶93. The written description then cites numerous examples of prior art references that teach audio- and/or video-sample-based content recognition and discusses various considerations in selecting a known algorithm. Ex. 1011, 584-85 (¶[0025]). (discussing for example Exhibits 2024, 2026, which are discussed above in Section III.B.2). It is the content identification results output by these well-known, existing content recognition systems that the disclosed invention builds upon. Ex. 2001 ¶94.

In one embodiment shown in Figure 4 (annotated below), a "Raw Play Stream" 410 of content identification results (green) is generated, which includes content identifiers obtained using existing content recognition technology. Ex. 1011, 590 (¶[0040]).

410 Raw Play Stream

row	log time	content ID
1	t_0	song 51
2	t_0+1	none
3	t_0+2	song51
4	t_0+3	song51
5	t_0+4	movie 81
6	t_0+5	song51
7	t_0+6	song51
8	t_0+7	song72
9	t_0+8	song72
10	t_0+9	none
11	t_0+10	song72
12	t_0+11	song72
13	t_0+12	song72
14	t_0+13	song65
15	t_0+14	song65
16	t_0+15	TVprgm58
17	t_0+16	song65
18	t_0+17	song65
19	t_0+18	song65
20	t_0+19	song65
21	t_0+20	song21
22	t_0+21	song21
23	t_0+22	song21
24	t_0+23	song21
25	t_0+24	song21
26	t_0+25	song21
27	t_0+26	ad49
28	t_0+27	ad49
29	t_0+28	song81
30	t_0+29	song35

ssd

450 Clean Play Stream

row	log time	content ID
1	t_0	song 51
2	t_0+1	song51
3	t_0+2	song51
4	t_0+3	song51
5	t_0+4	song51
6	t_0+5	song51
7	t_0+6	song51
8	t_0+7	song72
9	t_0+8	song72
10	t_0+9	song72
11	t_0+10	song72
12	t_0+11	song72
13	t_0+12	song72
14	t_0+13	song65
15	t_0+14	song65
16	t_0+15	song65
17	t_0+16	song65
18	t_0+17	song65
19	t_0+18	song65
20	t_0+19	song65
21	t_0+20	song21
22	t_0+21	song21
23	t_0+22	song21
24	t_0+23	song21
25	t_0+24	song21
26	t_0+25	song21
27	t_0+26	ad49
28	t_0+27	ad49
29	t_0+28	song35
30	t_0+29	song35

470 Channel Data for "radio 35"

row	start time	end time	content ID
1	C(0)	C(0)+6	song21
2	C(0)+6	C(0)+8	ad49
3	C(0)+8	C(0)+13	song35

480 Channel Data for "album 12"

row	start time	end time	content ID
1	C(0)	C(0)+7	song51
2	C(0)+7	C(0)+13	song72
3	C(0)+13	C(0)+23	song65
4	C(0)+23	C(0)+30	song29

490 Clean Play List

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

ssd = "sample sequence data"

FIG. 4

The process of obtaining a content identifier using existing content recognition technology was well-known and understood as applied to both audio and video data samples, as explained above in Sections III.B.1-2. Ex. 2001 ¶95. The technical innovation provided thus does not lie in applying these existing content recognition technologies to generate a content identification result. Rather, the innovation lies in what happens next. *Id*. For instance, after existing content recognition technologies are used to generate content identification results (green) as part of a "raw play stream," that play stream is then "scrubbed" to create "Clean Play Stream" 450 by correcting matches that are incorrect (red) or supply the identification of missing content (purple) based on, for example, an expected pattern of sample sequence results. Ex. 1011, 590-91 (¶¶[0040]-[0042]); Ex. 2001 ¶96.

This novel feature is agnostic as to whether the content identification results were generated based on an audio or video data sample—in either case the system simply relies on the content identifiers (e.g., the names of songs or TV programs) with no regard as to whether those identifiers were generated based on audio- or video-sample-based content recognition. Ex. 2001 ¶97. To illustrate this point, if the exact same scrubbing process is applied to a video-sample-based content recognition system, the "content ID" of Raw Play Stream 410 would identify video content (e.g., movie81, TVpgrm58) based on video data, rather than identifying audio-only content (e.g., song51) or audio-video content (e.g., TVpgrm58) based on audio data. However, the "scrubbing" process described would be the same, because that process begins with the content identification results (green) in the Raw Play Stream 410—the process is agnostic as to *how* those results are generated. Stated another way, whether audio- or video-sample-based content recognition is applied is incidental to the disclosed invention. The novel feature above applies in the same way—and is just as effective—as applied to content identification results generated based on video rather than audio data. Ex. 2001 ¶97.

Other disclosed novel features demonstrate this. Figure 5, for example, discloses an embodiment where a clean play stream of content identification results (green) is generated. Ex. 1011, 593 (¶[0047]). However, log times ("time at which capture of the analyzed sample began") and content offsets ("time offset from the

beginning of a particular piece of content") are also associated with each result and can be used to deduce play-altering actions of audience members (e.g., skipping forward or backward, or switching content). *Id.* at 593-95 (¶¶[0047]-[0052]).

510

Clean Playstream

row	log time	content ID	content offset
1	t_0	TVprgm 27	(0)
2	t_0+1	TVprgm 27	(0)+1
3	t_0+2	none	n/a
4	t_0+3	none	n/a
5	t_0+4	none	n/a
6	t_0+5	TVprgm 27	(0)+2
7	t_0+6	TVprgm 27	(0)+3
8	t_0+7	TVprgm27	(0)+9
9	t_0+8	TVprgm27	(0)+10
10	t_0+9	TVprgm27	(0)+9
11	t_0+10	TVprgm27	(0)+10
12	t_0+11	none	n/a
13	t_0+12	TVprgrm27	(0)+29
14	t_0+13	TVprgrm27	(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	TVprgrm28	(0)+2

560

Clean Play List

row	start log time	end log time	content ID	content offset
1	t_0	t_0+2	TVprgrm27	(0)
2	t_0+2	t_0+5	none	n/a
3	t_0+5	t_0+7	TVprgrm27	(0)+2
4	t_0+7	t_0+9	TVprgrm27	(0)+9
5	t_0+9	t_0+11	TVprgrm27	(0)+9
6	t_0+11	t_0+12	none	n/a
7	t_0+12	t_0+14	TVprgrm27	(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	TVprgrm28	(0)

570

Channel Data for TV C4

row	start time	end time	content ID
1	C(0)	C(0)+31	TVprgrm27
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	TVprgrm28

590

Clean Play List with Deduced Actions

row	log time	end log time	content ID and <deduced action>	content offset
1	t_0	t_0+2	TVprgrm27	(0)
2	t_0+2	t_0+5	<pause for 3 units>	n/a
3	t_0+5	t_0+7	TVprgrm27	(0)+2
4	t_0+7	t_0+7	<skip frwrd 5 units>	n/a
5	t_0+7	t_0+9	TVprgrm27	(0)+9
6	t_0+9	t_0+9	<skip bckwrd 2 units>	n/a
7	t_0+9	t_0+11	TVprgrm27	(0)+9
8	t_0+11	t_0+12	<fast frwrd 18 units>	n/a
9	t_0+12	t_0+14	TVprgrm27	(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	TVprgrm28	(0)

FIG. 5

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

technologies. Ex. 2010 ¶98. This methodology applies in the same way—with the same effectiveness—to content identification results generated based on video data samples. *Id.* ¶99. Indeed, just like the audio samples may have corresponding log times ("time at which capture of the analyzed sample began") and content offsets ("time offset from the beginning of a particular piece of content"), video samples may likewise have them. *Id.*

In another disclosed novel feature, the written description teaches in reference to Figure 4 (annotated below) that the channel an audience member is consuming content on can be identified using content identification results. Ex. 1011, 591 (¶[0043]). In this embodiment, the "Clean Play Stream" (element 450) is compared to known "Channel Data for 'radio 35'" (element 470) and known "Channel Data for 'album 12'" (element 480), which indicates that the rows 1-20 are associated with "album 12" (yellow) and rows 21-30 are associated with "radio35" (red). Ex. 2001 ¶100.

410 Raw Play Stream

row	log time	content ID
1	t_0	song 51
2	t_0+1	none
3	t_0+2	song51
4	t_0+3	song51
5	t_0+4	movie 81
6	t_0+5	song51
7	t_0+6	song51
8	t_0+7	song72
9	t_0+8	song72
10	t_0+9	none
11	t_0+10	song72
12	t_0+11	song72
13	t_0+12	song72
14	t_0+13	song65
15	t_0+14	song65
16	t_0+15	TVprgm58
17	t_0+16	song65
18	t_0+17	song65
19	t_0+18	song65
20	t_0+19	song65
21	t_0+20	song21
22	t_0+21	song21
23	t_0+22	song21
24	t_0+23	song21
25	t_0+24	song21
26	t_0+25	song21
27	t_0+26	ad49
28	t_0+27	ad49
29	t_0+28	song81
30	t_0+29	song35

ssd

450 Clean Play Stream

row	log time	content ID
1	t_0	song 51
2	t_0+1	song51
3	t_0+2	song51
4	t_0+3	song51
5	t_0+4	song51
6	t_0+5	song51
7	t_0+6	song51
8	t_0+7	song72
9	t_0+8	song72
10	t_0+9	song72
11	t_0+10	song72
12	t_0+11	song72
13	t_0+12	song72
14	t_0+13	song65
15	t_0+14	song65
16	t_0+15	song65
17	t_0+16	song65
18	t_0+17	song65
19	t_0+18	song65
20	t_0+19	song65
21	t_0+20	song21
22	t_0+21	song21
23	t_0+22	song21
24	t_0+23	song21
25	t_0+24	song21
26	t_0+25	song21
27	t_0+26	ad49
28	t_0+27	ad49
29	t_0+28	song35
30	t_0+29	song35

470 Channel Data for "radio 35"

row	start time	end time	content ID
1	C(0)	C(0)+6	song21
2	C(0)+6	C(0)+8	ad49
3	C(0)+8	C(0)+13	song35

480 Channel Data for "album 12"

row	start time	end time	content ID
1	C(0)	C(0)+7	song51
2	C(0)+7	C(0)+13	song72
3	C(0)+13	C(0)+23	song65
4	C(0)+23	C(0)+30	song29

490 Clean Play List

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

ssd = "sample sequence data"

FIG. 4

As with the other discussed novel features, this feature is agnostic as to whether the content identification results obtained are audio- or video-sample-based. *Id.* The only distinction when applying this feature to content identification results generated by video-sample-based content recognition systems is that the "content ID" of Raw Play Stream 410 would identify video content (e.g., movie81, TVpgrm58) based on video data, rather than identifying audio-only content (e.g., song51) or audio-video content (e.g., TVpgrm58) based on audio data, and the "channel" data would be associated with video content (e.g., a television channel or a movie) rather than audio-only content (e.g., radio35). *Id.*

As demonstrated above, each of these disclosed features relies on existing

content recognition systems that were well-known and understood in the art and that are incidental to the novelty of the invention. *Id.* ¶101. The "BACKGROUND" and "SUMMARY" plainly teach that the invention is applicable to audio- and video-sample-based content recognition systems, and a POSITA would understand that the invention applies equally to, and provides the same benefit for, both kinds of systems. *Id.* In fact, the very first statement in the "DETAILED DESCRIPTION" is:

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

Ex. 1011, 580 (¶[0015]). A POSITA would understand that the teachings of the "DETAILED DESCRIPTION" apply to both audio-sample-based and video-sample-based media measurement systems, and that the applicability of the teachings of the Figures 1-6 embodiment to video-sample-based content recognition systems is "readily apparent" and nothing more than the application of the generic principles of that embodiment to then-existing video data sampling systems referenced in the

"BACKGROUND" and "SUMMARY." Ex. 2001 ¶101.

5. The Challenged Claims

The Challenged Claims are fully supported by the common written description. Independent claim 1 is exemplary and is reproduced below.

1. A media measurement method comprising:

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

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

The only basis for Petitioner's contention that the common written description does not support the challenged claims is its inclusion of the word "video data samples" rather than "audio data samples." Pet., 15-16. However, this merely dictates the type of media sample that the "content identification results" are generated based upon—which was well within the knowledge of a POSITA as evidenced by the prior art. Indeed, the written description expressly teaches that: (1) the invention applies "existing" content recognition systems based on "audio or video data sample[s]" and such systems were well-known and understood (*see* Section III.B.1-2); (2) the invention applies to "media sample[s]," which refers to both audio and video data samples (*see* Section III.B.3); (3) the "DETAILED DESCRIPTION" teaches that algorithms for matching media signals (i.e., both audio and video signals) were well-known in the art and that applying the principles of operation of the audio embodiments to other embodiments would be "readily apparent" and (4) the novel feature disclosed in the detailed embodiments is entirely agnostic as to *how* the content identification results are generated (*See* Section III.B.4). A POSITA would thus not need or expect the inventors to disclose *how* content identification results are generated based on audio- or video-data samples to demonstrate possession of the invention, because that was within the knowledge of a POSITA based on existing technology. Ex. 2001 ¶105.

The first step of "using one or more computers to generate a play stream of

content identification results corresponding to a sequence of submitted data samples, the submitted data samples corresponding to **video** data samples" is thus clearly supported by the written description, as this involves nothing more than the application of existing video-sample-based content recognition technologies that are expressly discussed in the common written description to generate the "content identification results" upon which all other claim elements apply. Ex. 2001 ¶106. These other claim elements are agnostic to whether the content identification results are generated based on audio or video data, and thus they are clearly supported by the "DETAILED DESCRIPTION," which provides specific examples of applying these same steps based on content identification results. *Id.* Indeed, Patent Owner demonstrates that deducing play-altering actions using content offsets and log times as taught with respect to audio data samples operates in the same way, and provides the same benefit, as applied to content identification results generated based on video data samples. *Supra* Section III.B.4.

C. Petitioner's Contrary Position Lacks Merit and Evidentiary Support.

The common written description plainly teaches that the invention applies "existing" and "robust" content recognition technologies based on "audio or video data sample[s]," that the invention applies to "media sample[s]" (which includes both audio and video data samples), and that the improvement provided lies in the use of additional data beyond the results provided by those well-known and

understood technologies. *Supra* Sections III.B. The inventors were thus not required to disclose a detailed embodiment showing how the invention applies when relying on existing video-sample-based content recognition systems to demonstrate that they possessed the claimed invention. Ex. 2001 ¶107. Petitioner cannot legitimately dispute this, so it mischaracterizes these plain teachings and falsely portrays the invention's application to video-sample-based content recognition systems as somehow different or more complex than its application of audio-sample-based systems.

1. Petitioner's Attempt to Rewrite the "BACKGROUND" Should be Rejected

The common written description's express teaching that it applies "existing" and "robust" content recognition technologies, including those based on "audio or video data sample[s]," is sufficient in and of itself to convey that the inventors possessed the invention as applied to video data samples. Ex. 2001 ¶108. To downplay the significance of this evidence, Petitioner and its expert offer the Board the following unreasonable interpretation:

> 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."

Pet., 26-27 (citing Ex. 1002 ¶75). In other words, Petitioner and its expert argue that the Board should simply ignore what the passage expressly states and instead interpret the discussion of content identification based on "video data sample" as somehow referring to only an "audio data sample."

As an initial matter, Petitioner provides no evidence to support this assertion, other than the *verbatim* parroting of its expert. *Compare* Pet., 26-27 *with* Ex. 1002 ¶75. Indeed, neither Petitioner nor its expert provide any evidence where content recognition based on audio data is described as being from or related to "video data" or a "video data sample." In contrast, Patent Owner provides the Board with a multitude of objective references demonstrating that both audio- and video-sample-based content recognition systems were well-known and understood in the art at the time of the invention. *See supra* Section III.B.1-2 (discussing Exhibits 2005-2036). This includes Exhibit 2026, which is expressly discussed in the common written description. Indeed, it was well-understood that content recognition based on "audio data samples" involved analyzing audio data extracted from an audio signal to extract fingerprints (e.g., frequency patterns, tempo, and amplitude) and "video data samples" involved analyzing video data extracted from a video signal to extract

fingerprints (e.g., pixel patterns, color patterns, luminance, motion). Ex. 2001 ¶109. Petitioner offers no evidence to the contrary because no such evidence exists. *Id.* ¶¶109-110. In fact, Samsung's own patents demonstrate that it understands "video data" to refer to data sampled from a video signal, which is different than "audio data." *See e.g.*, Ex. 2036, Abstract, 3:47-49 (discussing "[u]ser media data, i.e., audio or video data that is saved inside an MPEG file"); Ex. 2037, 4:15-27 (discussing an apparatus for "coding video data" that includes "input means for receiving two or more **frames of video data representative of image sequences**").

Even setting aside Petitioner's lack of evidence, Petitioner's interpretation is illogical on its face for at least four reasons. First, the challenged independent claims include the same "video data sample" language referenced in the "BACKGROUND." Ex. 1001, cl. 1, 9 ("the submitted data samples corresponding to ***video data samples*** captured at a monitoring device"). If, as Petitioner proposes, the term "video data sample" in the "BACKGROUND" could somehow refer to the audio portion of a video data sample, that would necessarily mean that "video data samples" in the claims would have that same meaning. If that were the case, then Petitioner cannot plausibly contend that the claims lack written description support in the Prior Non-Provisional Applications, because Petitioner admits that the "DETAILED DESCRIPTION" provides examples of the invention as applied to content recognition systems that identify content based on audio associated with

video content (e.g., audio associated with a movie). Petitioner cannot have its cake and eat it too—if it is going to take the position that the reference to "robust" and "existing" content identification systems that rely on "video data sample[s]" somehow refers only to analyzing audio data associated with video content, then Petitioner necessarily forecloses its argument that the claims lack written description support in the Prior Non-Provisional Applications. Ex. 2001 ¶111.

Second, it is illogical that the "BACKGROUND" would refer to a "video data sample" if it was somehow only intending to refer to analyzing audio, because the same sentence *already refers to an* "audio … data sample." In other words, there would be no need to refer to a "video data sample" at all if that was intended to refer to analyzing only audio, because the reference to "audio … data sample" already conveys that those existing content recognition systems analyzed audio data of any type (including audio associated with video content, such as a movie). *Id.* ¶112.

Third, Petitioner attempts to justify its illogical interpretation of "video data sample" by arguing that allegedly "audio-sampling-based content recognition is the only type of content recognition described in the '330 application" and that "the '330 application does not describe any content recognition using a 'video data sample.'" Pet., 26-27. However, this is circular logic that falsely presumes that the term "video data sample" in the "BACKGROUND" does not actually refer to a video data sample and that the reference to "media sample[s]" in the "SUMMARY" does not

encompass both audio and video data samples. Indeed, the reference to "existing" and "robust" content recognition systems that analyze "audio or video data sample[s]" **does** in fact disclose a content recognition system that is video-sampling based, completely undermining Petitioner's position that "audio-sampling-based content recognition is the only type of content recognition" disclosed. Ex. 2001 ¶113. Petitioner also fails to explain why a POSITA would understand the "DETAILED DESCRIPTION" to limit what is plainly taught by the "BACKGROUND" and "SUMMARY," particularly given that the very first passage in the "DETAILED DESCRIPTION" explains that its disclosed embodiments are not intended to be limiting. *Id.* ¶114 (citing Ex. 1011, 580 (¶[0015]).

Fourth, the "DETAILED DESCRIPTION" describes the analysis of audio data associated with video content (e.g., the audio associated with a movie) but never refers to or suggests that this is content recognition based on "video data" or "video data samples." If Petitioner's position had any merit, a POSITA would expect the "DETAILED DESCRIPTION" to use the term "video data sample" in a manner consistent with that interpretation. *Id.* ¶115. It does not. The reason the written description never refers to the analysis of audio associated with a movie as involving a "video data sample" is because such analysis relies on "audio data samples"—"video data samples" are those taken from a portion of a video signal and used to

create video data fingerprints (e.g., luminance-based fingerprints). *Id*.

Try as it may, Petitioner is not permitted to rewrite what the written description plainly states to avoid the inescapable conclusion that the "BACKGROUND" expressly states the invention applies to both audio- and video-sample-based content recognition systems.

2. Petitioner Improperly Ignores the Teachings of the "SUMMARY"

The "SUMMARY" conveys to a POSITA that the invention applies to content recognition technology based on video data samples, particularly when it states "[s]ome embodiments of the present invention provide a media measurement system and method that enhances recognition (e.g. in terms of accuracy or efficiency) of the content of a **media** sample by analyzing information exogenous to the sample." Ex. 2001 ¶116 (citing Ex. 1011, 578 (¶ [0006])). Indeed, "media" refers to both audio and video samples and the immediately preceding passage from the "BACKGROUND" expressly teaches that the invention applies to content recognition systems based on audio and video data samples. *Supra* Section III.B.3.

Neither Petitioner nor its expert addresses this statement from the "SUMMARY," despite it clearly teaching that the invention applies to "media sample[s]" (i.e., audio and video samples), not merely "audio samples." *See generally* Pet., 24-29. However, the written description analysis requires consideration of all disclosures within the four corners of the specification—not

simply those that Petitioner believes support its invalidity theory. This particular disclosure clearly conveys that the invention was not intended to be limited to audio samples. If the inventors had not contemplated applying the invention to the "robust" and "existing" video-sample-based content recognition technologies, the statement from the "SUMMARY" would refer to "audio samples," not "media sample[s]." Ex. 2001 ¶117. The "SUMMARY" thus conveys possession of a video-data embodiment. This conveyance is even more apparent when viewed in the context of the immediately preceding passage from the "BACKGROUND" teaching that the invention applies to "existing" robust content recognition technologies that analyze "audio or video data sample[s]." *Id.*

MySpace, Inc. v. Graphon Corp., 672 F.3d 1250 (Fed. Cir. 2012) is instructive. The issue there was whether the claim term "database" should be interpreted to mean "a collection of data with a given structure that can be stored and retrieved," which would include both file (hierarchical) and relational systems, or whether "database" should be limited to only relational systems. *Id.* at 1255. In determining that "database" should be accorded its broader meaning, the Court explained:

> The "Summary of the Invention" section of the written description for all four of the patents-in-suit broadly describes the database as being a collection of user entries that "are automatically collected, classified and stored . . . in a searchable and retrievable form." '538 Patent col. 3

> ll. 2-3. All types of databases—relational, hierarchical, or otherwise—are searchable and retrievable. Thus, nothing in this language serves to limit the invention to a relational database.
>
> Further, the "Background of the Invention" section discusses the limitations of prior art systems in placing 'databases of various kinds' directly on the web, and suggests that the disclosed invention overcomes those disadvantages. *See, e.g.*, '538 Patent col. 1 l. 37 and col. 2 ll. 57-58. Taken together, this language indicates that the general term "database" in the claims can encompass any of the "various kinds" of databases known in the art at that time.

Id. at 1256. While the issue in *MySpace* related to claim construction, the Federal Circuit's logic wholly applies to the facts here. As in *MySpace*, where the "Background" discussed known "databases of various kinds" and taught disadvantages of those databases that the invention was to overcome, the "BACKGROUND" here clearly teaches that robust audio- and video-sample-based content recognition systems existed, and that the invention improves upon those systems by using data exogenous to a particular sample being analyzed. Likewise, just like in *MySpace*, where the "Summary" broadly described databases, the "SUMMARY" here "broadly describes" the invention as not simply applying to "audio samples," but rather "media sample[s]" which encompass both audio and video data samples. Taken together, the "BACKGROUND and "SUMMARY" here clearly convey that the inventors possessed the invention as applied to both audio-

and video-sample-based content recognition systems known at the time, just as those same sections in *MySpace* conveyed that "database" encompasses known database structures in the art at that time.[6]

3. Petitioner's Reliance on the Number of Times "Audio" and "Video" Are Used Is Misplaced.

Although Petitioner does not explicitly argue as much, Petitioner suggests that the number of times the written description refers to "audio" as compared to "video" somehow evidences that the written description does not support the challenged claims. Pet., 26. However, Federal Circuit precedent forecloses such logic. *See ScriptPro LLC v. Innovation Assocs., Inc.*, 833 F.3d 1336, 1341 (Fed. Cir. 2016) (reversing district court invalidity ruling based on lack of written description, explaining that "a specification's focus on one particular embodiment or purpose cannot limit the described invention where that specification expressly contemplates other embodiments or purposes"); *see also CenTrak, Inc. v. Sonitor Techs., Inc.*, 915

[6] Neither Petitioner nor its expert argue that the common written description fails to enable the invention as applied to video-sample-based content recognition systems. Given that the written description expressly teaches applying the invention to "existing" audio- and video-sample-based content recognition systems and that Petitioner does not allege a lack of enablement, Petitioner cannot demonstrate a lack of written description.

F.3d 1360, 1366-68 (Fed. Cir. 2019) (finding that "the district court leaned heavily on the fact that the specification devoted relatively less attention to the ultrasonic embodiment compared to the infrared embodiment" and determining that the district court erred "by not sufficiently crediting testimony from CenTrak's expert that the differences between IR and ultrasound, when used to transmit small amounts of data over short distances, ***are incidental to carrying out the claimed invention***.").

Indeed, the written description teaches that the disclosed invention relies on existing audio- or video-sample-based content recognition technologies, and the fact that the detailed description focuses on audio-sample-based content recognition is irrelevant to the fact that it covers both. Ex. 2001 ¶118. This is particularly true given that, as explained in Section III.B and further below, the invention is agnostic as to whether the content identification results are generated based on audio or video data samples. This is "incidental to carrying out the claimed invention" because the novelty relates to what the invention does *with* those results rather than merely how, or from what media data-type, those results are generated. *Id.*

4. Petitioner Mischaracterizes the Nature of the Invention and the State of the Art

The "BACKGROUND" and "SUMMARY" clearly convey possession of the invention as applied to both audio and video data samples using existing content recognition technologies, both of which were well-known and understood as evidenced by countless pieces of evidence cited herein. Petitioner cannot

legitimately dispute this, so it resorts to misdirection.

Petitioner argues that a POSITA would not have recognized the inventors to have possessed the invention as applied to video data samples because there are so-called "obstacles," "challenges," and "substantive differences" between audio- and video-sample-based content recognition and the "DETAILED DESCRIPTION" does not provide a detailed example of the disclosed system as applied to "video data samples." Pet., 20-21, 28. But Petitioner provides no legitimate support for this argument, and there is no question that audio- and video-sample-based content recognitions were well-known and understood. Even beyond the common written description explicitly saying as much, Patent Owner has provided the Board with dozens of objective examples from the prior art further proving this point. *See supra* III.B.2.

Given this, a POSITA would not need an explanation or detail as to *how* to generate content identification results based on video rather than audio samples and would not require or demand that the inventors explain how video data sampling works to determine that they possessed the invention as applied to video-sample-based content recognition systems. Ex. 2001 ¶¶119-120; *See Capon*, 418 F.3d at 1357 (determining that the "Board erred in refusing to consider the state of the scientific knowledge" and by requiring "a re-description of what was already known" in the art, particularly where what was known was not the "novel" aspect of

the invention). Indeed, the common written description does not even explain how *audio* data samples are used to generate content identification results—it simply refers to prior art references showing that both audio- and video-fingerprint matching techniques and algorithms were known in the art. Ex. 1011, 584-585 (¶[0025]). The common written description does not disclose how to perform audio- or video-sample-based content recognition because it is within the knowledge of a POSITA and the written description properly presumes as much. Ex. 2001 ¶121. As such, the express statements that the disclosed system "appl[ies] **existing** content recognition technologies" based on an "audio or **video** data sample" (Ex. 1011, 578 (¶[0005])) and "***enhance*s** recognition (e.g., in terms of accuracy or efficiency) of the content of a ***media*** sample by analyzing information exogenous to the sample" (Ex. 1011, 578 (¶[0006]), clearly conveys that the inventors possessed the invention as applied to both audio and video data samples. *Id.* ¶121.[7]

Petitioner also argues that the Board should not consider the fact that the common written description discusses numerous references (Ex. 1011, 584-85 (¶[0025])) that evidence that both audio- and video-sample-based content

[7] Petitioner's suggestion that video-sample-based content recognition was more complex and difficult than audio-sample-based content recognition is also misplaced. Ex. 2001 ¶¶122-125.

recognition systems were well-known in the art. Pet., 33-35. This argument mis-frames the issue before the Board by focusing on whether those references are "incorporated by reference." *Id.* at 34. As demonstrated, video-sample-based content recognition was well-known and understood and only incidental to the invention. A POSITA thus did not need any guidance or express teachings in the specification to understand how to generate an underlying content identification result based on a video data sample using these known technologies. Ex. 2001 ¶126. Petitioner does not allege a lack of enablement—it instead alleges that a POSITA would not understand the inventors to have possessed the claimed invention. The fact that the common written description states that "those skilled in the art will recognize that ***many such methods and algorithms exist***" and discusses prior art examples of video-sample-based systems (e.g., WO 2004/044820 (Ex. 2026)) may thus be considered by the Board as evidence that those technologies were in fact well-known and understood in the art, which further demonstrates that a POSITA would have understood the inventors to possess the invention as applied to such systems without a detailed description describing the same process from Figures 1-6 as applied to content identification results generated from video-samples.

Petitioner further argues that "a POSITA would not have understood that the '330 application's disclosures of embodiments relying on audio data would have been interchangeable with embodiments that instead use video data." Pet., 28. This

too is false. Patent Owner demonstrates that the disclosed invention is agnostic as to whether the content identification results are generated based on audio or video data samples. *Supra* Section III.B.4-5. The novel features disclosed apply in the same way and provide the same benefit, regardless of whether those identification results are audio- or video-sample-based. Ex. 2001 ¶127. In fact, numerous prior art references demonstrate that the applicability of inventions based on audio-sample based content recognition often applies equally to video-sample-based content recognition. *Supra* Section III.B.2.

Despite this, Petitioner disingenuously argues that a POSITA would not understand how to apply the disclosures of the common written description to video data samples because "capturing audio would operate in different environments and be subject to different obstructions, challenges, and limitations to a system that relies on capturing video." Pet., 28. Yet, capturing and matching audio and video samples was well-known and understood in the art, and so these "different environments" and "different obstructions, challenges and limitations" were *already addressed by those existing systems*. Ex. 2001 ¶128. Indeed, the "BACKGROUND" and "SUMMARY" make it clear that the inventors are not inventing a novel way of audio- or video-based fingerprint matching, but are instead applying those existing technologies to generate content identification results that are then used as the basis for the invention's novel combination of features.

The issue is not whether video data sample content recognition is different than audio data sample content recognition—there are obvious and well understood differences between these two types of content recognition systems. The issue is whether a POSITA would have understood that the inventors possessed the disclosed invention as applied to content identification results generated using **well-known and well-understood** audio- or video-sample-based content recognition systems, which already account for the purported differences raised by Petitioner. Even assuming "different obstructions, challenges and limitations" between audio- and video-sample-based content recognition systems, they have no effect on how the invention applies to content identification results generated by either type of system. Ex. 2001 ¶129.

Petitioner further suggests to the Board that video-sample-based content recognition was not well-known or understood. Pet., 20-21 (arguing that "government-funded efforts at the time demonstrated around 10-30% accuracy rate for video content recognition) (citing Ex. 1030-1032). The only evidence cited by Petitioner to support this statement is a TRECVID-sponsored *object recognition* competition. Ex. 2001 ¶¶130-131 (explaining Exhibits 1030 to 1032). Object recognition is fundamentally different than content recognition based on video sample matching, which is simply generating a video sample from a video signal and comparing it to a database of reference video samples associated with known

content. *Id.* ¶132. Regardless, Patent Owner has provided the Board with dozens of references demonstrating that video-sample-based content recognition was well-known and understood, far outweighing Petitioner's unexplained reliance on TRECVID.

Petitioner also complains that a POSITA would not have understood how the invention could be implemented based on a video-sample-based content recognition system because the common written description is purportedly silent as to "how the disclosed audio-based embodiment would have been implemented in the context of video in light of any unique challenges posed by video data." Pet., 21. Yet, Petitioner never identifies or explains any of these so-called "unique challenges" in implementing the invention based on video-sample-based content identification results. Ex. 2001 ¶133. This is because Petitioner cannot legitimately argue that a POSITA would not understand how the disclosed invention applies to content identification results generated based on video data samples—the invention applies in the same way as it does to results generated based on audio data samples. *Id.* ¶134.

Petitioner also argues that a POSITA would not have understood the inventors to have possessed the invention as applied to video-data samples because "multiple disclosures in the '330 application describe peculiarities related to audio content like 'mut[ing] a broadcast,' 'receiv[ing] a wireless or non-wireless audio signal,' 'captur[ing] audio energy,' and 'content types (e.g. speaking audio versus music

audio),' that have no clear analog in the video context." Pet., 28. As an initial matter, the fact that the common written description provides additional detail to account for issues specific to the "audio context" says nothing about whether the common written description **also** sufficiently demonstrates how the disclosed invention applies in the "video context." Indeed, Petitioner cherry-picks these so called "peculiarities" from the "DETAILED DESCRIPTION," while ignoring the teachings that a POSITA would clearly understand and recognize to be applicable to video-sample-based content recognition, such as generating content identification results and using log time and content offset data associated with those results to deduce play altering actions. Ex. 2001 ¶¶135-136.

Moreover, the claim language does not require these so-called "peculiarities" to be accounted for. Rather, the claims broadly encompass the concept of generating content identification results and then applying novel features that would be understood to be applicable to audio-sample-based and video-sample-based identification results alike (as evidenced by the nearly identical audio claims of Steuer791). The written description need only provide sufficient disclosure to demonstrate that the inventors possessed the invention commensurate with the scope of the claims. The common written description clearly does that here—it teaches that the invention applies existing "audio or video data sample" content recognition technologies and provides a detailed embodiment that a POSITA would readily

understand is equally applicable to content identification results based on video-data samples. The claims do not require, as Petitioner falsely suggests, that the so-called video "analogs" to these "peculiarities" be addressed or accounted for.

5. Petitioner's Arguments Regarding Alleged "New Matter" Are Wrong and Irrelevant.

New matter was not added to the specification of the '896 Patent (the '204 application) as alleged by Petitioner. Pet., 36-38. The common written description teaches that the disclosed invention applies "existing" audio- and video-sample based content recognition technologies to generate content identification results and, further, that the invention applies to "media sample[s]" (i.e., audio and video samples). A POSITA would thus understand, based on the common written description, that the disclosures relating to the detailed embodiment of Figures 1-6 are applicable to content identification results generated based on video data samples. Ex. 2001 ¶¶137-138. The passage first included in the '896 Patent thus does nothing more than elaborate or clarify what is already conveyed by the common written description—while that passage was "added" it is clearly not "new" matter. *See Schering Corp. v. Amgen Inc.*, 222 F.3d 1347, 1352 (Fed. Cir. 2000) (explaining that an application does not add new matter when "its original application supports the amended matter"). Indeed, the written description requirement is met "when 'the essence of the original disclosure' conveys the necessary information—'regardless of *how* it' conveys such information.'" *Inphi Corp.*, 805 F.3d at 1354 (emphasis in

original).

In addition to being wrong, Petitioner's new matter argument is also irrelevant. Even if the passage could be found to include some new matter, that would be immaterial to whether the common written description conveys possession of the claimed invention. Ex. 2001 ¶139. The additional passage cannot somehow alter or undo the fact that the common written description—in and of itself—conveys possession of the claimed invention as discussed in Section III.B. Petitioner cites no authority to support its contrary position that the added passage can be used to presume a lack of written description support or undermine the express teachings of the common written description. Petitioner also ignores the fact that patentees are entitled to, and routinely do, amend their specification to clarify disclosures without adding new matter. *See e.g.*, *TurboCare Div. of Demag Delaval Turbomachinery Corp. v. GE*, 264 F.3d 1111, 1118 (Fed. Cir. 2001) (explaining that an amendment "merely to clarify" what was originally disclosed is not new matter); *Schering Corp.*, 222 F.3d at 1352 (reversing district court determination that amendment to specification was "new matter," finding instead that it merely "clarifies" what was already conveyed by the written description). According to Petitioner's flawed logic, all clarifying amendments would necessarily be "new matter" because the act of amending would itself evidence the addition of new matter.

The added passage is also irrelevant because a determination of whether the

Prior Non-Provisional Applications provide adequate written description support for the challenged claims is limited to reviewing the four corners of *those* specifications for support—not the specification of the ’896 Patent. Consequently, because the common written description conveys possession of the invention, Petitioner has failed to demonstrate that the challenged claims are not entitled to the May 26, 2005 filing date because of a lack of written description support.

6. The Testimony of Petitioner’s Expert on the Issue of Written Description Should Be Given No Weight.

Sufficiency of written description is a question of fact. *General Hosp. Corp.*, 888 F.3d at 1371. The Board must thus weigh the credibility of the evidence before it, including the credibility of the expert testimony presented to the Board. The testimony of Petitioner’s proffered expert should be given no weight with respect to the alleged lack of written description for several reasons.

Petitioner’s expert offers testimony on key issues that is so implausible on its face that it demonstrates a clear and palpable bias. For example, he testifies that when the common written description states “[a]t the same time, the media measurement context provides opportunities to utilize data exogenous to a particular **audio or <u>video</u> data sample**,” it is “unclear” if that is actually referring only to audio data samples. Ex. 1002 ¶75. That Petitioner’s witness would offer testimony that is so plainly contradicted by the express text of the written description evidences a bias toward finding a lack of written description regardless of the evidence. Indeed, he

provides no evidence to the Board to support this implausible opinion, such as any evidence that *any* person has ever referred to audio-sample-based content recognition as being based on "video data" or a "video data sample."

Petitioner's expert also ignores portions of the specification that fundamentally undermine his opinions. For example, neither he nor Petitioner addresses the explicit teaching in the "SUMMARY" that the invention applies to "media sample[s]," which a POSITA would readily understand to refer to both audio and video samples. Ex. 2001 ¶¶116-117. This is particularly egregious given that the statement *immediately* follows the "BACKGROUND" statement that the invention applies existing "audio or video data sample"-based content recognition technologies. His failure to account for this critical teaching renders his entire analysis of the common written description unreliable.

Additionally, his testimony on written description merely parrots the arguments from the Petition. For example, every one of Petitioner's expert opinions that the common written description fails to support the claims is a copy-and-paste of the Petition, except for arbitrary insertions of statements like "in my opinion" and citation quotes (rather than mere citations). *Compare* Pet., 24-29 *with* Ex. 1002 ¶¶73-79. Petitioner's expert does not meaningfully expound or elaborate on *any* of the arguments set forth in the petition, including the vague unexplained statements that the disclosed invention would somehow operate differently when applied to audio

data samples versus video data samples. As explained in Section III.C.4 above, this argument has no merit; but even if it did, Petitioner provides no meaningful testimony or explanation to adequately support its argument. *Xerox Corp. v. Bytemark, Inc.*, IPR2022-00624, Paper 9 (Aug. 24, 2022) (*precedential*) (denying institution, holding expert declaration is entitled to little weight when it contains an exact restatement of the petition's unsupported, conclusory assertions without any additional supporting evidence or reasoning).

IV. Institution Should Be Denied Because the Petition Fails to Demonstrate Obviousness

Denial of institution is also warranted—with or without resolving the written description issue above—because Petitioner fails to show a reasonable likelihood that a POSITA would have found it obvious to modify Steuer750/Steuer791 to apply to the video-sample-based content recognition system of Neumeier.[8] To demonstrate obviousness, Petitioner must demonstrate that the asserted art is analogous, provide an articulated rationale for combining the art as proposed, and show a reasonable likelihood of success. *Corephotonics, Ltd. v. Apple Inc.*, 84 F.4th 990, 1003-1004

[8] Patent Owner assumes, for the sake of these arguments only, that Steuer750 and Steuer791 do not convey that the disclosed invention applies video data samples, as alleged by Petitioner.

(Fed. Cir. 2023). Petitioner fails to provide an articulated rationale or demonstrate a reasonable likelihood of success.

Petitioner begins its obviousness analysis by arguing "it would have been obvious to modify Steuer750's and Steuer791's disclosed media measurement methods such that video data samples are used instead of audio based on the teachings of Neumeier," followed by a discussion of the purported similarities between the references. Pet., 48-49. Petitioner then concludes: "Thus, Neumeier discloses content identification systems and methods in the same technical field as Steuer750 and Steuer791, including content identification based on captured data samples. Accordingly, a POSITA would have **had reason to consider the teachings of Neumeier** when implementing Steuer750 or Steuer791." *Id.* at 50. This amounts to nothing more than a statement that the asserted art is analogous—it is not an articulated rationale to combine the references.

Petitioner proceeds to address further purported similarities between Neumeier and Steuer750/Steuer791, ultimately concluding:

> In view of Neumeier's teachings, a POSITA would have been **motivated** to consider and draw from such teachings when implementing Steuer750's or Steuer791'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, Steuer791, and Neumeier 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.

Pet., 50-52. However, this is not an articulated rationale or motivation for combining the references as proposed—it is nothing more than a discussion of the similarities of the references followed by the conclusory, generic statement that a POSITA would have been motivated to combine them ***simply because they are allegedly similar***. Indeed, nowhere in this conclusion, or the lengthy discussion preceding it, does Petitioner actually explain why a POSITA would have been motivated to apply Steuer750/Steuer791 to video data samples. The mere fact that alleged prior art provides a feature that *could be* included as a purported "alternative" does not constitute a rationale or motivation to do so. And while Petitioner states there was a motivation "[i]n view of Neumeier's teachings," Petitioner does not identify any motivation or teaching in Neumeier for applying its video data teachings to the (allegedly exclusively) audio-sample-based system of Steuer750/Steuer791. Petitioner instead relies solely on the fact that Neumeier and Steuer750/Steuer791 allegedly teach similar concepts as the sole basis for the purported "motivation."

Petitioner next argues:

> **A POSITA would have found it obvious** and would have known how

> to modify Steuer750 and Steuer791 to incorporate Neumeier's teachings as proposed, **because** as discussed above, Neumeier **discloses a way of identifying content based on matching video data samples with video data stored in a database**. (EX1002, ¶ 109) Thus, a POSITA would have recognized that applying Steuer750's or Steuer791's use of audio data to video data **could have been** implemented 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 Neumeier.

Id. at 53. The first sentence of this passage is merely a statement that the combination would allegedly be obvious "because" Neumeier discloses video data sample matching. Again, there is no articulated reason provided for *why* a POSITA would have actually combined the references. The second sentence simply states that a POSITA "could have" applied the teachings of Steuer750/Steuer791 to video data. But the fact that a POSITA allegedly "could have" done so is not a motivation to combine. Indeed, even accepting this statement as true, the reasoning "seems to say no more than that a skilled artisan, once presented with the two references, would have understood that they *could be* combined. And that is not enough: it does not imply a motivation to pick out those two references and combine them to arrive at the claimed invention." *PersonalWeb Techs., LLC v. Apple, Inc.*, 848 F.3d 987, 994 (Fed. Cir. 2017) (emphasis in original) (citing *Belden Inc. v. Berk-Tek LLC*, 805 F.3d

1064, 1073 (Fed. Cir. 2015) ("[O]bviousness concerns whether a skilled artisan not only could have made but *would have been motivated* to make the combinations or modifications of prior art to arrive at the claimed invention.") (emphasis in original)). Petitioner's conclusory statement that the proposed modifications are "obvious" also lacks any supporting rationale—there is simply no articulated explanation as to *why* those modifications are "obvious."

Ultimately, the only arguments offered by Petitioner that even slightly resemble *actual* rationales to combine are statements lifted directly from the Manual of Patent Examining Procedure that the combination was allegedly (1) "obvious to try" and (2) the combination of "known technologies" according to known methods to achieve predictable results. Pet., 53-54.

Turning first to the "obvious to try" motivation, it fails because this rationale requires showing that, at the relevant time, there had been a ***recognized problem or need*** in the art (e.g., a design need or market pressure to solve a problem) and that that there had been ***a finite number of identified, predictable*** potential solutions to the recognized need or problem. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 421 (2007); *see also In re Index Sys.*, 576 F. App'x 976, 980 (Fed. Cir. 2014). Petitioner makes neither of these showings, but instead appears to have simply copied the general summary of the "obvious to try" rationale from the MPEP without actually having examined the actual requirements of that rationale. Instead, Petitioner

generically states the combination was "obvious to try."

With respect to the "known technologies" rationale, Petitioner is required to show that that the teachings of Steuer750/Steuer791 as applied to the video data system of Neumeier would have been recognized by a POSITA to yield an "***improvement***," that a POSITA would have recognized the teachings of Steuer750/Steuer791 ***to be applicable*** to video data samples, and that applying the technique of Steuer750/Steuer791 would have been recognized to yield ***predictable*** results. *KSR*, 550 U.S. at 417-18. The entirety of Petitioner's argument is the following sentence:

> The above modification would have been a mere combination of known technologies (e.g., the concept of collecting video data samples with Steuer750's/Steuer791's disclosed systems and methods for making content identifications), according to known methods (e.g., sampling pixel data of the video content), to achieve predictable results (e.g., generating content identification results by comparing video data samples with known content).

Pet., 54. Nowhere in this analysis does Petitioner explain ***why*** a POSITA would have recognized that the teachings of Steuer750/Steuer791—which Petitioner alleges does *not* teach or suggest its application to video data samples—would have improved, or even been applicable to, the teachings of Neumeier. Why would a POSITA have recognized Steuer750/Steuer791 to be applicable to video data samples given Petitioner's adamant contention that there were so many differences,

obstacles and challenges as between audio and video data sample recognition? Would a POSITA have even recognized Steuer750/Steuer791 to be a possible improvement for content recognition based on video data samples given these purported obstacles and differences? If so, why? Why would a POSITA have even considered applying Steuer750/Steuer791 to video-sample-based content recognition when it was well-known that video content could be identified based on audio-sample-based content recognition of the audio associated with video content? Petitioner does not address or explain any of this.

To the contrary, Petitioner repeatedly argues that there were difficult challenges and obstacles in applying the disclosed invention of Steuer750/Steuer791 to video data samples. *See e.g.*, Pet., 20-21 (arguing that there are "substantive differences" when "substituting video data samples for audio data samples" in the context of the disclosed invention, including "different challenges and obstacles faced with collection and recognition of audio versus video" and complaining that the provisional does not explain how to implement video to overcome "**unique challenges** posed by video data."); *see also id.* at 28 (representing that "a POSITA would not have understood that the '330 application's disclosures of embodiments relying on audio data would have been interchangeable with embodiments that instead use video data" and that a "POSITA 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."). Petitioner takes these positions to convince the Board that—despite the common written description expressly teaching that it applies to audio or video data samples—a POSITA would not have recognized that the inventors possessed such a system because they did not address all of the purported distinctions in applying the invention to content identification results generated based on video data samples rather than audio data samples.

Petitioner cannot have it both ways. It cannot, on one hand, argue that the common written description lacks support for the challenged claims due to all of these so called "challenges" and "obstacles" and turn around and argue that the invention is obvious without addressing or explaining how *any* of these so-called challenges or obstacles would be addressed or overcome by a POSITA in modifying Steuer750/Steuer791 to apply to the video data system of Neumeier. Petitioner complains, for example, that the common written description lacks support for the claims because "multiple disclosures in the '330 application describe peculiarities related to audio content like 'mut[ing] a broadcast,' 'receiv[ing] a wireless or non-wireless audio signal,' 'captur[ing] audio energy,' and 'content types (e.g. speaking audio versus music audio),' ***that have no clear analog in the video context***." Pet., 28. But, if these "challenges," "obstacles," and "peculiarities" would have been understood as Petitioner alleges (to support its lack of written description argument),

those same challenges, obstacles and peculiarities necessarily exist when attempting to modify Steuer750/Steuer791 to apply to video-sample-based content recognition systems as proposed. Because Petitioner bears the burden of demonstrating obviousness, its representation that these challenges would have foreclosed a POSITA from recognizing that the inventors possessed the invention based on the common written description necessarily forecloses Petitioner's unexplained, conclusory obviousness rationales.

Petitioner also fails to demonstrate a reasonable likelihood of success for a similar reason. Indeed, "obviousness cannot be sustained with mere conclusory statements; instead, there must be some articulated reasoning with some ***rational underpinning*** to support the legal conclusion of obviousness." *KSR Int'l*, 550 U.S. at 418 (quoting *In re Kahn*, 442 F.3d 997, 988 (Fed. Cir. 2006)). Petitioner's conclusory statement that "a POSITA would have had a reasonable expectation of success" simply because the Steuer750/Steuer791 and Neumeier systems have certain similarities is insufficient (Pet., 53), particularly given that Petitioner does not address or explain why there would have been a reasonable likelihood of success considering the many supposed challenges, obstacles, and peculiarities that Petitioner contends exist. Adopting Petitioner's own logic, Petitioner cannot demonstrate a reasonable likelihood of success "[w]ithout any" explanation of the "ways of dealing with [supposed] 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." Pet., 21. Consequently, even if Steuer750/Steuer791 and Neumeier can qualify as prior art, the Petition must still be denied because Petitioner fails to carry its burden of demonstrating obviousness.

V. Conclusion

For the reasons above, the Petition fails to show a reasonable likelihood of proving any challenged claim is unpatentable. Institution should be denied.

Dated: December 27, 2024

Respectfully submitted,

/*Stephen D. Zinda*/
Stephen D. Zinda (Reg. No. 67,272)
Counsel for Patent Owner

CERTIFICATE OF COMPLIANCE

The undersigned certifies that this document complies with the type-volume limitations set forth in 37 C.F.R. § 42.24 *et seq.* This document, excluding those portions excluded under § 42.24(a)(1), contains 13,999 words as calculated by the program used in creating it.

/*Stephen D. Zinda*/
Stephen D. Zinda

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. §§ 42.6(e)(4), the undersigned certifies that on December 27, 2024, a copy of the foregoing document (*Patent Owner's Preliminary Response*, including all Exhibits), was served via electronic mail on all attorneys of record at the email addresses provided for electronic service listed in Petitioner's mandatory notices, as set forth below:

Naveen Modi (Reg. No. 46,224)
Daniel Zeilberger (Reg. No. 65,349)
Quadeer Ahmed (Reg. No. 60,835)
Daniel J. Blake (*pro hac vice* to be requested)
PAUL HASTINGS LLP
2050 M Street NW
Washington, DC 20036

PH-Samsung-AMRH-IPR@paulhastings.com

/*Stephen D. Zinda*/
Stephen D. Zinda