

TABLE OF CONTENTS

I.	INTRODUCTION AND QUALIFICATIONS	1
II.	SUMMARY OF MATERIALS REVIEWED AND CONSIDERED.....	4
III.	LEGAL UNDERSTANDINGS.....	10
IV.	THE '836 PATENT.....	14
A.	Overview of the '836 Patent (Ex. 1001)	14
B.	The Challenged Claims	18
V.	LEVEL OF ORDINARY SKILL AND PERSPECTIVE APPLIED IN THIS DECLARATION.....	21
VI.	CLAIM CONSTRUCTION	23
VII.	STATE OF THE ART	23
A.	Head-mounted displays	24
B.	Creating Binaural Sound Using HRTFs.....	29
1.	Binaural Sound.....	29
2.	Development of HRTFs.....	33
3.	HRTF Databases	40
4.	Processing Binaural Audio	47
5.	Distance perception.....	50
C.	Combining 3D Sound With Images	52
1.	User Position and Orientation Tracking	54
D.	Processors Used For Rendering Binaural Sound	60
VIII.	PRIOR ART RELIED ON IN THE GROUNDS	63
A.	Finnegan (Ex. 1005).....	63

B.	Begault (Ex. 1006)	66
C.	Norris (Ex. 1007)	70
D.	Algazi (Ex. 1008)	73
E.	Sauk (Ex. 1009)	74
F.	Shivappa (Ex. 1010)	76
G.	Zhou (Ex. 1011)	82
IX.	PETITION GROUNDS	88
X.	GROUND 1: CLAIMS 8-10 AND 15-17 ARE UNPATENTABLE OVER FINNEGAN, ALONE OR IN VIEW OF BEGAULT	89
A.	Reasons Persons of Ordinary Skill in the Art Would Combine Finnegan and Begault	89
B.	Independent Claim 8	94
1.	8[pre]: “An electronic device, comprising:”	94
2.	8[a]: “a display that displays an image that occurs within a near-field distance from a head of a user”	94
3.	8[b]: “one or more processors that process sound for the image”	97
4.	8[c]: “to generate binaural sound that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”	99
5.	8[d]: “wherein the one or more processors include a digital signal processor (DSP)”	105
6.	8[e]: “that processes the sound for the image with far- field head-related transfer functions (HRTFs) while the electronic device is located at the near-field distance from the head of the user”	108

C.	Dependent Claims 9 and 10	110
1.	Claim 9: “The electronic device of claim 8 , wherein the image is an augmented reality (AR) image, and the display is included with electronic glasses worn on the head of the user.”	111
2.	Claim 10: “The electronic device of claim 8 , wherein the image is a virtual reality (VR) image, and the display is included with a head mounted display worn on the head of the user.”	113
D.	Independent Claim 15	114
1.	15[pre]: “An electronic system, comprising”	114
2.	15[a]: “a display that displays an image that is located within a near-field distance from a head of a user viewing the image”	114
3.	15[b]: “one or more processors that process sound for the image”	114
4.	15[c]: “with head-related transfer functions (HRTFs) to generate binaural sound”	115
5.	15[d]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”	115
6.	15[e]: “wherein the coordinate location of the binaural sound continues to occur at the far-field distance located behind the image when the head of the user moves away from viewing the image and the image remains within the near-field distance from the head of the user.”	115
E.	Dependent Claims 16 and 17	118

1.	Claim 16: “The electronic system of claim 15, wherein the display is an augmented reality (AR) display worn on the head of the user viewing the image, and the image is an AR image.”	118
2.	Claim 17: “The electronic system of claim 15, wherein the display is a virtual reality (VR) display worn on the head of the user viewing the image, and the image is an VR image.”	119
XI.	GROUND 2: CLAIMS 8-17 ARE UNPATENTABLE OVER NORRIS, ALONE OR IN VIEW OF BEGAULT AND ALGAZI, AND FURTHER IN VIEW OF SAUK.....	119
A.	Reasons Persons of Ordinary Skill in the Art Would Combine Norris and Begault	119
B.	Reasons Persons of Ordinary Skill in the Art Would Combine Norris and Algazi	120
C.	Reasons Persons of Ordinary Skill in the Art Would Combine Norris, Algazi, and Sauk	123
D.	Ground 2(a) Claims 8-10 and 13-17 are Unpatentable Over Norris, Alone or in View of Begault and Algazi	125
1.	Independent Claim 8	125
a.	8[pre]: “An electronic device, comprising:”	125
b.	8[a]: “a display that displays an image that occurs within a near-field distance from a head of a user”	126
c.	8[b]: “one or more processors that process sound for the image”	127
d.	8[c]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”	129

e.	8[d]: “wherein the one or more processors include a digital signal processor (DSP)”	133
f.	8[e]: “that processes the sound for the image with far-field head-related transfer functions (HRTFs) while the electronic device is located at the near-field distance from the head of the user”	137
2.	Dependent Claims 9, 10, 13, and 14	140
a.	Claim 9: “The electronic device of claim 8 , wherein the image is an augmented reality (AR) image, and the display is included with electronic glasses worn on the head of the user.”	141
b.	Claim 10: “The electronic device of claim 8 , wherein the image is a virtual reality (VR) image, and the display is included with a head mounted display worn on the head of the user.”	141
c.	Claim 13: “The electronic device of claim 8, wherein the image is an augmented reality (AR) image of an intelligent user assistant (IUA) that is located less than one meter away from the head of the user, and a voice of the IUA occurs at a coordinate location behind the AR image of the IUA.”	142
d.	Claim 14: “The electronic device of claim 8, wherein the image is a virtual reality (VR) image of an intelligent user assistant (IUA) that is located less than one meter away from the head of the user, and a voice of the IUA occurs at a coordinate location behind the VR image of the IUA.”	144
3.	Independent Claim 15	145
a.	15[pre]: “An electronic system, comprising”	145
b.	15[a]: “a display that displays an image that is located within a near-field distance from a head of a user viewing the image”	145

c.	15[b]: “one or more processors that process sound for the image”	145
d.	15[c]: “with head-related transfer functions (HRTFs) to generate binaural sound”	146
e.	15[d]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”	146
f.	15[e]: “wherein the coordinate location of the binaural sound continues to occur at the far-field distance located behind the image when the head of the user moves away from viewing the image and the image remains within the near-field distance from the head of the user.”	146
4.	Dependent Claims 16 and 17	148
a.	Claim 16: “The electronic system of claim 15, wherein the display is an augmented reality (AR) display worn on the head of the user viewing the image, and the image is an AR image.”	148
b.	Claim 17: “The electronic system of claim 15, wherein the display is a virtual reality (VR) display worn on the head of the user viewing the image, and the image is an VR image.”	149
E.	Ground 2(b): Dependent Claims 11 and 12 are Unpatentable Over Norris, Alone or in View of Begault and Algazi, and Further in View of Sauk	149
1.	Claim 11: “The electronic device of claim 8 further comprising: one or more sensors that detect rotation of the electronic device, wherein the binaural sound automatically switches to stereo sound in response to the one or more sensors detecting the rotation of the electronic device”	149

2.	Claim 12: “The electronic device of claim 8, wherein the binaural sound automatically switches to stereo sound in response to the electronic device detecting a change to a head orientation of the user that exceeds a threshold value in an azimuth direction.”	152
XII.	GROUND 3: CLAIMS 8-10 AND 15-17 ARE UNPATENTABLE OVER SHIVAPPA, ALONE OR IN VIEW OF ZHOU, AND FURTHER IN VIEW OF SAUK	153
A.	Reasons Persons of Ordinary Skill in the Art Would Combine Shivappa and Zhou.....	153
B.	Reasons Persons of Ordinary Skill in the Art Would Combine Shivappa, Zhou, and Sauk.....	155
C.	Ground 3(a): Claims 8-10 and 15-17 are Unpatentable Over Shivappa, Alone or in View of Zhou	157
1.	Independent Claim 8	157
a.	8[pre]: “An electronic device, comprising:”	157
b.	8[a]: “a display that displays an image that occurs within a near-field distance from a head of a user”	157
c.	8[b]: “one or more processors that process sound for the image”	159
d.	8[c]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”	160
e.	8[d]: “wherein the one or more processors include a digital signal processor (DSP)”	164
f.	8[e]: “that processes the sound for the image with far-field head-related transfer functions (HRTFs) while the electronic device is located at the near-field distance from the head of the user”	165

2.	Dependent Claims 9 and 10	167
a.	Claim 9: “The electronic device of claim 8 , wherein the image is an augmented reality (AR) image, and the display is included with electronic glasses worn on the head of the user.”	167
b.	Claim 10: “The electronic device of claim 8 , wherein the image is a virtual reality (VR) image, and the display is included with a head mounted display worn on the head of the user.”	168
3.	Independent Claim 15	169
a.	15[pre]: “An electronic system, comprising”	169
b.	15[a]: “a display that displays an image that is located within a near-field distance from a head of a user viewing the image”	169
c.	15[b]: “one or more processors that process sound for the image”	170
d.	15[c]: “with head-related transfer functions (HRTFs) to generate binaural sound”	170
e.	15[d]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line- of-sight from the user to the image while the display displays the image to the user at the near- field distance”	170
f.	15[e]: “wherein the coordinate location of the binaural sound continues to occur at the far-field distance located behind the image when the head of the user moves away from viewing the image and the image remains within the near-field distance from the head of the user.”	170
4.	Dependent Claims 16 and 17	172

Declaration for *Inter Partes* Review of USP 11,290,836

a.	Claim 16: “The electronic system of claim 15, wherein the display is an augmented reality (AR) display worn on the head of the user viewing the image, and the image is an AR image.”	172
b.	Claim 17: “The electronic system of claim 15, wherein the display is a virtual reality (VR) display worn on the head of the user viewing the image, and the image is an VR image.”	173
XIII.	Conclusion	173

I. INTRODUCTION AND QUALIFICATIONS

1. I have been retained by Meta Platforms, Inc., Meta Platforms Technologies, LLC, and Twisted Pixel Games, LLC (collectively “Meta” or “Petitioners”) to provide my opinion concerning the validity of certain claims of U.S. Patent No. 11,290,836 (“the ’836 patent,” Ex. 1001), in support of Petitioners’ Petition for *Inter Partes* Review of the ’836 patent.

2. I am a Professor at New York University (NYU) in the Music Technology program since 2006. I was the Associate Director of the Music Technology program 2009 – 2018. I served as Vice-Chair of the Music and Performing Arts Professions Department 2019-2020. From 2007 until 2022, I served as the Director of the Music Technology summer programs.

3. Prior to my position at NYU, I was an Audio Software Specialist (2000-2004) and Audio Research Scientist (2004-2006) at AuSIM Inc where I conducted research and development in 3D audio (binaural audio) for mission critical applications on projects related to the DoD, including the US Army, Navy, Air Force, and NASA. I was an Audio Research Scientist at Sennheiser in 2006.

4. I received a Bachelor of Music degree from McGill University in 1996, with an Honors degree in Computer Applications in Music, and a double major in Piano Performance. I received a Master of Music degree in Music Technology from New York University in 1998. My Master’s Thesis is titled “3D Audio Synthesis.” I

received a Doctor of Philosophy in Music Technology from Northwestern University in 2004. My dissertation is titled “Attention Redirection Based on the Spatial Presentation of Auditory Displays.”

5. At NYU, I have been teaching a graduate-level class titled “3D Audio” since 2007, as well as courses in Digital Signal Theory, Acoustics, Digital Audio Processing, Colloquy in Music Technology. In the past, I taught graduate-level computer programming courses. I am the faculty sponsor of the Immersive Audio Interest Group at NYU – a group focusing on the capture, analysis, synthesis, and reproduction of immersive auditory environments, including the capture and representation of spatial and 3D sound using binaural, transaural, multi-channel and soundfield techniques, audio display devices, virtual auditory environment simulation, games, characterization, and classification of spaces as well as augmented audio reproduction. I advise undergraduate, Master’s and PhD students in the subject areas of immersive, spatial and 3D audio.

6. My own research focuses on the simulation and applications of immersive and 3D audio including the capture, analysis and synthesis of auditory environments, HRTF measurement and analysis, and designing immersive experiences. Applications of my work include Augmented and Virtual Reality, gaming, sound recording and production, mission critical, and augmented acoustic sensing.

7. I have numerous publications, including a book titled “Immersive sound: the art and science of binaural and multi-channel audio” (Routledge Press), a book chapter, over 60 refereed journal, and conference publications. My publications cover the areas of the science and applications of spatial audio.

8. I have received over \$11M in funding from the National Science Foundation (NSF), National Institute of Health (NIH), Office of Naval Research, US Army, US Navy and NASA.

9. I have been a member of the Audio Engineering Society (AES) since 2004. The AES is the largest professional society in the world that is devoted to audio research, education, and applications. I have served on the Board of Governors of the AES in various roles since 2009. I am the Chair of the AES Events Coordination Committee, Co-Chair of the Laws and Resolutions Committee, and currently serve on the Board of Directors of the AES. I was President of the AES in 2020.

10. I served as Chair of AES Convention in 2009, 2017 and 2019, Chair of the International Community of Auditory Display conference in 2014, and workshops Co-Chair of the AES Conference on Audio for Virtual and Augmented Reality in 2022.

11. Over the course of my academic and professional career, I have received a number of awards. My *curriculum vitae*, which includes a more detailed summary

of those awards, my background, experience, and publications, is attached to the accompanying Petition as Ex. 1004.

II. SUMMARY OF MATERIALS REVIEWED AND CONSIDERED

12. The opinions contained in this declaration are based on the documents I reviewed and my knowledge and professional judgment. In forming the opinions expressed in this declaration, while drawing on my experience, I reviewed the following documents:

Ex.	Description
1001	U.S. Patent No. 11,290,836 (“’836 patent”)
1004	Curriculum Vitae of Dr. Agnieszka Roginska
1005	Daniel J. Finnegan et al., <i>Compensating for Distance Compression in Audiovisual Virtual Environments Using Incongruence</i> , Chi’ 16: Proc. of the 2016 CHI Conf. on Hum. Factors in Computing Sys. (2016) (“Finnegan”)
1006	Durand R. Begault, <i>3-D Sound for Virtual Reality and Multimedia</i> , NASA/TM—2000-209606 (Aug. 2000) (“Begault”)
1007	U.S. Patent No. 9,226,090 (“Norris”)
1008	V.R. Algazi et al., <i>The CIPIC HRTF Database</i> , IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2001) (“Algazi”)
1009	U.S. Patent No 7,876,903 (“Sauk”)
1010	U.S. Patent No. 10,979,843 (“Shivappa”)
1011	Zhiying Zhou et al., <i>An Experimental Study on the Role of Software Synthesized 3D Sound in Augmented Reality Environments</i> , 16 Interacting with Computs. 989 (2004) (“Zhou”)
1014	U.S. Patent No. 4,774,515 (“Gehring”)

Declaration for *Inter Partes* Review of USP 11,290,836

Ex.	Description
1015	U.S. Patent No. 2,955,156 (“Heilig”)
1016	U.S. Patent No. 9,041,622 (“McCulloch”)
1017	U.S. Patent No. 10,509,466 (“Miller”)
1018	U.S. Patent No. 9,767,618 (“Najaf-Zadeh”)
1019	U.S. Patent No. 9,232,336 (“Nishio”)
1020	U.S. Patent No. 9,578,440 (“Otto”)
1021	U.S. Patent No. 1,624,486 (“Fletcher”)
1022	U.S. Patent No. 1,855,149 (“Jones”)
1023	U.S. Patent No. 8,873,761 (“Fukui”)
1024	U.S. Patent No. 9,787,846 (“Srinivasan”)
1025	U.S. Patent Application Publication No. US 2012/0183161 A1
1026	U.S. Patent No. 9,530,426 (“Wright”)
1027	U.S. Patent Application Publication No. US 2017/0115488 A1 (“Ambrus”)
1028	U.S. Patent No. 7,710,654 (“Ashkenazi”)
1029	Matthias Geier et al., <i>The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods</i> (Convention Paper 7330), Audio Eng’g Soc’y 124th Convention (2008)
1030	V. Ralph Algazi & Richard O. Duda, <i>Headphone-Based Spatial Sound</i> , 28 IEEE Signal Processing Mag. 33 (2011) (“Algazi 2011”)
1031	S.S. Fisher et al., <i>Virtual Interface Environment Workstations</i> , Procs. of the Hum. Factors Soc’y, 32nd Annual Meeting 91 (1988)
1032	Wolfgang Fohl et al., <i>A System-On-Chip Platform for HRTF-Based Realtime Spatial Audio Rendering With an Improved Realtime Filter Interpolation</i> , 4 Int’l J. Advances Intelligent Sys. 309 (2011)

Ex.	Description
1033	Jason Jerald, The VR Book – Perception and Interaction Design for Virtual Reality (M. Tamer Özsu ed., 2015) (“THE VR BOOK”)
1034	Masayuki Morimoto & Yoichi Ando, <i>On the Simulation of Sound Localization</i> , 1 J. of Acoustical Soc’y Japan (E) 167 (1980)
1035	C. Phillip Brown & Richard O. Duda, <i>A Structural Model for Binaural Sound Synthesis</i> , 6 IEEE Transactions on Speech and Audio Processing 476 (1998)
1036	Douglas S. Brungart & William M. Rabinowitz, <i>Auditory Localization of Nearby Sources. Head-related transfer Functions</i> , 106 J. Acoustical Soc’y of Am. 1465 (1999)
1037	Kai-Uwe Doerr et al., <i>Evaluation of a Low-Cost 3D Sound System for Immersive Virtual Reality Training Systems</i> , 13 IEEE Transactions on Visualization and Comput. Graphics 204 (2007)
1038	<i>Virtual Reality</i> , Cryptologic Quarterly 21 (Dec. 1, 2011)
1039	Stephan Paul, <i>Binaural Recording Technology: A Historical Review and Possible Future Developments</i> , 95 Acta Acustica United with Acustica 767 (2009)
1040	David Romblom & Bryan Cook, <i>Near-Field Compensation for HRTF Processing</i> (Convention Paper 7611), Audio Eng’g Soc’y 125th Convention (2008)
1041	Amy Turner et al., <i>Can the Perception of Depth in Stereoscopic Images Be Influenced by 3D Sound?</i> , 7863 Stereoscopic Displays & Applications (2011)
1042	Elizabeth M. Wenzel et al., <i>Development of a Three-Dimensional Auditory Display System</i> , 20 SIGCHI Bull. 52 (1988)
1043	Elizabeth M. Wenzel et al., <i>Localization with Non-Individualized Virtual Acoustic Display Cues</i> , Chi’ 91: Proceedings of the SIGCHI Conf. on Hum. Factors in Comput. Sys. 351 (1991)

Declaration for *Inter Partes* Review of USP 11,290,836

Ex.	Description
1044	Frederic L. Wightman & Doris J. Kistler, <i>Headphone Simulation of Free-Field Listening. II: Psychophysical Validation</i> , 85 J. Acoustical Soc’y of Am. 868 (1989)
1045	Steven M. Kaplan, Wiley Electrical and Electronics Engineering Dictionary 335 (2004)
1046	Dmitry N. Zotkin et al., <i>Customizable Auditory Displays</i> , Proc. of the Inte’l Conf. on Auditory Display (July 2002)
1047	<i>Asserted U.S. Patent No. 11,290,836 “Providing Binaural Sound Behind an Image Being Displayed with an Electronic Device</i> , Plaintiff’s Infringement Contentions Exhibit 27, <i>Eight KHZ, LLC v. Meta Platforms, Inc.</i> , No. 6:22-cv-00575 (W.D. Tex. Mar. 3, 2023)
1048	<i>Asserted U.S. Patent No. 11,290,836 “Providing Binaural Sound Behind an Image Being Displayed with an Electronic Device</i> , Plaintiff’s Infringement Contentions Exhibit 18, <i>Eight KHZ, LLC v. Meta Platforms, Inc.</i> , No. 6:22-cv-00575 (W.D. Tex. Mar. 3, 2022)
1049	<i>Asserted U.S. Patent No. 11,290,836 “Providing Binaural Sound Behind an Image Being Displayed with an Electronic Device</i> , Plaintiff’s Infringement Contentions Exhibit 36, <i>Eight KHZ, LLC v. Meta Platforms, Inc.</i> , No. 6:22-cv-00575 (W.D. Tex. Mar. 3, 2023)
1050	Rasmus Konig Sorenson, <i>Review: Oculus Rift DK2 – Next-Level Virtual Reality</i> , flatpanelshd (Aug. 13, 2014), https://www.flatpanelshd.com/review.php?subaction=showfull&id=1407909173
1051	U.S. Patent No. 9,392,212 (“Ross”)
1052	Mark Wherry, <i>Audio Add-ons For Macbook Pro, Apple Notes, SoundonSound</i> (Mar. 2007), https://www.soundonsound.com/techniques/audio-add-ons-macbook-pro
1056	U.S. Patent No. 5,173,944 (“Begault ‘944”)
1057	U.S. Patent No. 5,438, 623 (“Begault ‘623”)

Ex.	Description
1059	Areti Andreopoulou & Agnieszka Roginska, <i>Towards the Creation of a Standardized HRTF Repository</i> , Audio Eng'g Soc'y Convention Paper 8571, Oct. 20-23, 2011
1060	Chan Jun Chun et al., <i>Deep Neural Network Based HRTF Personalization Using Anthropometric Measurements</i> , Audio Eng'g Soc'y Convention Paper 9860, Oct. 18-21, 2017
1061	Karen Collins, <i>Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design</i> (2008)
1062	Sungmok Hwang & Youngjin Park, <i>HRIR Customization in the Median Plane Via Principal Components Analysis</i> , AES 31st Int'l Conf., London, UK (2007)
1063	Donald H. Merson & John N. Bowers, <i>Absolute and Relative Cues for the Auditory Perception of Egocentric Distance</i> , 8 Perception 311-22 (1979)
1064	Matthew Fam Kai Liang, <i>The AweSOME 5: High-End Sound Cards Review</i> (Oct. 11, 1998), https://assets.hardwarezone.com/2009/reviews/sound/roundup/soundcards.htm
1065	Matthew Fam Kai Liang, <i>The AweSOME: 5(Part 2): Explore: Technologies Behind ...</i> (Sept. 27, 1998), https://assets.hardwarezone.com/2009/reviews/sound/roundup/sound2.htm
1066	Pavel Zahorik, <i>Assessing Auditory Distance Perception Using Virtual Acoustics</i> , 111 J. Acoustical Soc'y. Am. 1832-46 (2002)
1067	Korean Patent Application Publication No. KR 10-2011-0022252, including English Translation and Certification ("Moon")
1068	<i>Oculus: Complete Guide — History, Products, Founding, and More</i> , History-Computer.com (updated May 13, 2013), https://history-computer.com/oculus-history/

Ex.	Description
1069	Pat Lawler, <i>What's So Special About the Digital Signal Processor?</i> , Qualcomm: OnQ Blog (Dec. 5, 2013), https://www.qualcomm.com/news/onq/2013/12/whats-so-special-about-digital-signal-processor (“Qualcomm blog”)
1070	John D. Sutter, <i>Microsoft's Talking Paper Clip is Back</i> , CNN.com (Apr. 29, 2011), http://www.cnn.com/2011/TECH/web/04/29/microsoft.paper.clip/index.html#:~:text=Well%2C%20like%20it%20or%20not,2%3A%20Clippy's%20Second%20Chance.%22
1071	<i>Super Mario 64 Characters</i> , Mario.fandom.com, https://mario.fandom.com/wiki/Category:Characters_in_Super_Mario_Galaxy (last visited June 5, 2023)
1072	Mona Lalwani, <i>Surrounded By Sound: How 3D Audio Hacks Your Brain</i> , Theverge.com (Feb. 12, 2015), https://www.theverge.com/2015/2/12/8021733/3d-audio-3dio-binaural-immersive-vr-sound-times-square-new-york
1073	<i>Aureal Semiconductor</i> , Dosdays.co.uk, https://dosdays.co.uk/topics/Manufacturers/aureal.php (last visited June 5, 2013)
1074	Mike Andrawes, Audio Report – Comdex '99, Anandtech.com (Nov. 30, 1999), https://www.anandtech.com/show/417
1075	Elizabeth M. Wenzel et al., <i>A Virtual Display System for Conveying Three-Dimensional Acoustic Information</i> , 32nd Proc. of the Hum. Factors Soc'y Annual Meeting 86 (1988) (“Wenzel 1988b”)
1076	Jean-Marc Jot et al., <i>Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony</i> , 98th Proc. of the Convention of the Audio Eng'g Soc'y (1995) (“Jot”)
1077	Henrik Moller et al., <i>Head-Related Transfer Functions of Human Subjects</i> , 43 J. Audio Eng'g Soc'y 300 (1995) (“Moller”)
1078	David A. Burgess, <i>Techniques for Low Cost Spatial Audio</i> , ACM Symp. on User Interface Software and Tech. 53 (1992) (“Burgess”)

Ex.	Description
1079	Navarun Gupta et al., <i>HRTF database at FIU DSP Lab</i> , IEEE Int'l Conf on Acoustics, Speech, and Signal Processing (2010) ("Gupta")
1080	John C. Middlebrooks, <i>Individual Differences in External-Ear Transfer Functions Reduced by Scaling in Frequency</i> , 106 J. Acoustical Soc'y Am. 1480 (1999) ("Middlebrooks")
1081	Elizabeth M. Wenzel, <i>Localization in Virtual Acoustic Displays</i> , 1 Presence Teleoperators & Virtual Env'ts 80 (1992) ("Wenzel 1992")
1082	U.S. Patent No. 9,037,468 ("Osman")
1083	Durant R. Begault et al., <i>Design and Verification of HeadZap, A Semi-Automated HRIR Measurement System</i> , 120th Proc. of the Audio Eng'g Soc'y Convention (2006) ("HeadZap")
1084	Bill Gardner & Keith Martin, <i>HRTF Measurements of a KEMAR Dummy-Head Microphone</i> , MIT Media Lab Perceptual Computing – Tech. Rep. 280 (1994)
1085	David M. Howard & Jamie A.S. Angus, <i>Acoustics and Psychoacoustics</i> (3d ed. 2006)

13. My opinions are also guided by my appreciation and understanding of how a person of ordinary skill in the art would have understood the claims of the '836 patent at the time of the alleged invention, which I have been asked to assume is May 24, 2017.

III. LEGAL UNDERSTANDINGS

14. I am not an attorney, but I have been informed by counsel of certain legal standards and have applied those standards, described below, in formulating my opinions.

15. I understand that the claims of a patent define the limits of the patentee's exclusive rights. I understand that to determine the scope of the claimed invention, courts may construe claim terms, the meaning of which the parties may dispute. I understand that claim terms should generally be given their ordinary and customary meaning as understood by a person of ordinary skill in the art at the time of filing of the patent application.

16. I understand that claims must be construed in light of and consistent with the intrinsic evidence. In this context, I understand that intrinsic evidence includes the claims themselves, the written disclosure in the specification, and the patent's prosecution history, including prior art that was considered by the United States Patent and Trademark Office. I understand that the specification is always highly relevant to the claim construction analysis and often is the single best guide to the meaning of a disputed term. I understand that extrinsic evidence may also be considered when construing claims and may include, for example, technical dictionaries, technical publications and books, treatises, and expert testimony, though the intrinsic record will be more persuasive as to the meaning of a particular term.

17. I understand that a claim of a patent is anticipated when a single prior art document describes every element of the claimed invention. The reference must not only disclose all elements of the claim, but must also disclose those elements arranged or combined as in the claim.

18. I understand that a prior art reference can render a patent claim obvious to one of ordinary skill in the art if the differences between the subject matter set forth in the patent claim and the prior art are such that the subject matter of the claim would have been obvious at the time the claimed invention was made.

19. In analyzing obviousness, I understand that it is important to consider the scope of the claims, the level of skill in the relevant art, the scope and content of the prior art, the differences between the prior art and the claims, and any secondary considerations (or objective indicia of non-obviousness).

20. I understand that claimed subject matter is obvious when a person of ordinary skill in the art would have been motivated to combine preexisting elements and that combination yields no more than would have been expected.

21. I also understand that in assessing whether a claim is obvious, one must consider whether the claimed improvement is more than the predictable use of prior art elements according to their established functions. I understand that there need not be a precise teaching in the prior art directed to the specific subject matter of a claim because one can take account of the inferences and creative steps that a person of skill in the art would employ. I further understand that a person of ordinary skill is a person of ordinary creativity, not an automaton.

22. I understand that obviousness cannot be based on the hindsight combination of components selectively culled from the prior art. I understand that in

an obviousness analysis, neither the motivation nor the avowed purpose of the inventors controls the inquiry. Any need or problem known in the field at the time of the invention and addressed by the patent can provide a reason for combining elements. For example, I understand that it is important to consider whether there existed at the time of the invention a known problem for which there was an obvious solution encompassed by the patent's claims. I understand that known techniques can have obvious uses beyond their primary purposes, and that in many cases a person of ordinary skill can fit the teachings of multiple pieces of prior art together like pieces of a puzzle.

23. I understand that certain issues relating to validity must be judged from the perspective of a person of ordinary skill in the relevant art. I further understand that factors that may be considered in determining the level of skill of a person of ordinary skill in the art include, but are not limited to, (1) the type of problems encountered in the art; (2) prior art solutions to those problems; (3) rapidity with which innovations are made; (4) sophistication of the technology; and (5) the educational level of active workers in the field.

24. I have been informed and understand that a person of ordinary skill in the art is a hypothetical person who is presumed to be aware of all pertinent prior art, thinks along conventional wisdom in the art, and is a person of ordinary creativity.

Importantly, this hypothetical person is of ordinary skill in the art to which the claimed subject matter pertains.

25. I understand that, when there is a reason to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill in the art has good reason to pursue the known options within his or her technical grasp. I further understand that, if this leads to the anticipated success, it is likely the product not of innovation but of ordinary skill and common sense, which bears on whether the claim would have been obvious.

IV. THE '836 PATENT

A. Overview of the '836 Patent (Ex. 1001)

26. The '836 patent is entitled “Providing Binaural Sound Behind an Image Being Displayed With An Electronic Device.” I understand that the '836 patent is assigned to Eight Khz, LLC (“Patent Owner”).

27. The '836 patent is directed to methods, devices, and systems that purportedly “improve the user experience” by processing “sound for [an] image into binaural sound for a user.” '836 patent (Ex. 1001) Abstract. Specifically, the '836 patent purports to improve the provision of binaural sound by providing “binaural sound [that] has a sound localization point (SLP)” “behind [an] image while the image is located a near-field distance from a head of the user.” '836 patent at 2:65-3:4.

28. According to the '836 patent, “[o]ne problem with electronically generated binaural sound or three-dimensional (3D) sound rendering is that listeners have a difficult time determining a location where the sound originates when this location is close to the listener.” *Id.* at 2:11-24. For example, with “near-field” sounds where “the location of the origination of binaural sound is convolved to less than one meter from the person,” “it may be more difficult for the person to determine a location or direction of a binaural sound.” *Id.* at 2:14-19, *see also id.* at 27:23-25 (defining a “near-field distance” as “a distance less than one meter” and a “far-field distance” as “a distance greater than or equal to one meter.”). On the other hand, “a listener may have more success in external localization when the binaural sound is processed to originate from about one meter or more from the person (considered to be ‘far-field’).” *Id.* at 2:19-23.

29. The '836 patent states that “[h]ead related transfer functions (HRTFs), [which] describe how a sound wave changes as it interacts with the torso, head, and pinnae of the listener,” “do not vary substantially in the far-field range.” *Id.* at 2:24-27. However, “HRTFs can vary significantly in the near-field range.” *Id.* at 2:27-28. And these variations “can be significant depending on the frequency of the sound.” *Id.* at 2:28-31. Thus, “modeling near-field virtual auditory space with HRTFs can be quite complex.” *Id.* at 2:31-33. The consequence is that “when an electronic device attempts to convolve sound so the binaural sound originates to the listener in the near-

field range,” “[t]he listener may perceive the sound as originating from the wrong location or be unable to localize an origin of the sound.” *Id.* at 2:24-38. The ’836 patent further explains that “convolving sound with near-field HRTFs can be computationally complex and processor-intensive.” *Id.* at 2:39-40.

30. The ’836 patent describes “[a]nother problem with electronically generated binaural sound” that “occurs when the perceived location of where the sound originates does not match a location of an image of the sound,” such as when “[t]he listener [] hears the stream of sound as if the sound is originating from a certain point in space where an image is located in the physical or virtual environment around the listener.” *Id.* at 2:41-55. This occurs, for example, with a “virtual sound source” from “a video call or telephone software application, computer game, or a social VR (virtual reality) or AR (augmented reality) software application.” *Id.* The ’836 patents states that “[t]he effectiveness of the user experience is lost on the listener if the listener is not able to localize or hear the convolved or processed sound at the location of the image,” which occurs “when the image is located at a near-field distance from the listener or the image is occluded or not visible to the listener.” *Id.* at 2:53-66.

31. The ’836 patent purports to solve these problems discussed above with techniques to process sound “so the listener believes that the sound originates from [a] sound source when in fact the sound is convolved to a different location,” e.g., by

“trick[ing] the listener into believing that the sound originates from the sound source when in fact the sound is convolved to originate from a location that is behind the sound source.” *Id.* at 4:65-5:11 (disclosing “convolv[ing] [] sound so the listener believes that the sound originates from the sound source when in fact the sound is convolved to a different location”), 5:64-67 (providing that the “position to where the sound is convolved is not the location of the sound source” but instead “is behind the sound source from the point-of-view of the listener”), 6:26-48 (providing that a “listener [] perceives that the sound originates from the image even though the sound is convolved to originate from a position that is behind the image”). The ’836 patent provides that this can be accomplished by “process[ing] sound with HRTFs having coordinates of a far-field location even though the location of the source of the sound (e.g., an image being displayed) is a near-field distance from the listener.” *Id.* at 6:49-67.

32. The ’836 patent also discloses that such sound processing can be accomplished with a digital signal processor (or “DSP”). *See, e.g., id.* at 4: 4:26-31 (“One or more processors (such as a digital signal processor or DSP) processes or convolves the sound to a location that is not coincident with or not at the location of the sound source. For example, a DSP processes the sound to localize to a coordinate location that does not match a coordinate location of the sound source.”), 27:43-44

(“A processor (such as a digital signal processor or DSP) processes the sound with SLI (including a pair of HRTFs).”).

B. The Challenged Claims

33. The '836 patent has 18 claims, including independent claims 1, 8, and 15.

I understand that claims 8-17 are being challenged in Petitioners' petition for *inter partes* review.

34. Independent claim 8 recites:

8. An electronic device, comprising:
 - a display that displays an image that occurs within a near-field distance from a head of a user; and;
 - one or more processors that process sound for the image to generate binaural sound that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance, wherein the one or more processors include a digital signal processor (DSP) that processes the sound for the image with far-field head-related transfer functions (HRTFs) while the electronic device is located at the near-field distance from the head of the user.

'836 patent at cl. 8.

35. Claim 9 depends on claim 8 and recites: “9. The electronic device of claim 8, wherein the image is an augmented reality (AR) image, and the display is included with electronic glasses worn on the head of the user.” ’836 patent at cl. 9.

36. Claim 10 depends on claim 8 and recites: “10. The electronic device of claim 8, wherein the image is a virtual reality (VR) image, and the display is included with a head mounted display worn on the head of the user.” ’836 patent at cl. 10.

37. Claim 11 depends on claim 8 and recites: “11. The electronic device of claim 8 further comprising: one or more sensors that detect rotation of the electronic device, wherein the binaural sound automatically switches to stereo sound in response to the one or more sensors detecting the rotation of the electronic device.” ’836 patent at cl. 11.

38. Claim 12 depends on claim 8 and recites: “12. The electronic device of claim 8, wherein the binaural sound automatically switches to stereo sound in response to the electronic device detecting a change to a head orientation of the user that exceeds a threshold value in an azimuth direction.” ’836 patent at cl. 12.

39. Claim 13 depends on claim 8 and recites: “13. The electronic device of claim 8, wherein the image is an augmented reality (AR) image of an intelligent user assistant (IUA) that is located less than one meter away from the head of the user, and a voice of the IUA occurs at a coordinate location behind the AR image of the IUA.” ’836 patent at cl. 13.

40. Claim 14 depends on claim 8 and recites: “14. The electronic device of claim 8, wherein the image is a virtual reality (VR) image of an intelligent user assistant (IUA) that is located less than one meter away from the head of the user, and a voice of the IUA occurs at a coordinate location behind the VR image of the IUA.” ’836 patent at cl. 14.

41. Claims 15-17 are nearly identical to claims 8-10.

42. Independent claim 15 recites:

15. An electronic system, comprising:

a display that displays an image that is located within a near-field distance from a head of a user viewing the image; and
one or more processors that process sound for the image with head-related transfer functions (HRTFs) to generate binaural sound that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance, wherein the coordinate location of the binaural sound continues to occur at the far-field distance located behind the image when the head of the user moves away from viewing the image and the image remains within the near-field distance from the head of the user.

’836 patent at cl. 15.

43. Claim 16 depends on claim 15 and recites: “16. The electronic system of claim 15, wherein the display is an augmented reality (AR) display worn on the

head of the user viewing the image, and the image is an AR image.” ’836 patent at cl. 16.

44. Claim 17 depends on claim 15 and recites: “17. The electronic system of claim 15, wherein the display is a virtual reality (VR) display worn on the head of the user viewing the image, and the image is an VR image.” ’836 patent at cl. 17.

V. LEVEL OF ORDINARY SKILL AND PERSPECTIVE APPLIED IN THIS DECLARATION

45. As discussed above, I understand that issues relating to validity must be judged from the perspective of a person of ordinary skill in the relevant art.

46. With respect to the ’836 patent, a person of ordinary skill in the art at the time of the earliest claimed priority date, May 24, 2017, would have been familiar with the development and implementation of virtual and/or augmented reality systems, and underlying technology used therein.

47. A person of ordinary skill in the art would have had a bachelor’s degree in computer science/engineering, or a related field, with coursework in computer programming or virtual/augmented reality, and at least two years of industry experience, or academic research experience, in virtual/augmented reality, audio engineering, or an equivalent. In my view, more education can supplement relevant experience and vice versa.

48. At the time of the earliest claimed priority date for the '836 patent (May 24, 2017), I met the qualifications to be a person of ordinary skill in the art.

49. I also believe that I have a sufficient level of knowledge, experience, and education to provide expert opinions in the field of the '836 patent.

50. While my own level of skill exceeded that of the ordinary level of skill as of May 2017, I am well-acquainted with the actual performance and capabilities of a person of ordinary skill in the art as defined above. For example, during the relevant timeframe, I was supervising and evaluating the performance of such individuals, such as my students in the Bachelor's, Master's and PhD program, as well as adjunct faculty who were performing research in this area. Furthermore, my participation in professional organizations and in workshops and events exposes me to the state of the industry and the level of experience of individuals who, like I did, had at least the ordinary level of skill.

51. In forming and expressing my opinions in this declaration, I have applied the perspective of a person of ordinary skill in the art, as of May 2017. Unless otherwise stated, my opinions refer to the knowledge of one of ordinary skill in the art at the relevant time.

VI. CLAIM CONSTRUCTION

52. In my analysis, I have applied the plain and ordinary meanings of the claim terms in the '836 patent, as a person of ordinary skill in the art would have understood them in the context of that patent in approximately May 2017.

53. My analysis does not depend on any particular or precise construction of a term, because I did not find that to be necessary to provide my opinions regarding the relationship between the prior art and the claims of the '836 patent. The opinions that I provide in this declaration would remain true under any reasonable construction of the claim terms in the '836 patent.

VII. STATE OF THE ART

54. In this section, I provide overviews of technologies and prior art pertinent to the '836 patent and to the state of the art as of the '836 patent's earliest possible priority date of May 24, 2017. In later sections, I evaluate in detail how the prior art discloses and renders obvious the recitations of each limitation in the challenged claims, and the challenged claims as a whole.

55. The prior art references described below and discussions of what was known to a person of ordinary skill in the art provide a general description of the state of the art at the time of the invention, provide additional motivation to modify the primary references evaluated in the grounds of invalidity with the knowledge of a person of ordinary skill in the art or other references cited herein, and provide support

for why a person of ordinary skill in the art would have had a reasonable expectation of success in combining the teachings of the references. The additional prior art references identified in this section are exemplary in nature and are included to provide factual support for my opinions.

A. Head-mounted displays

56. Head-mounted displays (“HMDs”) for virtual reality (VR) environments were well-known before the date of the alleged invention. In fact, HMDs are “the most common visual interface to virtual environments.” NSA, *Virtual Reality* (Dec. 1, 2011) (Ex. 1038) at 25.

57. As early as 1988, NASA had already developed a Virtual Interface Environment Workstation (VIEW) system which provided both “a virtual auditory and stereoscopic image surround that [wa]s responsive to inputs from the operator’s position, voice, and gestures.” Fisher, S., *Virtual Environment Interface Workstations*, Proceedings of the Human Factors Society 32nd Annual Meeting, Anaheim, Ca., 91-95 (1988) (“Fisher”) (Ex. 1031) at 91. The HMD of the VIEW used “two passive, twisted nematic, monochromatic, liquid crystal display screens presented to each eye of the user through wide-angle optics” as well as headphones to “present[] a wide variety of binaural sounds to the operator.” *Id.* at 92-93.

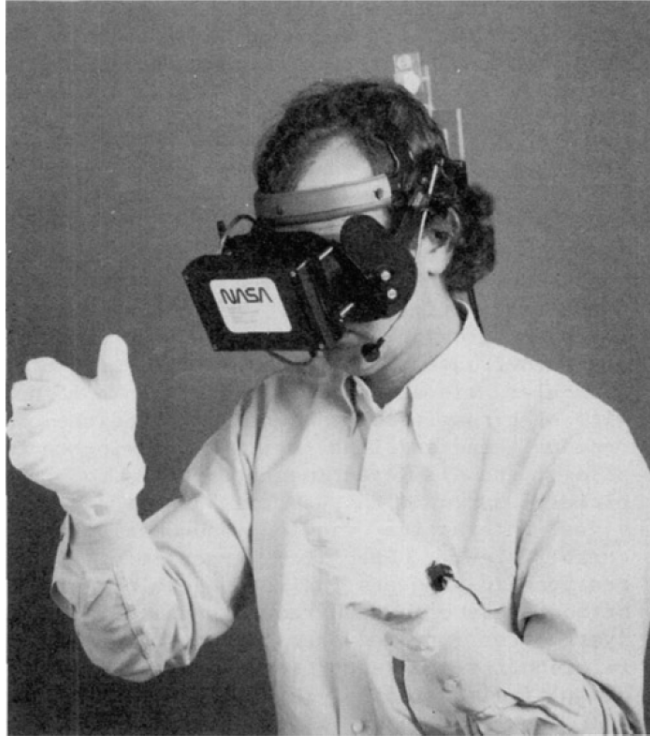
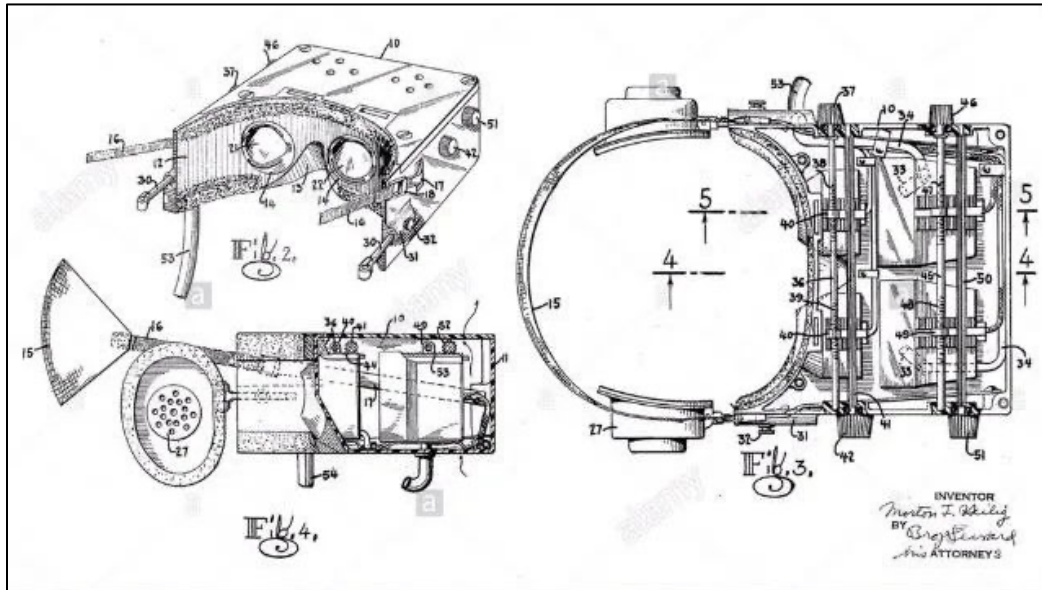


Figure 1. Head-coupled, stereoscopic display system with DataGlove tactile input technology for virtual object manipulation.

Id. at Fig. 1. The VIEW system utilized HRTF-based audio and “maximize[d] flexibility and portability by synthetically generating three-dimensional sound in realtime for delivery through headphones,” with measured HRTFs where “[p]robe microphones [we]re placed near the two eardrums of a listener seated in the anechoic chamber... The subject is at the center of a semicircular arc (1.4 m radius).” Ex. 1075, Wenzel, E. M., Wightman, F. L., & Foster, S. H., A Virtual Display System for Conveying Three-Dimensional Acoustic Information, *Proceedings of the Human Factors Society Annual Meeting*, 32(2), 86-90. (1988) (“Wenzel 1988b”) at 87.

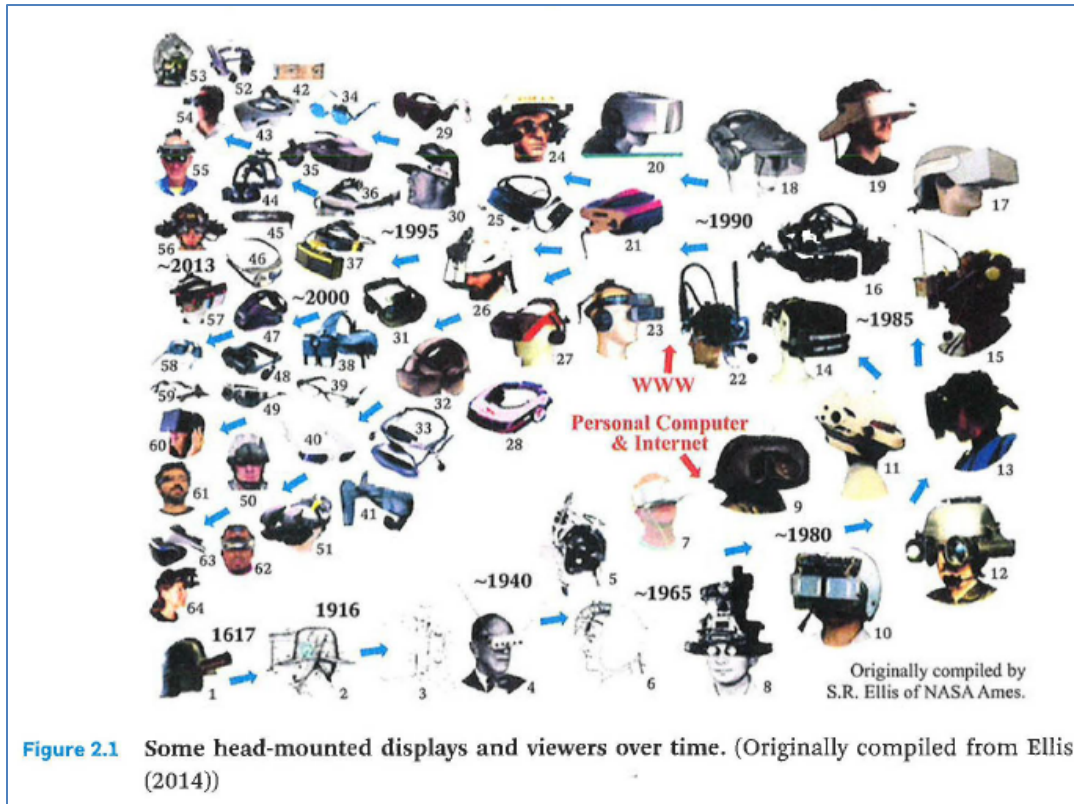
58. Additionally, HMDs were well-known even before NASA's VIEW system. For example, a 1960 patent disclosed the use of stereo displays to create 3D peripheral images, as well as stereo and binaural sound played through adjustable earphones:



59. U.S. Patent No. 2,955,156 ("Heilig") (Ex. 1015) at 3:18-21 ("By the independent feeding of right and left eyes with images taken by separate film or TV cameras equipped with wide angle lenses, a truly three dimensional peripheral image is created."), 1:37-40 ("Another object is to provide ear phones which are so designed that the outer ear is completely free and untouched, thus allowing the ear phones to operate fully as sound focusing organs."), 3:14-17, ("the spectator is given a complete sensation of reality, i.e., moving three dimensional images, which may be in color, with 100% peripheral vision, binaural sound, scents and air breezes"), 3:22-25 ("By

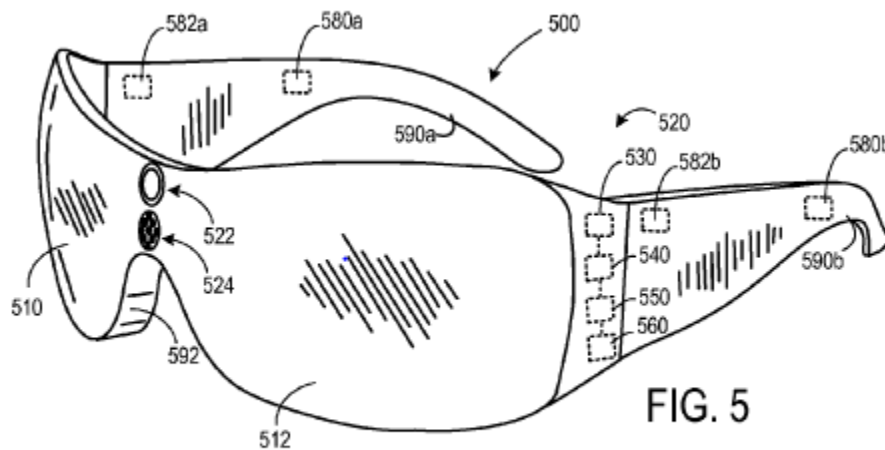
providing independent ear phones, which are fed separately from sounds recorded by independent microphone systems, excellent stereophonic or binaural sound 1.”), Figs. 2-4. I discuss sound localization and binaural sound in further detail in Section VII.B.

60. The following figure, from Jerald, *THE VR BOOK: PERCEPTION AND INTERACTION DESIGN FOR VIRTUAL REALITY* (2015) (“*THE VR BOOK*”) (Ex. 1033) depicts some of the many different types of HMDs that were known to persons of ordinary skill in the art before May 2017:



THE VR BOOK at Fig. 2.1.

61. As the figure above shows, head mounted displays can take many shapes and forms. For example, some may be a “near-eye” display device such as electronic glasses that are worn by the user, as shown below.



U.S. Patent No. 9,530,426 (“Wright”) (Ex. 1026), Fig. 5; *see also id.* Fig. 1; *see also* U.S. Patent No. 9,226,090 (Ex. 1007) (“Norris”) 31:52-32:48 (describing a “wearable electronic device” such as “electronic glasses with a display that provides an image or video” used for an “electronic call” that displays “a virtual image of the speaker”).

62. Persons of ordinary skill in the art have long known that electronic glasses, like the “near-eye display device” shown above, can be used for “augmented reality display[s]” and “virtual reality display[s].” Wright at 1:36-38; *see also* U.S. Patent No. 10,979,843 (Ex. 1010) (“Shivappa”) at 13:64 (disclosing that its HMD “may be a pair of virtual reality goggles.”); Zhou, Z., *An experimental study on the*

role of software synthesized 3D sound in augmented reality environments. Interacting with Computers 16, 5, 989–1016 (2004) (Ex. 1011) (“Zhou”) at 1004-09 (describing a virtual game environment in which players wearing electronic glasses “seamless[ly] transition from AR to VR”), Fig. 1 (showing the electronic glasses), Fig. 11 (showing the “AR to VR” transition).

63. Persons of ordinary skill in the art also knew and understood for years before May 2017 that HMDs display images on lenses located inches away from a user’s eyes, to immerse the user in the images displayed. As the above-discussed references and images show, that is how HMDs work.

B. Creating Binaural Sound Using HRTFs

1. Binaural Sound

64. Processing sound to generate binaural sound, including with head-related transfer functions (HRTFs), was well-known to persons of ordinary skill in the art long before May 2017. Binaural simulation and the impact of individualized HRTFs were studied at length in the 1980s and 1990s. *See* Wightman, *Headphone simulation of free-field listening. II: Physiological validation*, 85 J. Acoust. Soc. Am., 866-878 (Feb. 1989) (“Wightman”) (Ex. 1044); *see also* Begault at 57 (disclosing studies by Wightman & Kistler).

65. To begin, binaural sound, also referred to as 3D audio, virtual acoustics, or spatialized audio, refers to sound that is processed to “provide the listener with a

three dimensional virtual audio environment.” U.S. Patent No. 7,876,903 (“Sauk”) (Ex. 1009) at 1:15-16. This type of audio allows a listener to be immersed in an environment and allows for a more realistic experience.

66. Binaural sound that emanates from different spatial locations around a listener’s head is different from stereophonic sound that is heard in each ear independently and monophonic sound that is heard in both ears simultaneously. These different spatial locations to which binaural sound localizes, what the ’836 patent calls sound localization points (or SLPS), allow a user to “perceive various virtual audio sources as located in various x,y,z locations in world space.” U.S. Patent No. 10,509,466 (“Miller”) (Ex. 1017) at 23:53-59.

67. Methods of generating and managing binaural sound have been known for decades. By the 1980s, processors such as the Convolvotron were capable of processing multiple sound sources in real-time using HRTFs. *See* Begault, *3-D Sound for Virtual Reality and Multimedia*, NASA (August 2000) (Ex. 1006) (“Begault”) at 165-70. In the 1990s, companies such as Aureal and Creative Labs incorporated positional audio through HRTF processing through sound cards with hardware acceleration to achieve spatial audio, primarily used for gaming. *See* Ex. 1064, Liang, M., *THE AweSOME 5: High-End Sound Cards Review* (1998), available at <https://assets.hardwarezone.com/2009/reviews/sound/roundup/soundcards.htm>; *see* Ex. 1065, Liang, M., *THE AweSOME (Part 2): Technologies behind ...* (1998),

available at
<https://assets.hardwarezone.com/2009/reviews/sound/roundup/sound2.htm>; Ex.
1073, DOS Days Aural Semiconductor, available at
<https://dosdays.co.uk/topics/Manufacturers/aural.php>; Ex. 1061, Collins, K. *Game
Sound: An Introduction to the History, Theory, and Practice of Video*, MIT Press
2008).

68. Publicly available processing techniques and technologies have been available since the 1990s. For example, real-time binaural processing has been available in the Max/MSP object named *Spat~* since its first version in 1995. *See* Jot, J.M., *Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony*, Proceedings of the 98th Convention of the Audio Engineering Society, Paris, (Feb. 1995) (“Jot”) (Ex. 1076) at 1-3. Max/MSP is visual programming language and toolset for music and multimedia creation.

69. In addition, over 30 years ago, the VIEW HMD discussed above included a “binaural auditory display” that was “capable of presenting a wide variety of binaural sounds to the operator via headphones using sound-synthesis technology developed for music synthesizers.” Fisher at 93; *see also* Begault at 170-174, (discussing and describing the 3D audio (i.e., binaural sound) processes and capabilities of NASA’s VIEW system).

70. Other binaural or 3D audio systems were developed in the 1980s as well. For example, U.S. Patent No. 4,774,515 (“Gehring”) (Ex. 1014), which issued in 1988, described several techniques for producing virtual auditory cues in military cockpits. *See, e.g.*, Gehring at 2:67-3:4 (“The localization achieved through spectral shaping of the sound provides true three-dimensional sound to the pilot. The pilot would know, from the position of the sound, which way to manipulate the aircraft to return the sound to the “normal” position.”), 2:30-40 (“The aural indications are generated in such a manner that to the user the sound appears to emanate from an external source, that is, one situated outside his own head to thereby provide localization. However, by placing the sound immediately adjacent the user’s ears, that is, within the user’s ‘sacred space’, urgency may be imparted to the individual as by changing the intensity or frequency of the sound such that the individual cannot ignore the three-dimensional aural signal but must deal with it in a manner to return it toward the predetermined norm”).

71. Also, in addition to describing the 3D audio capabilities of the VIEW system, the Begault reference identified above described several other convolution devices that generated and used binaural sound for military aerospace research. *See* Begault at 168 (explaining that the AL-204 was an “early DSP-based convolution device” used to provide auditory location coordinates and “spatialized audio” for flight control and other onboard systems).

2. Development of HRTFs

72. Persons of ordinary skill in the art knew and understood for years before May 2017 that binaural sound can be processed and generated “through the use of HRTFs.” Algazi, V. R., *Headphone-Based Spatial Sound*, IEEE Signal Processing Magazine, 33-42 (Jan. 2011) (“Algazi 2011”) (Ex. 1008) at 38. In fact, the use of HRTFs is typically a “key component” of a 3D or binaural sound system. Begault at 41. Indeed, I have been using HRTFs in my research and work in the industry, to generate binaural sound, since at least as early as 1996, and I also have been teaching others (e.g., my students) how to do this since 1999, while teaching a computer programming course at Northwestern University while I was a doctoral student there.

73. HRTFs themselves were described at least as early as 1980. *See* Morimoto, *On the simulation of sound localization*, 1 J. Acoust. Soc. Jpn (E) 1, 3, 167–174 (1980) (“Morimoto”) (Ex. 1034) at Abstract (“Sound localization both in the horizontal and the median planes was simulated by the aid of a digital computer through two loudspeakers located in the traverse plane. The simulation was performed by the use of the measured head-related transfer function (HRTF) for two ears”).

74. By 1989, researchers at the University of Wisconsin found that after measuring HRTFs at a large number of positions around a subject listener, recorded HRTFs could be used as digital filters to simulate sound localized to different

positions around the subject. *See* Ex. 1044 at 870 (“The source locations were chosen from a list of 144 potential positions, which were those at which each subject’s HRTFs had been measured.”), 869 (“Production of each stimulus involved passing a shaped burst of Gaussian noise, spectrally contoured according to the algorithm described above, through two digital filters, one for the left-ear stimulus, and the other for the right-ear stimulus.”).

75. Similarly, the AL-204 3D audio device Begault discusses, which was “an early four-channel DSP-based convolution device that used HRTFs,” was manufactured in 1988. Begault at 168-69.

76. HRTFs are transfer functions that represent how sound reaches the ears of a listener. *See* Ex. 1045, WILEY ELECTRICAL AND ELECTRONICS ENGINEERING DICTIONARY, 335 (2004), defining “head-related transfer function” as follows:

Its abbreviation is HRTF. A transfer function that takes into account factors involved in sound reaching the ears of a listener, for sound localization and the determination of the arriving spectrum of sound. Variables such as the dimensions of the ear canal, the exterior part of the ear, head, and torso, and how these modify the spectrum of the sound heard are combined with the azimuth, elevation, range, and frequency

of the sound source, the slightly different time at which a sound arrives at each ear, and so on. Each listener has a unique HRTF.¹

¹ See also Begault at 41 (“The spectral filtering of a sound source before it reaches the ear drum that is caused primarily by the outer ear is termed the **head-related transfer function (HRTF)**. The **binaural HRTF** (terminology for referring to both left- and right-ear HRTFs) can be thought of as a frequency-dependent amplitude and time-delay differences that result primarily from the complex shaping of the pinnae.”) (bold in original); Ex. 1072, Lalwani, *Surrounded by sound: how 3D audio hacks your brain* (February 12, 2015), available at <https://www.theverge.com/2015/2/12/8021733/3d-audio-3dio-binaural-immersive-vr-sound-times-square-new-york> (“The technique at the heart of binaural audio can be traced back to Clement Ader, a 19th-century French engineer. In 1881, Ader devised the Theatrophone, a telephonic system of transmission to broadcast a Paris Opera show. Pairs of microphones were systematically spaced in front of the stage, covering the breadth from left to right. Signals from the show were then transmitted via telephone receivers to listeners on the other end. With a pair of receivers, one mounted on each ear, listeners could hear the show from their designated suites at the gallery of Palais de l’Industrie”).

77. HRTF-based sound rendering uses techniques “where the outer ears (the pinnae) are either directly implemented or modeled as digital filters.” Begault at x. These digital filters can be designed to contain the directionally-dependent cues used by humans to locate sound.

78. The primary auditory cues used by people include: “1) the interaural time difference (ITD); 2) the interaural level difference (ILD); 3) monaural spectral cues that depend on the shape of the outer ear or pinna; 4) cues from torso reflection and diffraction; 5) the ratio of direct to reverberant energy; 6) cue changes induced by voluntary head motion; 7) familiarity with the sound source.” Algazi 2011 at 34.

79. The composite of the ITD, ILD and the spectral coloration characteristics are captured in Head-Related Transfer Functions (HRTF). The HRTF is the frequency-domain representation of the Head-Related Impulse Response (HRIR). Due to physical differences between individuals, HRTFs vary in both general shapes and detail. *See, e.g.,* Middlebrooks, J.C., *Individual Differences in External Ear Transfer Functions Reduced by Scaling in Frequency*, J. Acoust. Soc. Am., 106:1480-1492 (1999) (Ex. 1080) (“Middlebrooks”); Moller, H, *Head-related transfer function of human subjects*, J. Audio Eng. Soc., 43 (5) (1995) (Ex. 1077) (“Moller”); Shaw, E.A.G., *External ear response and sound localization*, In *Localization of Sound: Theory and Applications*, pp. 30-41 (1974) (“Shaw”); Wenzel,

E.M. *Localization in virtual acoustic displays*, Presence, 1:80-107 (1992) (Ex. 1081) (“Wenzel 1992”).

80. The most common way of acquiring HRTFs is through acoustic measurement. Acoustic measurements of HRTFs have been done for decades and have resulted in HRTFs databases, such as CIPIC. HRTFs are measured by playing a flat-frequency spectrum signal (such as a sine-sweep, maximum length signal (MLS), Golay code, etc.) through a loudspeaker at different locations around a person’s head and recording the signal at the person’s left and right ear. *See* Begault, D.R, *Design and verification of HeadZap, a semi-automated HRIR measurement system*, Proceedings of the 120th Audio Engineering Society Convention (May 2006) (Ex. 1083) (“HeadZap”). The result is a left- and right-ear impulse response for each location around a person’s head, or a Head-Related Impulse Response (HRIR) – which is the time-domain equivalent of the Head-Related Transfer Function (HRTF). The HRIRs are filter sets which can be used to create binaural audio through, for example, convolution processes (as described below).

81. As Algazi 2011 explains, “[e]xcept for source familiarity, all of these cues stem from the physics of sound propagation and vary with azimuth, elevation, range, and frequency.” *Id.* So by “filtering a digitized sound source with” HRTFs, “one can potentially place sounds anywhere in the virtual space about a headphone listener.” *Id.* (“[T]he HRTF alters the spectrum and timing of the input signal in a

location-dependent way that is recognized to a greater or lesser degree by a listener as a spatial cue.”); *see also* U.S. Patent No. 7,710,654 (“Ashkenazi”, Ex. 1028) at 11:61-12:8 (“By artificially introducing such delays and frequency response, a monaural signal, is transformed into a binaural signal, having spatial location characteristics. The delay and frequency response which associate a spatial audio source location with each ear are described by a Head Related Transfer Function (HRTF) model.”); Fohl, *A System-On-Chip Platform for HRTF-Based Realtime Spatial Audio Rendering with an Improved Realtime Filter Interpolation*, International Journal on Advances in Intelligent Systems, Vol. 3 No. 3-4, 309-317 (2011) (Ex. 1032) (“Fohl”) at 310 (“The human auditory system uses (at least) three binaural properties of a sound signal to determine the direction of the source: Inter-aural Intensity Differences (IID), Inter-aural Time Differences (ITD), and the angular variation of the spectral properties of the sound. . . . These three properties are completely represented by the Head Related Transfer Functions (HRTF) for given azimuth and elevation angles.”).

82. The relationship between the physical characteristics of the human head, torso, and pinnae, and the acoustic characteristics known as HRTFs have been known and studied by persons of ordinary skill in the art for decades.

83. For example, in 1907, Lord Rayleigh laid the foundation of human sound localization with his Duplex Theory, which included two fundamental cues of

localization: (i) differences in arrival times between the ears, i.e., interaural time differences (or “ITDs”) and (ii) differences in amplitude or level of the sound between the ears, interaural level differences (or “ILDs”). Begault at 31.

84. Later, moving beyond ITDs and ILDs, researchers determined that the physiological shape of the human head, torso, and especially the pinnae (i.e., how sound reflects, diffracts, and is generally filtered by the geometry of a person’s head, face, and ears before entering the ear canal) result in spectral filtering characteristics depending on the direction where the sound is coming from.

85. For example, in 1927, Fletcher and Sivian patented a “Binaural telephone system” using “transmitters are mounted in a dummy or artificial head which produces a sound shadow similar to that of a human head.” *See* U.S. Patent No. 1,624,486 (“Fletcher”) (Ex. 1021) at 1:1-9. Likewise, in 1927, W. Bartlett Jones developed a system to capture, record and reproduce “binaural” signals using an artificial head and simulated human pinnae. U.S. Patent No. 1,855,149 (“Jones”) (Ex. 1022) at 6:19-36 (explaining that “calculations for adjustment may be readily dispensed with to produce a predetermined effect. A special pick-up device may be provided and this device and the origin which it is to pick up are preferably isolated from other sounds. A pick-up device which acts as a human head with respect to the origin and the cars is employed, and this is preferably made as a dummy head with microphonic ears. By moving such a device relatively to the origin the desired effect

will be obtained. In Fig. 8 there is shown a plan view of a dummy head 47 with two microphones 48- and 49 serving as substitutes for ears. The microphone-ears are placed upon the head in positions and in angular relations comparable to human ears on their head.”), 6:85-87 (“For phonographic reproduction two records may be made, one for the right ear effect and one for the left ear effect.”).

86. And in 1931, Fletcher developed a more sophisticated binaural recording device, nicknamed “Oscar,” at Bell Labs. *See Paul, Binaural Recording Technology: A Historical Review and Possible Future Developments*, *Acta Acustica* united with *Acustica* 95, 767-788 (2009) (Ex. 1039) (“Paul”) at 769-770 (“Based on the knowledge gathered with ‘Oscar’ [Fletcher] devised a system for “auditory perspective” intended to be used in cinemas, as “auditory perspective” was a topic of considerable interest in around 1900.”).

3. HRTF Databases

87. Well before the priority date of the ’836 patent, and as the above references indicate, it was “well known that HRTFs vary significantly from person to person.” Algazi at 99. However, because in many cases “the taking of individual HRTF impulse response measurements is impractical,” persons of ordinary skill in the art knew and understood that it can be “better to use measurements that represent the best compromise between the HRTF features of a given population.” Begault at

116 (“In most 3-D sound applications some type of generalized HRTF set will be necessary.”).

88. Thus, as Begault explains, designers of 3D sound systems in most cases used HRTFs from “a library of HRTF measurements from which particular sets c[ould] be selected on the basis of either user preference or behavioral data.” Begault at 112. Such libraries or databases were well-known and frequently used by persons of ordinary skill in the art to process sound for images with HRTFs before May 2017. *See, e.g.*, Algazi at 99 (“several studies have employed the public-domain dataset for the KEMAR mannequin”), *id.* (disclosing, in 2011, that “[t]he CIPIC Interface Laboratory at U.C. Davis ha[d] measured HRTFs at high spatial resolution for more than 90 subjects”), 102 (noting the “public availability of the CIPIC HRTF database”). Indeed, “in most cases, nonindividualized (generalized) HRTFs [we]re used.” Begault at 45; *see also* Algazi at 102 (“A uniform database of HRTFs enables the study of person-to-person differences and of the relation of temporal and spectral characteristics of the HRTF to the anthropometric data.”).

89. In addition to the CIPIC database (measured at 1m distance), other public HRTF measurement databases and libraries have been readily available and would have been known by a person of ordinary skill. These were measured in the “far-field” – i.e. at a distance of 1 meter or greater. A few examples of standard, public HRTF libraries include the KEMAR, LISTEN, MARL, and FIU databases.

See Gardner, B., HRTF Measurements of a KEMAR Dummy-Head Microphone, MIT Media Lab Perceptual Computing - Technical Report (1995) (measured at 1.4 meters) (Ex. 1084); Warusfel, O., LISTEN HRTF Database (2003), available at <http://recherche.ircam.fr/equipes/salles/listen/index.html>; Andreopoulou A., *Towards the creation of a standardized HRTF repository*, Proceedings of the 131st Audio Engineering Society Convention (2011) (Ex. 1059), (“Andreopoulou”); Gupta, N., *HRTF database at FIU DSP lab*. International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 169–172. IEEE (2010) (Ex. 1079) (“Gupta”); see also Wenzel 1988b at 87 (describing NASA’s VIEW system and its use of HRTFs measured at 1.4 meters).

90. I further note that most existing HRTF libraries and databases contained and still contain HRTFs that were measured at what the ’836 patent calls “far-field” distances, i.e., distances greater than or equal to 1 m. See, e.g., Romblom, D., *Near-Field Compensation for HRTF Processing*, Proceedings of the 125th Convention of the Audio Engineering Society (2008) (Ex. 1040) (“Romblom”) at 1 (“When binaural synthesis is chosen to simulate 3D sound sources over headphones, the Head-Related Transfer Functions (HRTF) used have most often been measured at a single distance in the far-field of the listener”). Such HRTFs were frequently captured and used by persons of ordinary skill in the art because, as opposed to near-field HRTFs, “far-field” HRTFs are “essentially independent of range,” (i.e. distance), and the “far-field

range dependence is easily obtained” with basic computations, i.e. “merely by adding the propagation delay and the inverse range dependence.” Algazi 2011 at 35. Further, sound processing is “more complicated” with “near-field” HRTFs and when it is “close to the head.” *Id.* It is more complicated because once sounds are close to the head, “the HRTF changes substantially with distance at distances less than 1 m.” Brungart, *Auditory localization of nearby sources. Head-related transfer functions*, J. Acoust. Soc Am 106, 1465-1479 (1999) (“Brungart”) (Ex. 1036) at 1465. This is in direct contrast to far-field HRTFs that are “virtually independent of distance for sources beyond 1 m.” *Id.*

91. Thus, it was well-known before May 2017 that a far-field HRTF could readily and advantageously be used for sounds of various distances, and one can then alter other cues, e.g., “turn[] up the volume,” to adjust user perception. Romblo at 1 (“it is possible to compensate the existing data using the well known law of ‘1/r spreading roll off’”); *see also* Brungart at 1465 (explaining that far-field HRTFs are “virtually independent of distance”); Algazi 2011 at 35 (explaining that for far-field HRTFs, the “inverse Fourier transform of the HRTF is the head-related impulse response (HRIR). If $h(t)$ is the head-related impulse response for a distant source and c is the speed of sound, then the anechoic pressure response to an impulsive velocity source at a distance r is proportional to” $1/r$). It was also known that for near-field sounds within one meter, the same $1/r$ spreading roll off could be used along with

“complementary methods for simulating near-field sources using common far-field HRTFs,” such as “using a ‘difference filter’ that compensates for spectral differences between far-field and near-field sources” or using a “cross-ear” lookup model to “select the correct HRTF position.” Romblom at 2; *see also* Brungart at 1470 (explaining that “[t]he high-frequency response of the HRTF is generally lower relative to the low-frequency response of the HRTF when the source is close than when the source is more distant. ... The high-frequency response is generally 4–6 dB lower relative to the low-frequency response when the source is at 0.12 m than when the source is at 1 m (Fig. 4). This relationship implies that a sound source at a fixed angle relative to the head will appear to be low-pass filtered as the source approaches the head.”)

92. So, to be clear, well before the earliest alleged priority date of the ’836 patent (May 24, 2017), processing sound with HRTFs—including to render binaural sound to particular locations in virtual environments corresponding to virtual images, and including with “far-field” HRTFs for “near-field” sources—was well-known to persons of ordinary skill in the art. *See, e.g.*, THE VR BOOK at 240 (“Multiple HRTFs from different directions can be interpolated to create an HRTF from any direction relative to the ear ... resulting in a realistic spatialized audio cue to the user.”); Begault at 41 (“The use of HRTF spectral shaping is typically featured as the key component of a 3-D sound system.”), 105-53 (describing in detail methods of

processing sound with HRTFs to generate binaural audio for virtual environments—“3D Sound for Virtual Reality and Multimedia”), 165-66 (discussing NASA’s VIEW system and its spatialized audio features); Wenzel, E.M., *Development of a three-dimensional auditory display system*. SIGCHI Bulletin, 20, 52-57 (1988) (“Wenzel 1988a”) (Ex. 1042) (describing NASA’s VIEW system and its spatialized audio features); Wenzel, E.M., *Localization of nonindividualized virtual acoustic display cues*, Proceedings of the CHI’91 ACM Conference on Computer-Human Interaction. New Orleans, 351-359 (1991) (“Wenzel 1991”) (Ex. 1043) at 359 (discussing the “[s]ynthesis of virtual sources [that] involves the digital filtering of stimuli using filters based on acoustic Head-Related Transfer Functions (HRTFs) measured in human ear canals); *see also* Burgess D. A. , *Techniques for low cost spatial audio*, in ACM Symposium on User Interface Software and Technology, pp. 5359 (1992) (Ex. 1078) (“Burgess”) at 55 (“Once measured in this manner, the HRTF can be expressed as a set of convolution operators, or finite impulse response (FIR) filters. For each position (azimuth, elevation), there is a pair of filters: one for each ear. To place a sound in a given direction, simply apply the corresponding filter pair to the soundstream. The filters are typically measured at 10 to 20-degree intervals, and filters for intermediate angles can be bilinearly interpolated from the empirical set (Wenzel, et al, 1988)”).

93. The following additional references further demonstrate and confirm how well-known sound processing with HRTFs was, including for generating binaural audio in implementations using AR and VR HMDs:

- Zotkin, *Customizable Auditory Displays*, Proc. 2002 Int’l Conf. on Auditory Display, 1-20 (2002) (“Zotkin”) (Ex. 1046) at 6-9 (describing an HRTF audio rendering process for AR and VR HMDs with head tracking that “was capable of making people think that the sound was coming from the external source, even though it was being rendered at the headphones,” which ran “in real-time on a typical office PC” and processed “multiple virtual sound sources at different positions” with HRTFs in parallel and “no noticeable latency”).
- U.S. Patent No. 9,787,846 (“Srinivasan”) (Ex. 1024) at 2:3-23, 5:34-67, 6:31-60, 11:48-12:5 (describing an HMD for mixed reality applications with an “audio position processor” that could be “configured to apply a head related transfer function (HRTF) using []relative position/orientation information as an input parameter to generate multiple audio output channels from a mono audio input channel”), Figs. 3, 7 (depicting same).
- U.S. Patent No. 9,767,618 (“Najaf-Zadeh”) (Ex. 1018) at 4:41-65 (describing HMD with “circuitry [] and applications for providing 3D audio” processed using HRTFs), Fig. 4.

- U.S. Patent No. 10,509,466 (“Miller”) (Ex. 1017) at 23:48-59 (describing HMD that “spatially render[s]” audio “to make the user perceive various virtual audio sources as located in various x,y,z locations in world space,” which used “a software algorithm that controls both the audio output as well as head-tracking inputs and a head-related transfer function (HRTF) to simulate acoustics”);
- U.S. Patent Pub. No. 2012/0183161 (“Agevik”) (Ex. 1025) at [0044] (describing an HMD that could “apply [a] selected HRTF to an audio signal (e.g., from an audio player, radio, etc.) to generate an output signal” for binaural sound).

4. Processing Binaural Audio

94. Numerous methods of processing binaural audio using HRTFs would have been known to persons of ordinary skill in the art before May 2017, with which persons of ordinary skill in the art would have understood and known how to generate and process binaural audio.

95. An HRTF is the frequency domain representation of an impulse response that is acoustically measured (or simulated) between a sound source and the listener’s ears. These impulse responses, as represented in an HRTF, contain all the acoustic properties of a sound that is located in space. As explained above, the HRTF is the directionally dependent acoustic fingerprint between a sound source and a

listener. The left and right-ear HRTF for a specific location contains all the information a listener needs to *decode* the location of a sound at that location. All this information is contained in the HRTF, and more specifically, in an HRTF's audio cues: ITD, ILD and spectral coloration. *See, e.g.,* Begault Chpt. 2.

96. An HRTF can be applied to an audio signal through convolution processes. In general, convolution is a mathematical operation where two functions are combined to produce a third function, and in signal processing, convolution processes involve the combination of two signals to produce a third signal. Binaural audio can be generated through processes that convolve the left- and right-ear Head-Related Impulse Response (time domain representation of the Head-Related Transfer Function) of a specific location with a sound, to create a binaural sound at that location. For example, to create a binaural sound of a dog barking to the right of a listener, persons of ordinary skill in the art would understand that one could take a mono sound of a dog bark and apply a convolution process to convolve it with left and right-ear impulse response measured at 90 degrees.² As Begault summarizes:

“What filtering does is multiply the spectra of two waveforms, which is equivalent to the convolution of the time domain representations of the

² Convolution in the time domain is equivalent to multiplication in the frequency domain.

waveforms. Stated mathematically, $X(z) \bullet H(z)$ in the frequency domain is equivalent to $x(n) * h(n)$ in the time domain (where $*$ symbolizes convolution). This is exactly what the HRTF in natural spatial hearing does; in effect, the spectra of the pinnae is multiplied with the spectra of the sound source.”

Begault at 106.

97. A sound that has been convolved with a left/right HRTF effectively represents a signal which is acoustically equivalent to a sound that would be placed at the location of the HRTF, in azimuth, elevation and distance. Using an HRTF that was measured in the far field (at a distance of 1 meter or greater), and convolving it with a sound, would result in the binaural sound being placed virtually in space in the far field, at the location of the HRTF. Thus, it would have been obvious to a person of ordinary skill in the art that convolving a sound with an HRTF measured at 1 m (or greater) would result in a perceived “far field” binaural sound.

98. Furthermore, most publicly available HRTF databases in 2017 were measured at a distances of at least 1 m, i.e. in the far field. *See, e.g.*, Algazi (HRTFs measured at 1 meter); Gardner (HRTFs measured at 1.4 meters); Warusfel, (describing the LISTEN HRTF database); Gupta (describing the HRTF database at FIU DSP lab). Any person of ordinary skill using such databases would process signals and render binaural signals resulting in sounds in the far field.

5. Distance perception

99. Distance perception has been studied and documented since the 1970s. Studies have concluded that acoustical factors such as direct-to-reverberant ratio, spectrum, and intensity are principal factors when identifying distance. *See generally* Ex. 1085, Howard, D., Angus, J. (2006) *Acoustics and Psychoacoustics*, Focal Press; *see also* Begault at ch. 3. In anechoic conditions (without reverberation), most notable articles documenting distance perception of live and recorded sources have focused on distance perception starting at a distance of 3 feet. *See, e.g.*, Begault at 76 (citing Gardner, M B, *Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space*, 45 47-53 *Journal of the Acoustical Society of America* (1969), which included experiments at locations of 10 feet to 30 feet).

100. An important study on distance perception was published in 1979 by Merson and Bowers where they presented sounds through loudspeakers located at 55cm, 100cm, 200cm, 400cm, and 800cm. Ex. 1063, Mershon, D., *Absolute and relative cues for the auditory perception of egocentric distance*, *Perception*, vol. 8, pp. 311-322 (1979) at 315-317. They observed that “The closer presentations produced overestimations in median distance, which shifted to underestimations by the 800 cm distance. *Id.* at @@”. The results of individual observers were quite consistent with this pattern; thus, for the first presentation, 78% of forty observers reported a distance greater than 55 cm for the nearest speaker and 82% of another

forty observers reported a distance less than 800 cm for the farthest speaker.” *Id.* at 315-317. Their results showed that for a source presented in the near-field at 55cm, the perceived distance was at 101cm, leading to an over-estimation of the distance of the sound source. *Id.* at @@.

101. Corroborating the above results was a study by Zahorik, where binaural sounds were presented over headphones at twelve (12) distances of 30cm, 43cm, 61cm, 86cm, 122cm, 172cm, 244cm, 345cm, 488cm, 690cm, 975cm, and 1379cm. *See Ex. 1066, Zahorik, Assessing auditory distance perception using virtual acoustics, J. Acoust. Soc. Am., vol. 111, pp. 1832-1846 (2002).* Zahorik’s results proved “that far physical distances are underestimated and near distances are overestimated are consistent with an influence of a perceptual organization factor known as *specific distance tendency*...” and that “[t]his factor describes a tendency towards a specific ‘default’ value of perceived distance under conditions where all distance cues are removed.” This describes the exact condition when sounds are processed using HRTFs, and distance cues such as direct-to-reverberant sound ratio, intensity are removed. Zahorik continues that “the value of specific distance tendency is estimated to be approximately 1.5m.” These studies state that listeners have a tendency to localize any sound source that is in the near field to a location of 1.5m.

102. A person of ordinary skill would have understood the background of localization and distance perception, and would have understood the specific distance

tendency as taught by Zahorik. Studies by Zahorik are part of the academic curriculum and readings in classes related to spatial audio, such as the 3D Audio class I have been teaching at NYU since 2007.

C. Combining 3D Sound With Images

103. With respect to AR and VR devices, persons of ordinary skill in the art also knew and understood before May 2017 that auditory distance cues and a listener's ability to accurately localize binaural sound could be impacted by visual stimuli. *See, e.g.*, Zhou at 1001-02 (demonstrating that 3D audio “led to 2.5 times as accurate” of listeners correctly determining “relative depth judgments” and “significantly enhanced” listeners’ ability to distinguish the relative depth of two competing visual stimuli). This is true even in single camera AR systems (as opposed to “stereo video images”) where a “depth confusion problem” may occur when two virtual objects of different sizes are placed at different distances (i.e. a situation where a larger virtual object placed at a farther distance would appear to be the same size on the screen as a smaller object placed closer to the user). *Id.* at 995-996, Fig. 2.

104. As discussed above, the '836 patent aims to “trick[] the listener into believing that [a] sound originates from [a] sound source when in fact the sound is convolved to originate from a location that is behind the sound source” and its associated image. '836 patent at 4:65-5:11; Section IV.A ('836 patent overview); *see also* Begault at 69 (discussing this effect, in 2000, and stating that, at that time,

“[w]hat would be particularly relevant to virtual reality applications” are “studies of the interaction between visual objects and aural distance cues in a virtual world”); Finnegan, D, *Compensating for Distance Compression in Audiovisual Virtual Environments Using Incongruence*, In Proceedings of ACM CHI conference on Human Factors in Computing Systems (CHI’16), 1-13 (2016) (“Finnegan”) (Ex. 1005) at 11 (describing a study evaluating incongruent visual and audio stimuli and explaining that the study’s results show that in a virtual environment, e.g., in a virtual conference setting, “speech could be rendered incongruently to [] respective 3D avatars” “to compensate for the ventriloquist effect”) (citing additional reference discussing this effect).

105. Similarly and relatedly, it well was well known before May 2017 that “[c]ompression,” “a phenomenon where humans tend to underestimate the distance between themselves and target objects,” occurs in virtual environments. Finnegan at Abstract. “Distance compression is a multisensory phenomenon, where both audio and visual stimuli are often compressed with respect to their distances from the observer.” *Id.*; see also Turner, *Can the perception of depth in stereoscopic images be influenced by 3D sound?*, Proceedings of SPIE – The International Society for Optical Engineering, 1-10 (February 2011) (“Turner”) (Ex. 1041) at Abstract (explaining that humans “can hear small differences in forward depth between sound sources” and that “it could be practical to extend the apparent depth in a stereoscopic

image by using 3D sound” as a “controlled way to compensate for the depth budge limits on 3D displays.”).

106. Turner found that “a 0.25 meter difference between the image stimulus depth and the auditory stimulus depth led participants to report they were seeing the stimulus closer than in the case when the image stimulus and auditory stimulus were at the same depth.” Turner at 11.

107. Finnegan likewise explained that many previous publications “highlighted a perceived compression of distance” in virtual environments. Finnegan at 2. This is what prompted Finnegan to further explore incongruent multisensory environments, e.g., environments where audio and visual stimuli are placed at different distances, as a way to compensate for distance compression. Finnegan at 11.

1. User Position and Orientation Tracking

108. For years before May 2017, persons of ordinary skill in the art knew that movement of AR and VR HMDs could and should be detected and tracked, to keep track of a user’s position and to better manage binaural sounds in AR and VR environments.

109. In fact, “head tracking” is a “core element[] for VR experiences.” THE VR BOOK at 263. Persons of ordinary skill in the art knew before May 2017 that, to create appropriate imagery of a virtual environment and to properly spatialize

associated binaural sound as a user wearing an HMD moves, an AR or VR system must sense, detect, and track a user's head positioning. *See* THE VR BOOK at 99 (“Auditory perception is quite complex and is affected by head pose”), 156 (“[u]se binaural cues with head pose to give a sense of sound location”), 189 (“rendering needs to be tightly coupled to head pose with minimal delay”), 318 (“head tracking must be accurate, precise, fast, and well calibrated for the virtual world to appear stable”).

110. As Begault taught in 2000, “[i]f one of the perceptual requirements is to have virtual sound images remain fixed in virtual space independent of where one's head moves, a six degree-of-freedom” (or “6DoF”) “tracker system is necessary, consisting of a source, sensor, and a hardware unit housing power supply, electronics, and RS-232 interface to send positional data to a host computer.” Begault at 133.

111. Figure 4.37 in Begault depicts how “[p]ositional data about the user's position in relationship to [a] source can [] be obtained in terms of pitch, yaw, and roll, and x, y, and z,” so that a “3-D audio subsystem then can compensate for the relative position of source and sensor to give the illusion of a sound that remains fixed in space independent of head movement”:

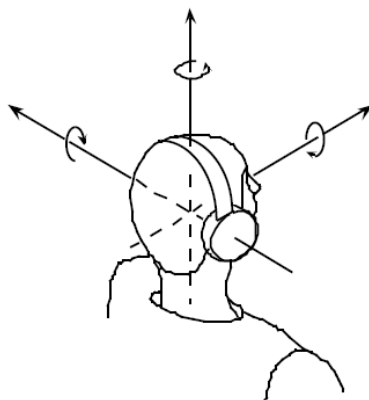


Figure 4.37. Coordinates for pitch-yaw-roll and x, y, z in a head-tracked system. *Courtesy of Crystal River Engineering.*

Id. at 133, Fig. 4.37.

112. Figure 4.38 in Begault “shows the method used by a [then] commercially available 3-D sound system for real-time placement of a virtual source to a fixed position, independent of the listener’s movement”:

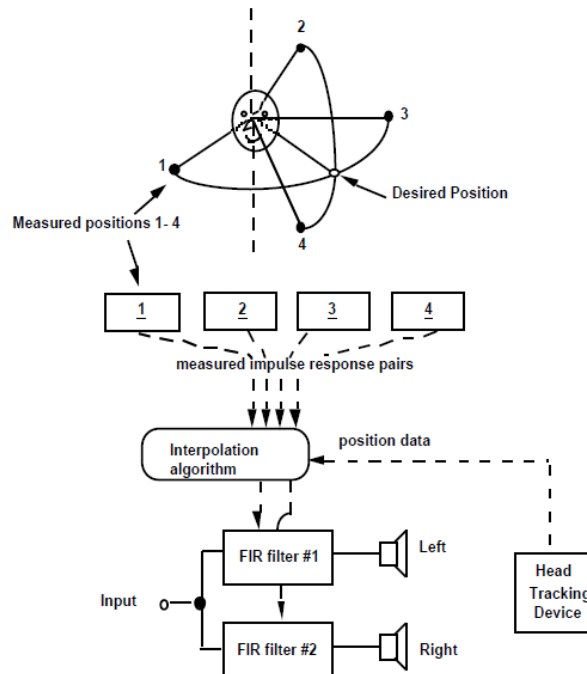


Figure 4.38. Interpolation procedure used to compensate for head movement of the listener. To simulate moving sound sources, the position data is obtained algorithmically instead of from a head tracker.

Id. at 133, Fig. 4.38. As Begault explains: “The head tracking system sends positional information to the host computer dedicated to the 3-D sound subsystem. An interpolation algorithm constantly checks this position against the two nearest azimuth and elevation positions in memory. The HRTF used at any particular moment is then the interpolation between these prestored values.” *Id.* at 133.

113. By 2015, major players in the AR and VR industry had described numerous ways of tracking movement of an HMD and peripherals in six degrees of freedom (“6DoF”), i.e., tracking movement of the HMD and controllers in three-

dimensional space, and having related image and audio processing take place on an HMD.

114. For example, U.S. Patent Pub. No. 2017/0115488 (“Ambrus”) (Ex. 1027) describes a VR HMD that includes a “head tracking system that utilizes one or more motion sensors, such as sensors 708 on HMD device 700, to capture pose data and thereby enable position tracking, direction and orientation sensing, and/or motion detection of the user’s head and/or HMD device 700.” Ambrus, [0074]-[0078], [0080]-[0082] (tracking head movements to determine head orientation and gaze direction).

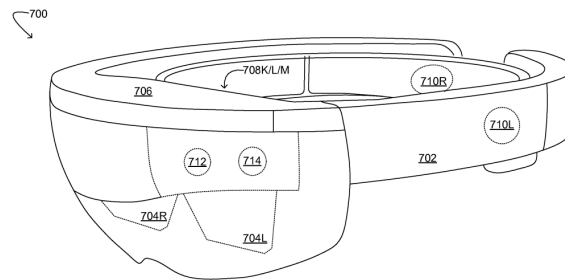
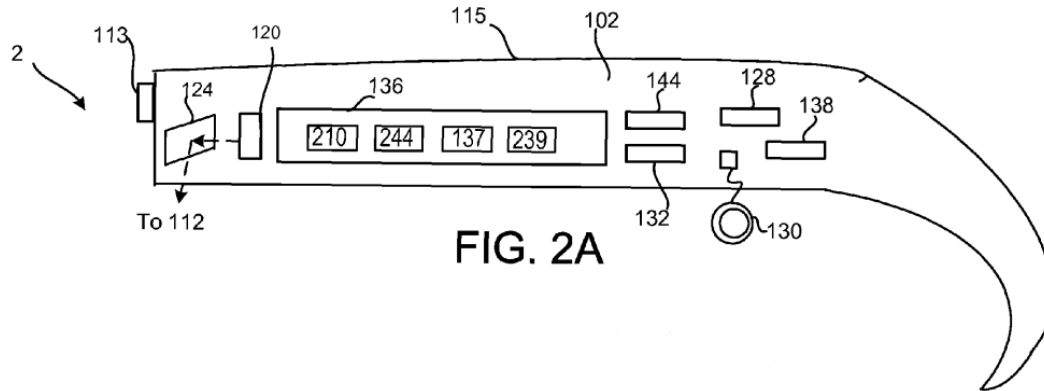


FIG. 7

Id. at Fig. 7; *see also id.* at [0083] (“The HMD device may include a microphone system ... that captures audio data. In other examples, audio may be presented to the user via one or more speakers ... on the HMD device 700.”).

115. Similarly, U.S. Patent No. 9,041,622 (“McCulloch”) (Ex. 1016) described “inertial sensors . . . for sensing position, orientation, and sudden accelerations of head mounted display device 2.” McCulloch at 8:17-39.



Id., Fig. 2A; *see also id.* at 5:4-6 (explaining that “[t]he frame 115 includes ... a microphone 110 for recording sounds and transmitting audio data to control circuitry 136.”); 14:12-22 (“The computing environment 54 also stores data in image and audio data buffer(s) 199. The buffers provide memory for receiving image data captured from the outward facing capture devices 113, image data captured by other capture devices if available, image data from an eye tracking camera of an eye tracking system 134 if used, buffers for holding image data of virtual objects to be displayed by the image generation units 120, and buffers for both input and output audio data like sounds captured from the user via microphone 110 and sound effects for an application from the 3D audio engine 304 to be output to the user via earphones 130.”); 14:59-67 (“A sound to be played for the user in accordance with the application 166 can be uploaded to a sound library 312 and identified to the 3D audio engine 304 with data identifying from which direction or position to make the sound seem to come from. The device data 198 makes available to the application 166

location data, head position data, data identifying an orientation with respect to the ground and other data from sensing units of the display device 2.”).

D. Processors Used For Rendering Binaural Sound

116. Processors that could be used for processing and rendering binaural or spatial audio was well-known to persons of ordinary skill in the art before May 2017.

117. Persons of ordinary skill in the art knew that processors could be used for the sound processing I described above. *See, e.g.*, McCulloch at 13:65-67 (describing “a computing environment” comprising an “audio processing engine”); Srinivasan at 2:3-23, 5:34-67, 6:31-60, 11:48-12:5 (describing an HMD for mixed reality applications with an “audio position processor” that could be “configured to apply a head related transfer function (HRTF) using []relative position/orientation information as an input parameter to generate multiple audio output channels”).

118. In addition, skilled artisans knew and understood that binaural sound processing is “almost always best accomplished digitally, using audio DSP chips to manipulate signal sources.” Begault at 100.

119. A DSP (digital signal processor) is a specialized processor with better power performance, speed, and efficiency than a general-purpose processor, and persons of ordinary skill in the art knew that DSPs could be advantageously used to process digitally stored information such as visual and audio data. *Id.* Audio DSP chips were known to process instructions very quickly and often in real time. *Id.*

DSPs also provide flexibility and can be programmed by using powerful instruction sets and for complex operations. *Id.* (explaining that DSPs execute instructions “*very quickly*,” “[o]ne can select a DSP that is most appropriate for a specific operation or type of digital filtering,” DSPs can be programmed with a “powerful instruction set,” and DSPs can be optimized for “specific complex operation[s]”); *see also* U.S. Patent No. 9,578,440 (“Otto”) (Ex. 1020) at 13:50-53 (Explaining that its “DSP strategy” provides optimal performance in terms of directivity of the sound beam created and of the stability of the binaural rendering at higher frequencies”); Brown, *A Structural Model for Binaural Sound Synthesis*, IEEE Transactions on Speech and Audio Processing, vol. 6, no. 5, 476-488, (Sept. 1998) (“Brown”) (Ex. 1035) at Abstract (providing a model for synthesizing binaural sound” and noting that the “the model permits efficient implementation in DSP hardware, and thus facilitates real-time operation”).

120. As discussed above, in the 1980s, the Convolvotron from Crystal River Engineering (“CRE”), which could spatialize audio, became commercially available. The Convolvotron system could produce real time synthesis of complex acoustic environments and 3D sound and could spatialize four sound sources with custom HRTFs. *See* Begault at 165-67 (explaining that the Convolvotron used “four-way interpolation” to “obtain a synthetic HRTF for an arbitrary virtual acoustic position” where “signal processing can be tailored for the individual user.”). The Convolvotron

had two boards: a “Spectrum board” and a “convolution engine card.” *Id.* at 167. The spectrum board was a modified “TMS320-C25 system board” that included “a Texas Instruments TMS320 DSP chip” that “functions primarily to perform the interpolation calculations for the convolution engine card, as well as performing input data buffering and clock signal generation.” *Id.* The “convolution engine” is a “custom-made board that contains four INMOS A-100 cascadable DSPs.” *Id.* Crystal River Engineering also made a “low-cost version that spatializes two sources in real-time” and uses “a Motorola 56001 DSP rather than the A-100.” *Id.* at 168.

121. Thus, by May 2017, the use of DSPs for rendering binaural sound, including for AR and VR environments and images, was extensive. *See, e.g.*, Shivappa at 5:6-20, 27:63-28:24, 28:58-29:22 (disclosing an audio processing device using processor 106 to generate an output spatialized audio signal and also discloses that the processor can be a DSP); Sauk at 7:24-28 (disclosing a “sound environment manager” that processes sound with HRTFs that “can be implemented by” “one or more digital signal processors”); U.S. Patent No. 9,232,336 (“Nishio”) (Ex. 1019) at 10:25-33 (“For reference’s sake, the normalization processing circuit 10 generates the normalized head related transfer function HN of the cross talk component with respect to the assumed sound source direction position by performing a similar computation processing also with regard to the cross talk component and stores this in a predetermined storage unit. It should be noted that the respective signal

processing in the normalization processing circuit 10 can be performed mainly by a DSP (Digital Signal Processor)"); U.S. Patent No. 8,873,761 ("Fukui") (Ex. 1023) at 15:63-66 ("All the signal processing in the block diagram in FIG. 1 for explaining the embodiment of the measurement method of head related transfer functions can be performed in a DSP (Digital Signal Processor)"); Otto at 10:52-58 ("The DSP for the binaural mode involves the convolution of the audio signal to be reproduced with a set of digital filters representing a Head-Related Transfer Function (HRTF).").

VIII. PRIOR ART RELIED ON IN THE GROUNDS

122. In this section, I provide brief overviews of the prior art references relied upon in the grounds in the Petition and discussed herein, some of which were mentioned above in discussion of the State of the Art. Each of these references is discussed further in later sections of this declaration in the context of claims 8-17 of the '836 patent.

A. Finnegan (Ex. 1005)

123. I have been informed that Finnegan is prior art because it was published in May 2016.

124. Finnegan teaches methods for compensating for distance compression in virtual environments containing both audio and visual stimuli by using incongruent stimuli.

125. Incongruent stimuli means that images and their associated sounds are presented at different distances from a user by, for example, placing a sound source directly in front of or behind an image.

126. As Finnegan explains, “a key requirement for a sense of presence in Virtual Environments (VEs) is for a user to perceive space as naturally as possible.” Finnegan at Abstract. And it was known in the industry that “[c]ompression”—“a phenomenon where humans tend to underestimate the distance between themselves and target objects”—occurs in virtual environments. *Id.* Indeed, as Finnegan explains, “[d]istance compression is a multisensory phenomenon, where both audio and visual stimuli are often compressed with respect to their distances from the observer,” and in HMDs, there is “significant distance compression error,” posing a challenge for VR applications to “effectively simulate environments which require spatial perception similar to that of the real world.” *Id.*

127. Finnegan details an experiment “designed to capture distance estimates provided by participants within” a virtual environment, to test whether “rendering [] audio at an incongruent position further from the observer than the visual stimulus (IV)” —in other words, positioning sound behind an associated image—“will result in more accurate distance perception (DV) compared to conditions where both stimuli are at the same position (congruent conditions).” Finnegan at 6.

128. For the experiment, Finnegan used a VR HMD, specifically an Oculus Rift Development Kit 2. *Id.* And audio was rendered using a custom plugin built for the Unity Game Engine, and binaural headphones, which the user used to hear the generated 3D audio. *Id.*

129. Through the VR HMD and headphones, users were provided with perceived visual stimuli and both (i) congruent 3D sound (sound located at the same location as the visual stimuli) and (ii) incongruent 3D sound (e.g., sound behind the visual stimulus). *Id.* at 7.

130. The visual stimuli were presented at perceived distances ranging from 0.5 to 10 meters in front of the user. *Id.* The “the visual component of a target” acted as “an anchor” and the test involved “systematically positioning the audio component of the target further away from the user.” *Id.*

131. Finnegan found that “intentionally rendering the auditory and visual components of objects incongruently to one another could improve distance perception in a virtual environment.” *Id.* at 11.

132. Finnegan concluded that its findings “have implications for applications and the design of” virtual environments, and that “compensating for distance compression can inform engineers in developing tools for VE designers to enhance the mapping between the designer’s intentions and the user’s perceptions.” *Id.* at 11-12.

B. Begault (Ex. 1006)

133. I have been informed that Begault is prior art its first edition was published in 1994 by Academic Press, i.e., over 21 years before the '836 patent's earliest alleged priority date. And Begault was later published by NASA in August 2000 and added to the NASA Technical Reports Server system in 2013.

134. Begault is titled "3-D Sound for Virtual Reality and Multimedia."

135. Begault describes processes for producing "3-D sound" for VR.

136. Begault discloses that the terms "3-D sound," "binaural audio," and "spatialized sound" are "equivalent designations" in that they all "refer to techniques where the outer ears (the pinnae) are either directly implemented or modeled as digital filters." Begault at x. Begault further discloses that "[b]y filtering a digitized sound source with these filters," one can "place sounds anywhere in the virtual space about a headphone listener." *Id.* Furthermore, Begault teaches that "[t]he combination of 3-D sound within a human interface along with a system for managing acoustic input is termed a 3-D auditory display." *Id.*

137. Begault explains that "[e]arly on in the development of virtual reality, it was seen that sound should have some type of interactive, spatial character imposed on it." Begault at 12. Begault teaches that 3-D sound "improve[s] the quality and ease of interaction within a human interface, communication system, or office or home environment," and also "contribute[s] to the sense of immersivity in a 3-D

environment.” *Id.* at 13. Begault further discloses that “[o]ne might be able to work more effectively within a VR environment if actions [a]re accompanied by appropriate sounds that seemingly emit from their proper locations.” *Id.*

138. In its Chapter 2, Begault describes in detail HRTFs and localization with HRTF cues. *See id.* at Chpt. 2.

139. Begault teaches that “[t]he spectral filtering of a sound source before it reaches the ear drum that is caused primarily by the outer ear is termed the head-related transfer function (HRTF),” and “[t]he use of HRTF spectral shaping is typically featured as the key component of a 3-D sound system.” *Id.* at 41.

140. Begault explains that “binaural HRTF” is “terminology [used] for referring to both left- and right-ear HRTFs” and “can be thought of as a frequency-dependent amplitude and time-delay differences that result primarily from the complex shaping of the pinnae.” *Id.*

141. Begault further discloses that “inclusion of the binaural HRTF is fundamental for a 3-D sound system,” and Begault discusses implementation details for HRTF processing and generating binaural sound. *Id.* at 99; *see also id.* at 105-06 (“The basis of 3-D sound processing lies in imitation of the spatial cues present in natural spatial hearing, where the binaural HRTF imposes spatially dependent spectral modifications and time delays on an incoming wavefront to the hearing system. In virtual spatial hearing, it is possible to imitate these modifications using

a DSP as a digital filter. One can obtain HRTF data by several methods (described later) and then use digital filters—specifically, two different $h(n)$ s, termed a binaural impulse response—to produce a desired virtual acoustic source position.”), 118 (“A review of the literature shows that KEMAR’s HRTFs have been used many times in both 3-D sound applications and localization studies (e.g., Kendall, Martens, and Wilde, 1990).”), 139 (“Some of these methods have been worked into 3-D sound systems, where HRTF processing is applied not only to the direct sound but to the indirect sound field as well, resulting in a synthetic spatial reverberation.”).

142. Begault discloses examples of in which “distributed 3-D audio devices” were “integrated within a dedicated virtual reality application,” such as the Convolvotron 3-D audio device within NASA’S VIEW system. *Id.* at 170.³

³ See also THE VR BOOK at Fig. 2.13 (photograph of NASA’s VIEW system):



Figure 2.13 The NASA VIEW System. (Courtesy NASA/S.S. Fisher, W. Sisler, 1988)

143. Begault also discusses “Computer Workstation 3-D Audio,” where an “audio signal is brought into the computer and then spatialized,” explaining that “communication source[s] can be usefully arranged in virtual space.” *Id.* at 172. As Begault teaches, this feature “becomes particularly important in multiconversation teleconferencing,” or “spatial audio teleconferencing,” or “an audio window system.” *Id.*; *see also id.* at 173-74 (providing further detail on “audio windows”).

144. Begault further discloses that audio spatialization is not limited to teleconferencing environments, but can be used with “all types of sonic input,” which can be “directionalized to a specific location, controlled by the user.” *Id.* at 172; *see also id.* at 172-73 Fig. 5.12 (presenting and describing “a GUI for arranging [] various sound inputs in virtual auditory space in a manner configurable by each individual listener,” which are interpreted “as commands to a multiinput, 3-D sound spatialization subsystem”).

145. Begault disclosed using digital signal processors (“DSPs”) for spatializing sound and emphasized the utility of DSPs, explaining that “[t]he 3-D audio component of an integrated audio subsystem ... can contain anywhere from one to as many DSPs as can be accommodated,” and “[w]ithin a virtual reality or multimedia application, 3-D sound processing is almost always best accomplished digitally, using audio DSP chips to manipulate signal sources.” *Id.* at 100-101.

C. Norris (Ex. 1007)

146. I have been informed that Norris is prior art because it issued on December 28, 2015. I also have been informed that Norris, like the '836 patent, is assigned to Patent Owner.

147. Norris discloses methods and systems that use sound localization points (SLPs) to simulate locations in empty space from where the origins of voices occur (or are localized) during electronic calls. Norris at Abstract (“During an electronic call between two individuals, a sound localization point simulates a location in empty space from where an origin of a voice of one individual occurs for the other individual”), 2:50-52 (“Example embodiments include systems, apparatus, and methods that provide a sound localization point (SLP) to a listener”), 1:19-21, 3:1-2, Fig. 1.

148. Norris states that a “sound localization point can exist in a three-dimensional (3D) position described by an azimuth or horizontal angle, elevation or vertical angle, and distance from the listener.” *Id.* at 3:46-50. Sound can be processed, including with “HRTFs,” so “sound localization occurs for the listener to emanate from or originate at the sound localization point.” *Id.* at 4:21-29; *see also id.* at 28:24-32 (“HRTFs are functions of frequency and three spatial variables. In far field distances greater than about one meter, the HRTF attenuates inversely with range and can be measured and/or mathematically estimated. Once HRTFs are known or

calculated for discrete angular increments (such as increments of fifteen degrees (15°) in the azimuth position), HRTFs can be estimated or interpolated for other angular positions (e.g., other angular positions, θ , in the far field HRTF of $H(f, \theta, \phi)$.”). Norris explains that “if a position and a shape of the ears of a listener are known or estimated to a suitable degree of accuracy, binaural sound delivered to this listener can be altered in order to create, move, and/or adjust the sound localization point.” *Id.* at 3:55-59.

149. Norris discloses use of HRTFs from a “database of common approximate profiles,” including “known HRTFs that are measured or calculated from or tested and refined by individuals and dummies.” *Id.* at 21:47-50 (“The stored HRTFs and other localization profile data may be from previous use by the individual or copied from another individual or from a database of common approximate profiles.”), 23:49-52 (“HRTFs for individuals are stored and retrievable from memory, such as an anonymous database. These HRTFs include known HRTFs that are measured or calculated from or tested and refined by individuals and dummies.”), 24:12-21 (“Consider an example in which known HRTFs are stored for individuals in a database. These HRTFs are mapped to one or more physical attributes and/or anatomical characteristics that include, but are not limited to, a size and shape of the head, a size and a shape of a torso and shoulders, an amount and a style of hair, clothing worn (e.g., whether a person wears a hat), height, weight, etc.”).

150. Norris's systems comprise "wearable electronic devices" (or "WEDs"), which include "head-mounted displays" and "electronic glasses or eye wear" that display "augmented reality imagery" on their lenses *Id.* at 30:3-43 (describing an "electronic system" that includes a "wearable electronic device" and which can "provide images and/or video" "through the display of the wearable electronic device"), 41:48-59 (describing "wearable electronic device" as a device that "is worn on or attached to a person" and providing examples of such device devices, including "electronic watches," "head-mounted displays, electronic eyeglasses or eye wear (such as glasses in which augmented reality imagery is projected through or reflected off a surface of a lens)"), Fig. 25 (depicting WED 2510), 37:29-39 (describing Fig. 25); *see also id.* at 9:4-17 (explaining that an electronic call can be "commenced with an electronic device that includes" "a wearable electronic device (such as electronic glasses[])"), 40:59-41:9 (describing an electronic device that includes "a sound localization system" and "an imagery system," such as "an optical projection system, a virtual image display system, virtual augmented reality system, and/or a spatial augmented reality system"); *see also id.* at 42:23-32 ("The methods in accordance with example embodiments are provided as examples, and examples from one method should not be construed to limit examples from another method. Further, methods discussed within different figures can be added to or exchanged with methods in other figures. Further yet, specific numerical data values (such as specific

quantities, numbers, categories, etc.) or other specific information should be interpreted as illustrative for discussing example embodiments. Such specific information is not provided to limit example embodiments.”).

D. Algazi (Ex. 1008)

151. I have been informed that Algazi is prior art because it was published in 2001.

152. Algazi explained that “[h]ead-related transfer functions (HRTFs) capture the sound localization cues created by the scattering of incident sound waves by the body, and play a central role in spatial audio systems.” Algazi at 99. And Algazi disclosed “a public-domain database of high-spatial-resolution head-related transfer functions measured at the U.C. Davis CIPIC⁴ Interface Laboratory.” *Id.*

153. Algazi further disclosed that “[t]he CIPIC Interface Laboratory at U.S. Davis [] measured HRTFs at high spatial resolution for more than 90 subjects,” including a “a public-domain subset for 45 subjects (including KEMAR with large and with small pinnae).” *Id.* Algazi explained that all of those HRTFs “were measured with the subject seated at the center of a 1-m radius hoop whose axis was aligned with the subject’s interaural axis” and “[s]ound source location [] specified by the azimuth angle θ and elevation angle ϕ in interaural-polar coordinates.” *Id.*

⁴ “CIPIC” stands for Center for Imaging Processing and Integrated Computing.

154. The CIPIC database is one of first publicly available HRTF database measured on humans, and is one of the most widely used HRTF databases by the audio community. It has been used as part of standard HRTF repositories (Ex. 1059, Andreopoulou), and in HRTF customization research (Ex. 1062, Hwang, S., *HRIR Customization in the Median Plane Via Principal Components Analysis*, 2007; Chun, J.C., *Deep Neural Network Based HRTF Personalization Using Anthropometric Measurements*; Proceedings of the 143rd Convention of the Audio Engineering Society, 2017 (Ex. 1060)).

E. Sauk (Ex. 1009)

155. I have been informed that Sauk is prior art because it issued on January 25, 2011.

156. Sauk is entitled “Method and apparatus for creating a multi-dimensional communication space for use in a binaural audio system,” and describes that while most development efforts in binaural audio have been directed to “simulating a virtual audio environment,” binaural audio may also be used as a cue to “provide the user with understandable information regarding their surroundings.” Sauk (Ex. 1007), Title, 2:34-64, 3:30-47.

157. Sauk discloses a “multi-dimensional communication space” where sound relevance can be signified by how sound is processed, for example, as binaural, stereo, or mono sound. Sauk at Title, 11:26-40 (describing “selective filtering

techniques” which “combines a plurality of audio dimensions such as 3-D (binaural), 2-D (stereophonic) and 1-D (monophonic).

158. Sauk also describes a “method and apparatus for producing, combining, and customizing virtual sound environments to provide the user with understandable information regarding their surroundings.” *Id.* at 2:45-48.

159. Specifically, Sauk describes defining a “predetermined audio enhancement” by “selecting an audio reproduction format based on a source of enunciated data,” by, for example, selecting from “monophonic audio, stereophonic audio,” and “binaural audio environments,” and “defining a plurality of information relevance levels.” *Id.* at 3:30-38. Then, “the predetermined audio enhancement comprises selectively applying the audio reproduction format in accordance with a particular relevance level specified by the metadata.” *Id.* at 3:38-41

160. As Sauk explains, “[t]hose skilled in the art will appreciate that relevant audio information can in some instances become diluted with unwanted background sounds.” *Id.* at 10:33-35. And humans can “focus on selected voiced (or voiced originating from a particular location) even though the background voices may be louder,” a phenomenon known as the “cocktail party effect.” *Id.* at 10:35-43.

161. In accordance with these principles, Sauk describes an audio environment operating in a “Selective Relevance Mode” using “a plurality of audio dimensions such as 3-D (binaural), 2-D (stereophonic), and 1-D (monophonic).” *Id.*

at 11:25-29. Because users can perceive differences when audio is provided in different dimensions, i.e. 3D, 2D, and 1D, Sauk explains that in its selective relevance mode, “changing an audio format for sounds from binaural to monophonic or stereophonic audio signifies a different level of relevancy or importance for such audio.” *Id.* at 11:29-40. Thus, as certain sounds become less “relevant,” such as when a sound source is no longer in view, the system can “change the less relevant audio to a monophonic (1-D) or stereophonic (2-D) dimension.” *Id.* at 11:29:36.

162. Sauk further discloses that as audio processing changes from binaural to monophonic or stereophonic sound, localization cues are removed and decibel level is adjusted. *Id.* at 11:40-41.

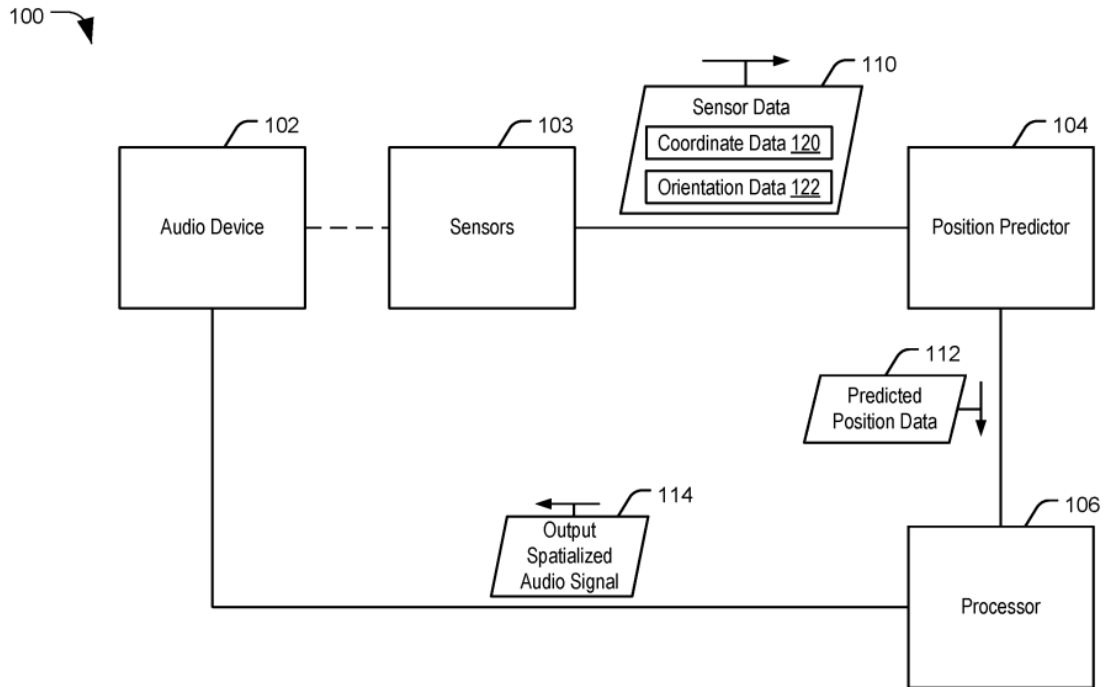
F. Shivappa (Ex. 1010)

163. I have been informed that Shivappa, which issued on April 13, 2021, is prior art because it was filed on April 8, 2016, before the '836 patent's earliest possible May 2017 priority date.

164. Shivappa discloses creating spatialized audio signals using 3D sound rendering techniques that allow a user to perceive 3D sound, including the direction and distance of a sound source. *Id.* at 5:22-36; *see also id.* at 24:49-53 (“[T]he audio processing device may be configured to process object-based 3D audio (e.g., audio having audio sources represented as individual objects with 3D coordinates and directivity)”). Those techniques include “performing binauralization” and “head

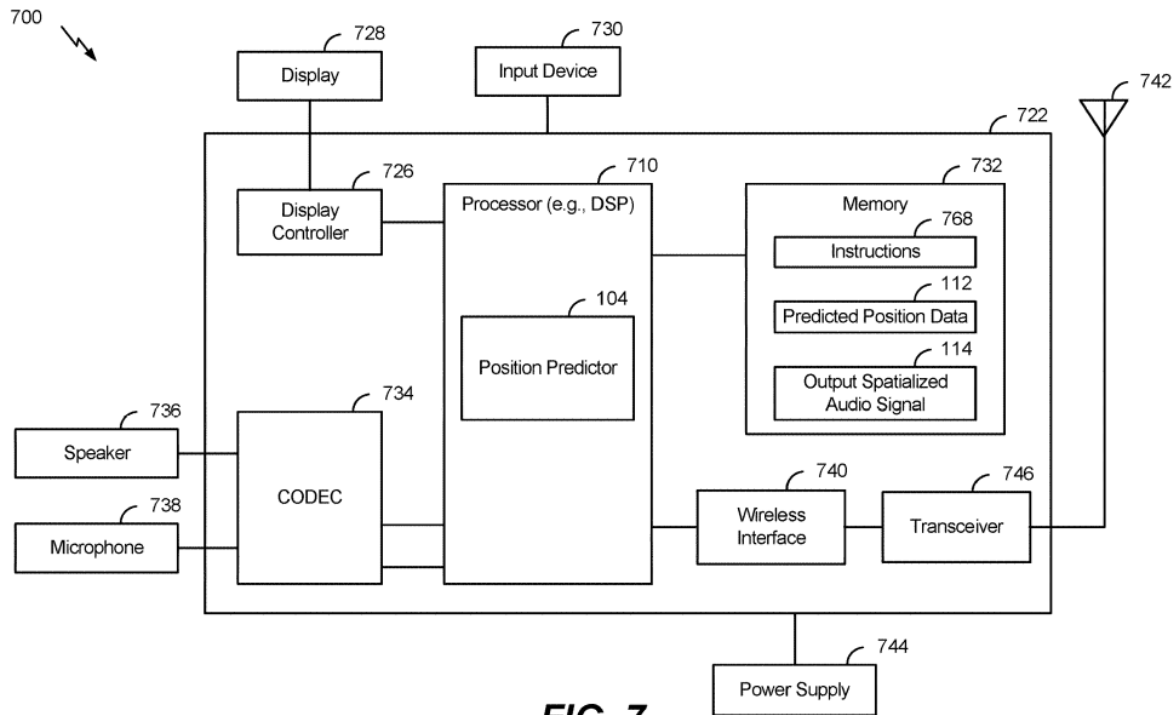
related transfer function (HRTF) filtering.” *Id.* at 7:60-8:4 (explaining that “the processor 106 (or an additional component of the audio processing device 100) may be configured to perform 3D sound rendering to generate a 3D output audio signal, and the 3D output audio signal may be used to generate the audio output at the audio device 102” and that “[p]erforming 3D sound rendering” includes “performing binauralization” and “head-related transfer function filtering”), 23:29-52 (explaining that “performing binauralization may include convolving the audio data” with “BRIRs (or head related transfer functions (HRTFS))”).

165. Shivappa discloses an audio processing device 100 that creates spatialized audio using processor 106 based on predicted position data of the user or audio device 102 and uses sensors 103 for detecting changes in those positions:

**FIG. 1A**

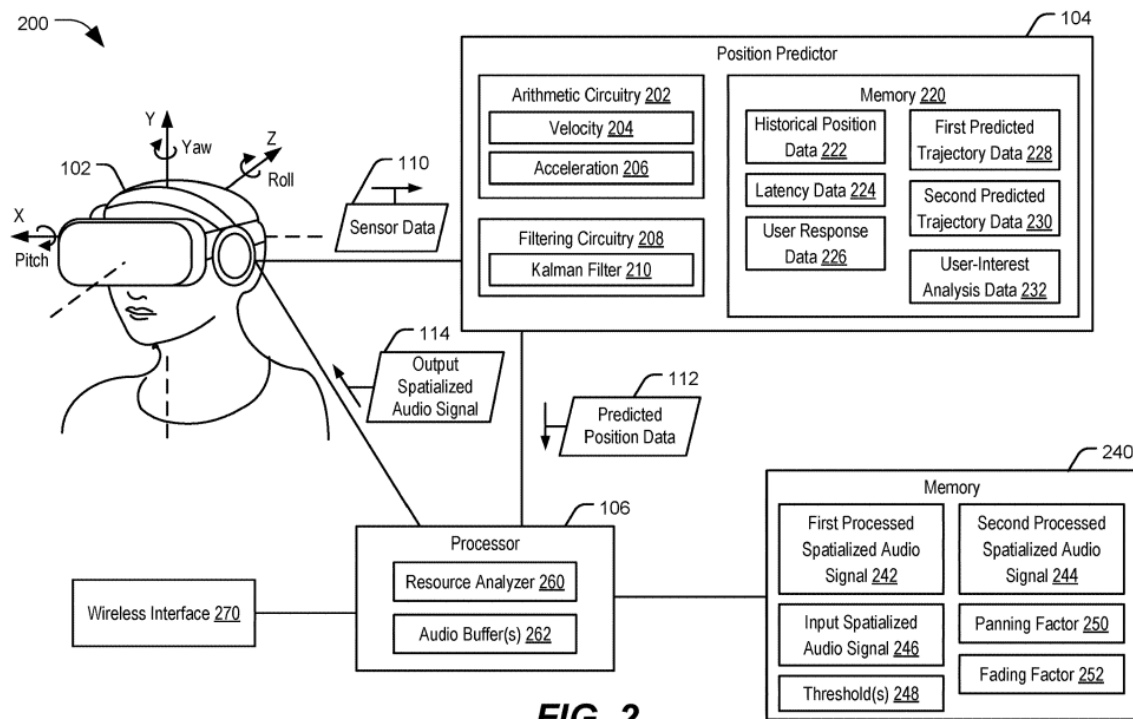
Id. at Fig. 1A, 5:6-20 (“audio processing device 100” includes “audio devices 102,” “sensors 103, a position predictor 104, and a processor 106” and is “configured to generate an output spatialized audio signal based on the predicted position data.”).

166. Shivappa’s Figure 7 depicts a device 700 with a processor 710 such as a “DSP” (i.e., a digital signal processor) that processes position data and creates spatialized audio signals based on predicted position data of a user or audio device:

**FIG. 7**

Id. at Fig. 7, 27:63-28:24, 28:58-29:22. As shown, this device also includes a display 728 for showing video and images, and a speaker 736. *Id.* at 28:10-57 (detailing components of the device shown in Fig. 7).

167. Shivappa discloses and shows that its audio device can be “integrated with a head mounted display (HMD)” of a VR or AR system:

**FIG. 2**

Id. at Fig. 2; *see also id.* at 5:37-45 (“audio device 102 includes (or is integrated within) a head mounted display (HMD) or a virtual reality (VR) system or an augmented reality (AR) system” where the VR system may “display media content, such as movies” or “provide interactive content, such as video games, presentations, virtual meetings, etc.”), 5:46-6:3.

168. Shivappa further explains that the “visuals and the audio output by the virtual reality system may change based on the movement of the user.” *Id.* at 5:45-48. “For example, if a user rotates his or her head to the side, the visuals output by the virtual reality system may change to represent a view of the side, and the audio output by the virtual reality system may change such that sounds that previously were

perceived as emanating from the side appear to emanate from a forward direction after the rotation.” *Id.* at 5:48-54.

169. Sensors on the HMD determine the location and orientation of the audio device, and by extension, the user. *Id.* at 6:51-55. These sensors can include “an accelerometer, a gyro sensor, an orientation sensor, a linear position sensor, a proximity sensor, a motion sensor, an angular position sensor, a global positioning system (GPS) sensor, an ultrasound sensor, or any other sensor(s) capable of determining a translational position (e.g., a location in a coordinate space, such as x-y-z coordinates), an orientation (e.g., pitch, yaw, and roll angles, as further described with reference to FIG. 2), or both.” *Id.* at 6:37-40.

170. Then, the processor processes an input spatialized audio signal to create a spatialized audio signal output so that, if the user’s head moves or rotates or if the user moves in a virtual environment, processor 106 changes the spatialized audio so the perceived direction or distance of one or more sound sources in the 3D space changes based on that movement. *Id.* at 8:28-42; *see also id.* at 6:7-19 (“For example, the user may be listening to a concert and may perceive a flute to be playing to the left. If the user turns to the left, the sound may change such that the user perceives the flute to be playing from in front of the user.”).

G. Zhou (Ex. 1011)

171. I have been informed that Zhou is prior art because it was published in 2004.

172. Zhou describes a study on the effects of 3D sound in an AR context. Zhou Abstract (describing the goals of the study as, among others, “(1) to access the impact of 3D sound on depth perception in a single-camera AR environment, (2) to study the impact of 3D sound on task performance and the feeling of ‘human presence and collaboration’, (3) to better understand the role of 3D sound in human-computer and human–human interactions”); *see also id.* at 990 (“3D sound was reported to increase the immersive experience in virtual reality (VR) environments and hence provide a more realistic virtual environment).

173. Zhou discloses that the “[e]xact spatial 3D sound in virtual and augmented environments requires perceptual externalization of the sound source, which can only be brought about by using head related transfer functions (HRTFs). *Id.* at 991.

174. Zhou further discloses that software development kits (SDKs), “which have already incorporated the general HRTFs database and simple model of environment to generate 3D sound” can be used to generate 3D sound. *Id.* And Zhou teaches that, to do this, an SDK must “update the relative position of [a] virtual sound source to [a] head co-ordinate system.” *Id.*

175. Zhou studied “the impact of 3D sound on improving depth perception in AR systems” and a user’s “perceptual feeling to [an] exaggeration” of 3D sound at various depths. *Id.* at 992 (“We investigated the impact of 3D sound on improving depth perception in AR systems with single-camera settings” and “We investigated the subject’s perceptual feeling to [an] exaggeration” and “the impact of 3D sound” on “locating correct augmented information.”).

176. Zhou’s experiment used “a head mounted display (HMD) through which the user view[ed] the real world, either directly, or via a video camera attached to the front of the display” with “[v]irtual objects ... superimposed on the real world.” *Id.* at 993-94. An “OpenAL (Loke entertainment Software” was used “to generate the 3D sound. *Id.* at 994.⁵ “[A]ugmented virtual image[s] w[ere] viewed through a Daeyang Cy-Visor DH-440VP HMD,” which presented “the same 640 X 480 pixel image to both eyes,” and the “headphone of the HMD play[ed] back the 3D sound

⁵ See also Doerr, *Evaluation of a Low-Cost 3D Sound System for Immersive Virtual Reality Training Systems*, IEEE Transactions on Visualization and Computer Graphics, Vol. 13, No. 2, 204-212 (March/April 2007) (Ex. 1037) at 205 (describing the open-source OpenAL library, an HRTF software-based 3D sound system for “consumer-level 3D audio hardware interfaces” for immersive VR systems).

generated by the computer.” *Id.* Both visual and audio sources were updated “25 times per second” *Id.*

177. Zhou used two experiments to test the impact on 3D sound. In the first experiment, Zhou used an “image-processing technology to retrieve 3D position information from a fiducial marker” to “track the position” of a “virtual sound source which is attached” to the marker. *Id.* at 994. Zhou explains that “[t]he 3D position of the marker relative to [the] camera can be obtained by computing the TCM from the camera co-ordinates to the 7 marker co-ordinates. However, the 3D position of the marker we obtain is based on the camera co-ordinates. Thus, in order to obtain the 3D position of the marker in user’s head co-ordinates (3D sound synthesizer’s co-ordinates), simple translations from the camera to the center of user’s inter-aural axis should be considered.” *Id.*

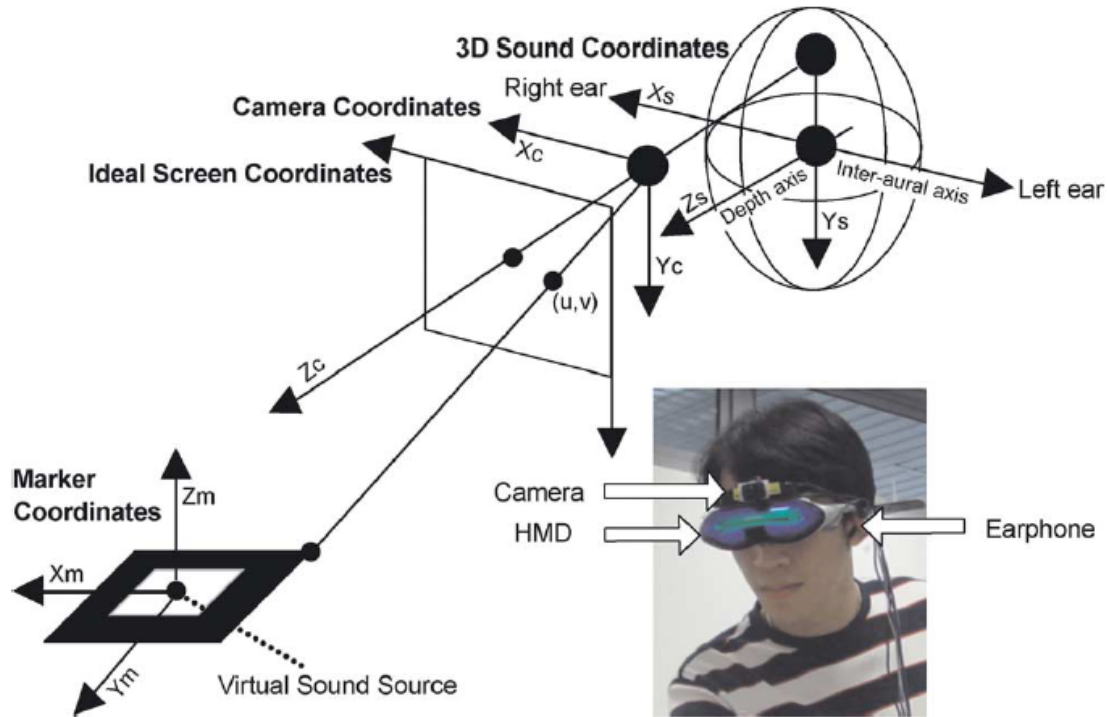


Fig. 1. In order to place a virtual object in the image of the three dimensional scene, the Euclidean transformation between the camera co-ordinates (X_c , Y_c , Z_c) and the marker co-ordinates (X_m , Y_m , Z_m) is estimated. This calculation is based on the imaged position of the corners of the square marker. The imaged position of one of these corners is depicted with pixel co-ordinates (u,v). To obtain the transformation matrix from the 3D sound co-ordinates to marker co-ordinates, translations should be taken from the camera co-ordinates to the 3D sound co-ordinates (X_s, Y_s, Z_s). The picture at the right bottom shows the physical setting of our single camera system where camera is mounted to the front of HMD and 3D sound plays back into the earphones.

Fig. 1. As shown above, “to attach the virtual sound source on the [fiducial] marker, the TCM [transformation matrix] from the 3D sound co-ordinates to marker co-ordinates could be calculated by simply applying translations to the TCM obtained before.” *Id.* Thus, once the “3D position of the marker in the 3D sound co-ordinates are [] fed into OpenAL (Loke Entertainment Software) for synthesis,” 3D sound “appears to emanate from the marker.” *Id.* at 995; *see also id.* at Fig. 2.

178. As Zhou explains, in AR settings, when “two markers of different size are located in different depth[s],” it is difficult to determine the depth and size of the

objects “lacking cues.” *Id.* at 996, Fig. 2. Zhou’s first experiment “requires the subject to tell which of the augmented objects is closer (less deep) to her” under “vision plus 3D sound” and “vision only” conditions. *Id.* at 997. This experiment was designed to determine “whether 3D sounds could be helpful to improve depth perception” in an AR environment. *Id.* at 995.

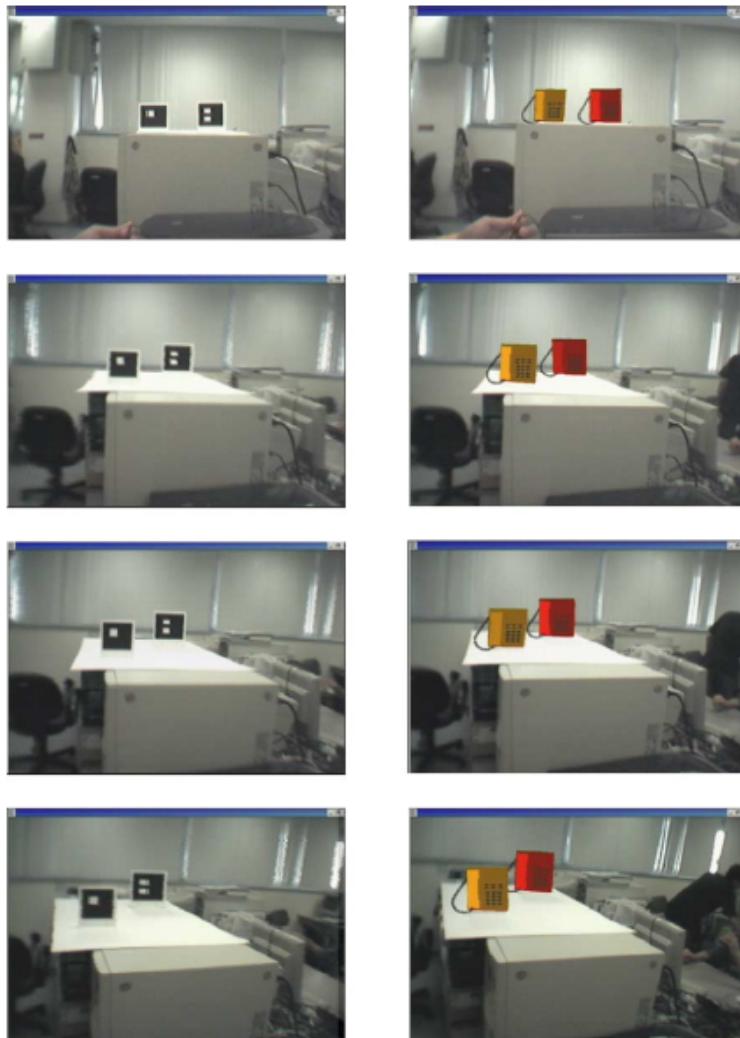


Fig. 2. The depth confusion problem in a single-camera AR setting. The left column shows the image sequences of a scene captured by a moving camera while the right column shows the augmented view seen by the user. (1) In the top left frame, two markers of different size are located in different depth which is hard to tell because of lacking cues. In the following frames, we know that the left marker is nearer to the camera with the cues obtained from the motion, background and lighting environment. (2) Right column shows that the augmented objects occlude everything behind them hence weaken the effect of motion cue. Without background information (in this special case, the objects are located on the same board), relative depth is hard to tell.

Id. at Fig. 2.

179. In the experiment, users “were required to tell the relative depth of the two markers upon which virtual objects were augmented” *Id.* at 999. The “relative position of the two markers was chosen carefully so that the two markers seem to be approximately the same size in the camera’s view,” which “will cause depth confusion” and “hence [Zhou could] decide the effectiveness of adding 3D sound to the virtual objects.” *Id.* at 1000. “[S]ound sources [we]re always placed around 1.5 m away” and “in front of the listener” for this first experiment. *Id.* at 993. And Zhou found that “3D sound” led to “2.5 times as accurate” results compared to tests without 3D sound. *Id.* at 1001.

180. Zhou’s second experiment involved a game with the objective of “rescu[ing] the princess” where three “stages were used” to investigate the effects of 3D sound. *Id.* at 1006. Users could “walk around in a three by three meter region” and saw “a virtual land augmented on the physical ground through a small rectangular virtual window displayed” on a tracked wand. *Id.* at 1007. A small virtual plane was “attached to the wand” and functioned as “an avatar” for the users, allowing them to “explor[e] the game space by manipulating the plane.” *Id.* Players “move[d] freely in the physical space” and in different stages, “transition[ed] from” an “AR environment to VR environment.” *Id.*

181. To test “the impact of 3D sound on task completion time and subjective feelings in this AR environment, 3D sounds were applied in this experimental game.” *Id.* at 1008. Zhou found that “3D sound helps significantly to shorten ... the searching task completion time in this AR environment.” *Id.* at 1010.

IX. PETITION GROUNDS

182. The following table summarizes the grounds that I understand are presented in Petitioners’ petition for *inter partes* review:

<u>Ground</u>	<u>Relevant Prior Art</u>	<u>Claims</u>
Ground 1 (Section IX)	Finnegan, alone or in view of Begault	Claims 8-10 and 15-17
Ground 2(a) (Section XI.D)	Norris, alone or in view of Begault and Algazi	Claims 8-10 and 13-17
Ground 2(b) (Section XI.E)	Norris, alone or in view of Begault and Algazi, and further in view of Sauk	Claims 11 and 12
Ground 3(a) (Section XII.C)	Shivappa, alone or in view of Zhou	Claims 8-10 and 15-17

183. In analyses provided in Sections IX-XII below, I discuss each limitation in the challenged claims, and I explain how the prior art identified in the table above discloses and renders obvious each limitation, and each claim as a whole.

X. GROUND 1: CLAIMS 8-10 AND 15-17 ARE UNPATENTABLE OVER FINNEGAN, ALONE OR IN VIEW OF BEGAULT

184. In this section, I explain why every limitation of claims 8-10 and 15-17, and each of claims 8-10 and 15-17 as a whole, would have been obvious to persons of ordinary skill in the art over Finnegan, alone or in view of Begault.

A. Reasons Persons of Ordinary Skill in the Art Would Combine Finnegan and Begault

185. A person of ordinary skill in the art would have considered it obvious to, and would have good reason to, use DSPs in a system like Finnegan's, including as Begault described.

186. Finnegan disclosed that “[i]n binaural environments, *digital signal processing* provides abilities to alter the intensity, frequency and reverberation present in the audio signal” Finnegan at 3. And persons of ordinary skill in the art understood that DSPs could be advantageously used to process 3D sounds for virtual images. *See* Section VII.D. This was routine, commonplace, and straightforward for skilled artisans.

187. Begault disclosed the use of digital signal processors (“DSPs”) for spatializing sound sources. Begault explained, among other things, that “[t]he 3-D audio component of an integrated audio subsystem ... can contain anywhere from one to as many DSPs as can be accommodated,” and that “[w]ithin a virtual reality

or multimedia application, 3-D sound processing is almost always best accomplished digitally, using audio DSP chips.” Begault at 100-101.

188. Persons of ordinary skill in the art would have been motivated to implement DSPs for 3D sound processing for images because DSPs were known to be advantageous for that type of processing. DSPs are fast and have better power performance, speed, and efficiency than a general-purpose processor. *Id.* at 100 (explaining that DSP’s work “*very quickly*—i.e., in real-time”) (emphasis in original). Persons of ordinary skill in the art would understand that this speed that allows for real-time processing is especially helpful in HMDs where processing of binaural sounds depends on the location and orientation of the sound with respect to a user’s head, both of which may be moving.

189. DSPs also excel at signal processing tasks like HRTF processing for 3D audio (i.e., binaural sound). *See* Otto at 13:50-53 (Explaining that its “DSP strategy” provides optimal performance in terms of directivity of the sound beam created and of the stability of the binaural rendering at higher frequencies”); Brown at Abstract (providing a model for synthesizing binaural sound” and noting that the “the model permits efficient implementation in DSP hardware, and thus facilitates real-time operation”) They are reconfigurable and flexible for different types of signal processing. Begault at 100. (explaining that “[o]ne can select a DSP that is most appropriate for a specific operation or type of digital filtering,” DSPs can be

programmed with a “powerful instruction set,” and DSPs can be optimized for “specific complex operation[s]”). For example, the Convolvotron discussed in Section VII.D used a “custom-made board that contains four INMOS A-100 cascable DSPs.” *Id.* Crystal River Engineering also made a “low-cost version that spatializes two sources in real-time” and uses “a Motorola 56001 DSP rather than the A-100.” Begault at 168.

190. In the late 1990s DSPs were commonly used to process HRTF-based 3D audio. The Crystal River Engineering technology developed for NASA’s Virtual Environment Workstation VIEW project was the foundation for the technology used by Aureal Semiconductor for 3D audio processing. In 1997, Aureal Semiconductor (later acquired by Creative Labs) released their Vortex AU8820 (later expanded to the AU8830 Vortex 2) audio accelerator chipset to improve the quality and performance of their A3D technology. The A3D technology aimed at delivering HRTF-processed 3D audio over headphones. It later enhanced their processing, adding the 100MHz Motorola DSP56362. *See* Ex. 1074, Andrawes, M., *Audio Report – Comdex ’99* (1999), available at <https://www.anandtech.com/show/417/2>.

191. Further, DSPs are appropriately sized for use in a system like Finnegan’s. *See* OnQ Blog, *What’s So Special about the Digital Signal Processor*, <https://www.qualcomm.com/news/onq/2013/12/whats-so-special-about-digital-signal-processor> (December 5, 2013) (Ex. 1069) (“Qualcomm blog”) (explaining that

“[t]he DSP plays a key role in mobile heterogeneous computing. New applications offer breakthrough experiences and increase our engagement with our mobile devices. Many of these experiences—such as sound and image enhancements as well as advanced camera and sensor capabilities—include signal-processing tasks that the DSP excels in executing at low power. And saving power not only extends the battery life of the mobile device, it also keeps your phone cooler and provides a better user experience”).

192. Also, persons of ordinary skill in the art would have had a reasonable expectation of success in combining Finnegan’s device and system with Begault’s DSPs because DSPs were frequently and routinely used by persons of ordinary skill in the art for processing binaural sound in 3D audio applications long before May 2017. *See, e.g.*, Shivappa at 5:6-20, 27:63-28:24, 28:58-29:22 (disclosing an audio processing device that can be a DSP to generate an output spatialized audio signal); Sauk at 7:24-28 (disclosing processing sound with HRTFs that “can be implemented by” “one or more digital signal processors”); Nishio 10:25-33 (disclosing that processing using HRTFs can be performed mainly by a DSP (Digital Signal Processor)); Fukui at 15:63-66 (“All the signal processing ... of the measurement method of head related transfer functions can be performed in a DSP (Digital Signal Processor)"); Otto at 10:52-58 (“The DSP for the binaural mode involves the

convolution of the audio signal to be reproduced with a set of digital filters representing a Head-Related Transfer Function (HRTF).”).

193. And persons of ordinary skill in the art understood that because binaural sound processing requires fast operations (i.e. in real time), binaural sound processing is “almost always best accomplished digitally, using audio DSP chips to manipulate signal sources.” Begault at 100. Furthermore, it was known that DSPs could be used with Mac computers, like the Macbook used in Finnegan. *See* Ex. 1052, *Audio Add-ons For Macbook Pro*, available at <https://www.soundonsound.com/techniques/audio-add-ons-macbook-pro> (disclosing external DSP compatible with Macs). Thus, this combination would have been obvious to try and would have been well within a person of ordinary skill’s common sense. It also would have required no more than combining prior art elements, i.e. the use of a particular type of processor, according to known methods in the industry, i.e. processing of HRTFs to generate binaural sound. *See also* Section VII.D (discussing known sound processing techniques and processors).

B. Independent Claim 8

1. 8[pre]: “An electronic device, comprising:”⁶

194. Finnegan teaches the preamble of claim 8. At a minimum, Finnegan renders the preamble obvious.

195. Finnegan discloses use of an “Oculus Rift Development Kit 2 HMD,” an electronic device, with audio rendered using a custom plugin “built for the Unity Game Engine,” a “Macbook Pro,” and “a pair of Sennheiser HD201 Lightweight-Over-Ear Binaural Headphones.” Finnegan at 2.

2. 8[a]: “a display that displays an image that occurs within a near-field distance from a head of a user”

196. Finnegan teaches, or at a minimum renders obvious, limitation 8[a].

197. To begin, Finnegan presented images to users with an Oculus Rift HMD. Finnegan at 2. And for reasons I explained above in Section VIII.A (Finnegan Overview), persons of ordinary skill in the art would have known and understood that an HMD displays images on its lenses, inches from a user’s eyes, which is “within a near-field distance from a head of a user,” i.e., less than a meter away.

198. In fact, Finnegan depicts images presented on the lenses of its HMD—which users viewed “through the HMD viewport”—in its Figure 1:

⁶ This declaration uses the same bracketed identifiers used in Petitioners’ petition.

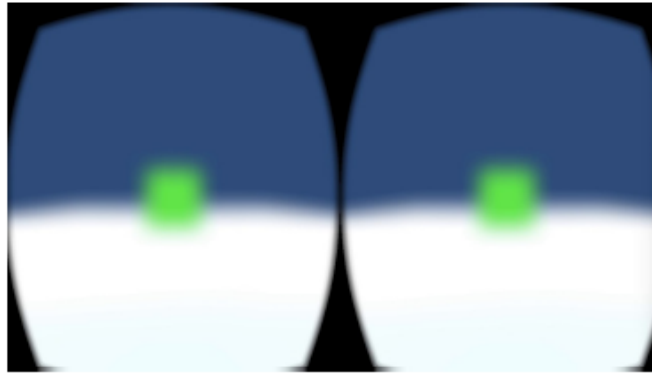


Figure 1: An example screenshot of an audiovisual condition in our experiment, with the visual noise at the highest level (AV3).

Finnegan at 7, Fig. 1; *see also id.* at 6 (disclosing that the “virtual reality (VR) head mounted display” displays images, e.g., “target[s],” to users, and that “[a]n example screenshot of what the participants saw inside the headset is shown in Figure 1.”).

199. As the following images of the Oculus Rift DK2 demonstrate, the lenses in the Rift DK2 were well within 1 m from a user’s head:



Ex. 1050; *see also* THE VR BOOK, Fig. 3.3 (presenting another image of an Oculus Rift). The Oculus Rift was a predecessor to the Oculus Quest. *See* Ex. 1068, *Oculus: Complete Guide — History, Products, Founding, and More*, available at

[https://history-computer.com/oculus-](https://history-computer.com/oculus-history/#:~:text=Android%2C%20and%20iOs.-,Gear%20VR,was%20powered%20by%20Samsung%20smartphones.)

[history/#:~:text=Android%2C%20and%20iOs.-,Gear%20VR,was%20powered%20by%20Samsung%20smartphones.](https://history-computer.com/oculus-history/#:~:text=Android%2C%20and%20iOs.-,Gear%20VR,was%20powered%20by%20Samsung%20smartphones.)

200. I further note that claim 8 appears to require this configuration for HMDs, notwithstanding the '836 patent specification's disclosures relating to "tricking" a user into believing incongruent sound originates at an image's location. See '836 patent, 6:26-48 ("The listener thus perceives that the sound originates from the image even though the sound is convolved to originate from a position that is behind the image."), 21:48-22:43. Specifically, with respect to the location of a displayed image and the location of the electronic device displaying that image, claim 8 requires (i) "a display that displays an image that occurs within a near-field distance from the head of a user," (ii) that the "display displays the image to the user at the near-field distance," and (iii) that "the electronic device is located at the near-field distance from the head of the user." '836 patent at cl. 8. In other words, the electronic device must be at the same near-field distance at which the display displays the image to the user.

201. Consequently, if all that is required for limitation 8[a] is the display of an image on an HMD's lenses, that is disclosed in Finnegan and would have been obvious to persons of skill in the art over Finnegan (or any VR HMD reference).

202. Additionally, if the pertinent “near-field distance from a head of a user” for limitation 8[a] is the distance from the head of a user to the location of a *perceived* image—which is more consistent with the ’836 patent’s specification but is not how Patent Owner interprets this limitation or what claim 8 recites—Finnegan taught this too.

203. In Finnegan’s experiments, users wore Oculus HMDs, as noted, and images were displayed to users at *perceived* locations ranging from 0.5 m to 9.5 m in front of the users while associated auditory stimuli were presented at congruent and incongruent locations. Finnegan at 7. And any distance greater than or equal to 0.5 m and less than 1 m that Finnegan tested was a “near-field” distance as defined by the ’836 patent (’836 patent at 27:23-25). Thus, Finnegan taught *perceived* near-field distances for displayed images as well.

3. 8[b]: “one or more processors that process sound for the image”

204. Finnegan discloses, or at a minimum renders obvious, limitation 8[b].

205. Finnegan disclosed that its virtual environment was “simulate[d]” using a “2.9GHz Intel i7 processor” in a “Macbook Pro (13-inch, Mid 2012 model).” Finnegan at 6.

206. Finnegan further disclosed that “digital signal processing” in “binaural environments” allows “alter[ing of] the intensity, frequency and reverberation

present in [an] audio signal,” and that “[t]hrough digital signal processing techniques and geometric manipulations, visual and auditory distance cues can be modified to alter the impression of the virtual space.” *Id.* at 3.

207. From at least these disclosures, persons of skill in the art would have known and understood that Finnegan used its 2.9GHz Intel i7 processor to process sound for an image. At a minimum, Finnegan’s disclosures would have rendered the claimed “one or more processors processing sound for [an] image” obvious.

208. As I explained in Section VII.D above, in the State of the Art section of this declaration, the use of processors to process sound for images was routine and conventional before May 2017. *See* Section VII.D.

209. I further note that persons of ordinary skill in the art also would have known Finnegan’s teachings could be applied with hardware other than an HMD, e.g., a handheld portable electronic device, which could and would display images at near-field distances, e.g., while held in front of a user, including with SLPs for binaural sound positioned behind such images per Finnegan’s disclosures. Indeed, it was known that binaural sound was used in various types of devices, including in handheld devices that include a display. *See* U.S. Patent No. 9,037,468 (“Osman”) (Ex. 1082) at 4:37-67 (describing a “sound localization function” like a “Head-Related Transfer Function” for localizing sound); 8:65-9-17 (describing an “augmented-reality environment with realistic delivery of sound” with a virtual scene

“seen through a screen of the [handheld] device”), 9:49-56 (explaining that as the player moves the “portable device 408a, the view of the virtual scene changes as if the player were holding a camera into the virtual world.”) Figs. 2, 4.

4. 8[c]: “to generate binaural sound that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”

210. Finnegan discloses, or at a minimum renders obvious, limitation 8[c].

211. Persons of ordinary skill in the art would have known and understood that Finnegan’s processing generated, or at a minimum obviously generated, “binaural sound.” Finnegan produced “binaural audio” that played through “Binaural Headphones” for its tests. Finnegan at 2, 6. Further, Finnegan taught that its system “integrate[d] the SoundScape Renderer (SSR), a GPL licensed software implementation for binaural audio [1].” *Id.* 6 (citing “[1],” “Jens Ahrens, Matthias Geier, and Sascha Spors. 2008. The Sound-Scape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods. In *Audio Engineering Society Convention*. Audio Engineering Society.” [(“Ahrens”) (Ex. 1029)]).

212. Further, as Finnegan states, “the visual component of an object will tend to produce distance compression,” so “[i]f we place the audio component at a greater distance from the observer,” then “the auditory sensory signal should be integrated

with the visual sensory signal to produce a combined distance estimate that is closer to the intended distance.” *Id.*

213. Moreover, Finnegan disclosed, or at a minimum rendered obvious, the use of SLPs at far-field distances.

214. Finnegan’s tested range for visual and auditory stimuli was “approximately the 10 meters in front of the observer.” *Id.* at 7. More specifically, Finnegan used a staircase algorithm for a range of sound locations from “0.5 to 9.5 meters in front of the observer.” *Id.* Thus, all the incongruent audio that Finnegan positioned at distances of 1 m or greater—which would occur in numerous instances in Finnegan’s tests ranging from 0.5 m to 9.5 m—would be at locations where a listener localizes sound (SLPs with coordinate locations) at “far-field” distances. *See* Finnegan at 7.

215. Further, the sounds Finnegan positioned at “far-field” distances were “behind” the images Finnegan “concurrently” displayed on an Oculus Rift HMD’s lenses, which were well within 1 m (i.e., at “near-field” distances) from the user for reasons discussed in the context of limitation 8[a]. *See* Section X.B.2; *see also* Finnegan at 6 (“The stimuli presented to each participant were the same, and consisted of a visual cube, an auditory pink noise burst, or both concurrently.”), Fig. 1.

216. Persons of ordinary skill in the art would have known and understood that the “incongruent” sounds that Finnegan tested were, in addition to being behind

tested images, along a line-of-sight from the user to such images. To test Finnegan's hypothesis of whether "incongruent" sounds would be "integrated with the[ir] visual sensory signal[s]," persons of ordinary skill in the art would have known that Finnegan's 3D sounds should be presented both behind the tested visual stimuli *and* along an observer's line-of-sight. Finnegan at 3. This would be the best alignment for controlling Finnegan's test of whether incongruence between sounds and visual stimuli impacted a user's ability to determine distances to the visual stimuli.

217. To this end, Finnegan's test setup secured the direction of participants' lines-of-sight so that participants were looking straight at visual stimuli as well as the locations for the displayed images' incongruent sounds. Finnegan's participants locked in their head positions by placing their chins on a "chin rest before beginning the experimental conditions." *Id.* at 7.

218. Additionally and alternatively, if the *perceived* distance to an image is the pertinent "near-field" distance for the displayed image of limitation 8[c], Finnegan disclosed far-field SLPs located behind such images. Finnegan discloses rendering audio "at an incongruent position further from the observer than [a] visual stimulus (IV)"—i.e., to a "sound localization point (SLP) with a coordinate location ... behind the image"—to facilitate "more accurate distance perception (DV) compared to conditions where both stimuli are at the same position (congruent conditions)." Finnegan at 6.

219. When Finnegan tested “incongruent conditions,” “the visual stimulus was at the reference distance with the audio stimulus offset by [an] incongruent positioning function.” *Id.* As Finnegan explains, other skilled artisans had “investigated distance compression in virtual auditory environments,” and their work “provided the following exponential function for describing the degree to which humans compress distance”:

$$\hat{y} = k\phi^\alpha \quad (1)$$

Id.

220. From this, Finnegan derived its Equation 2, reproduced below, which is the “incongruent positioning function” that Finnegan used to “systematically position the audio component of a virtual object incongruently to”—and behind—“its visual component”:

$$\phi = \left(\frac{\hat{y}}{k} \right)^{\frac{1}{\alpha}} \quad (2)$$

Id. at 5, 7.

221. The $\hat{y}$ variable in this function “is the perceived target position” that “represents the desired distance” for the “the observer to perceive.” The ϕ variable is the “actual target position.” *Id.* And, again based on work of other skilled artisans relating to distance compression compensation, Finnegan set the variables α and k to

“the values 2.22 and 0.61 respectively.” Finnegan at 5. This way, with a known desired distance from an observer (i.e., the tested distances between 0.5 m and 9.5 m to visual stimuli), and α and k set at 2.22 and 0.61, respectively, Finnegan could use Equation 2 to determine the distances from observers to audio stimuli (ϕ) presented behind associated images.

222. The table below shows on the right (rounded to nearest hundredth) the distances to incongruent audio stimuli (ϕ) that result from application of Finnegan’s Equation 2—Finnegan’s “incongruent positioning function”—being applied to distances to visual stimuli ($\hat{y}$) less than or equal to 1 m (on the left, at 0.1 m intervals).

Distance to Visual Image ($\hat{y}$)	Distance to Incongruent Audio (ϕ)
0.5 m	0.91 m
0.6 m	0.99 m
0.7 m	1.06 m
0.8 m	1.13 m
0.9 m	1.19 m
1.0 m	1.25 m

223. As this table demonstrated, persons of ordinary skill in the art would have readily understood that Finnegan presented to users virtual images displayed at near-field distances while concurrently presenting audio with coordinate locations (SLPs) at far-field distances. As shown (and highlighted), that configuration was present in Finnegan’s testing for any tested distances to visual stimuli exceeding 0.6

meters. Also, Finnegan provides that it tested distances from 0.5 m to 9.5 m (Finnegan at 7), and, consistent with that disclosure, Finnegan's figures present results for distances from observers below 1 m, including, for example, results for a distance of approximately 0.75 m. *See* Finnegan at Fig. 3 (Panel C). Applying Equation 2, reproduced above, to a 0.75 m distance to a visual stimulus solves to a distance to an incongruent audio source of 1.098 m, which is greater than 1 m and thus a far-field distance in the context of the '836 patent.

224. I further note that, as explained above, there is no question that Finnegan presented visual and audio stimuli together—Finnegan expressly discloses that it presented stimuli “consisted of a visual cube, an auditory pink noise burst, or both concurrently.” Finnegan at Fig. 1, 6.

225. And for the reasons explained earlier in this section, persons of ordinary skill in the art would have known and understood that the “incongruent” sounds that Finnegan tested were, in addition to being behind tested images, along a line-of-sight from the user to such images.⁷

⁷ I also note that persons of ordinary skill in the art would have understood that Finnegan's teachings could be applied to devices other than HMDs, for example, handheld portable electronic devices such as smartphones, which could and would

5. 8[d]: “wherein the one or more processors include a digital signal processor (DSP)”

226. Finnegan discloses, or at a minimum renders obvious, limitation 8[d].

227. As I explained for limitation 8[b], Finnegan disclosed, or at a minimum rendered obvious, one or more processors processing sound for an image. *See* Section X.B.2; *see also* Finnegan at 6 (disclosing that its testing used an Oculus Rift connected to “a Macbook Pro ... with a 2.9GHz Intel i7 processor”). And based on Finnegan alone—e.g., Finnegan’s disclosure that, “[i]n binaural environments, ***digital signal processing*** provides abilities to alter the intensity, frequency and reverberation present in the audio signal” (Finnegan at 3)—it would have been obvious to persons of ordinary skill in the art to process sound with a DSP in a Finnegan-type system, to implement the incongruent virtual environments Finnegan taught.

228. As I explained in Section VII.D above, for more than a decade before May 2017, persons of ordinary skill in the art understood that DSPs can be advantageously used to process 3D sounds for virtual images. *See* Section VII.D. This was routine and straightforward for skilled artisans.

display images at near-field distances (e.g., in user’s hands), and, per Finnegan’s teachings, SLPs behind the images on their screens.

229. For example, Begault disclosed in 2000 the use of digital signal processors (“DSPs”) for spatializing sound sources, including in virtual environments. *See, e.g.*, Begault at 100-101. Begault explained, among other things, that “[t]he 3-D audio component of an integrated audio subsystem ... can contain anywhere from one to as many DSPs as can be accommodated,” and that “[w]ithin a virtual reality or multimedia application, 3-D sound processing is almost always best accomplished digitally, using audio DSP chips.” *Id.*

230. Begault is just one of many references that demonstrate that persons of ordinary skill in the art would have known and been motivated to implement a system like the one Finnegan disclosed with a DSP to process sound for images. *See* Section VII.D (collecting and discussing such references, including Shivappa, Sauk, Otto, Brown, Nishio, and Fukui).

231. Persons of ordinary skill in the art would have been motivated to implement DSPs for 3D sound processing for images because DSPs were known to be advantageous for that type of processing. DSPs are fast, excel at signal processing tasks like HRTF processing for 3D audio (i.e., binaural sound), and are reconfigurable and flexible for different types of signal processing. Begault at 100. (explaining that “[o]ne can select a DSP that is most appropriate for a specific operation or type of digital filtering,” DSPs can be programmed with a “powerful instruction set,” and DSPs can be optimized for “specific complex operation[s]”); Qualcomm blog (“DSPs

deliver high performance at lower clock frequencies, which saves power”). They also are appropriately sized for use in a system like Finnegan’s. *See* Section VII.D; *see also* Section X.A (reasons to combine).

232. For these reasons, limitation 8[d] would have been obvious to persons of ordinary skill in the art based on Finnegan alone.

233. Additionally, limitation 8[d] would have been obvious to persons of ordinary skill in the art over Finnegan in view of Begault.

234. As noted, Begault describes and prescribes using DSPs for 3D audio processing. Begault at 100-101. And as discussed above, persons of ordinary skill in the art would have known that Begault’s DSPs would be advantageous for processing 3D sound for images due to, among other things, their speed, reconfigurability, and size.

235. Persons of ordinary skill in the art, therefore, would have known and been motivated to implement a Finnegan-type system with a Begault DSP for sound processing, e.g., with the DSP replacing Finnegan’s Intel processor or being added to the system Finnegan’s Macbook used to render its virtual environment, including audio.

236. Persons of ordinary skill in the art would have reasonably expected to succeed in using a DSP, e.g., as taught by Begault, in a Finnegan-type system because doing so would have required no more than combining known elements (e.g., HMDs

and DSPs) and known functions (e.g., HRTF processing) through known methods. *See* Begault at 100-101 (“[w]ithin a virtual reality or multimedia application, 3-D sound processing is almost always best accomplished digitally, using audio DSP chips” and explaining that “[t]he 3-D audio component of an integrated audio subsystem ... can contain anywhere from one to as many DSPs as can be accommodated”); Section VII.D.

6. 8[e]: “that processes the sound for the image with far-field head-related transfer functions (HRTFs) while the electronic device is located at the near-field distance from the head of the user”

237. Finnegan discloses, or at a minimum renders obvious, limitation 8[e].

238. Finnegan discloses an “Oculus Rift Development Kit 2 HMD” with audio processed with “a custom plugin that was built for the Unity Game Engine.” Finnegan at 2.

239. Finnegan also explains that its system “integrate[d] the SoundScape Renderer (SSR), a GPL licensed software implementation for binaural audio [1], with Unity for spatial audio rendering over headphones.” Finnegan at 6 (citing “[1],” Ahrens).

240. A person of ordinary skill in the art would have known and understood that the SoundScape Renderer Finnegan utilized processed sounds with HRTFs. The Ahrens article Finnegan cites expressly states that the SoundScape Renderer “use[d]

Head Related Transfer Functions (HRTFs) to reproduce [a] soundfield at the listeners ears.” Ahrens (Ex. 1029) at 3. Also, as I explained in Section VII.B, in the State of the Art section of this declaration, sound processing with HRTFs to produce 3D audio and binaural sound was well-known to persons of ordinary skill in the art for years before May 2017.

241. Persons of ordinary skill in the art also would have known and understood that the HRTFs from the SoundScape Renderer used in Finnegan’s testing were (or at a minimum were obviously) far-field HRTFs. As discussed above, the 3D sounds Finnegan generated to users were positioned at far-field distances. *See* Section X.B.4. And persons of ordinary skill in the art would have known and understood the benefits of using far-field HRTFs versus near-field HRTFs, the latter of which require “more complicated” processing. Algazi at 35; *see also* Section VII.B.2.

242. In addition, persons of ordinary skill in the art also would have known and understood that the Oculus Rift HMD that Finnegan used would be worn on the user’s head, such that the HMD, including its lenses, would be located at a near-field distance as claimed:



Ex. 1050; *see also* THE VR BOOK, Fig. 3.3 (presenting another image of an Oculus Rift).

243. Moreover, limitation 8[e] also would have been obvious to persons of ordinary skill in the art over Finnegan in view of Begault.

244. Begault described HRTFs and “localization with HRTF cues” and explained that “[t]he use of HRTF spectral shaping is typically featured as the key component of a 3-D sound system.” *Id.* at Chpt. 2, 99. Begault further taught the use of DSPs for processing HRTFs to spatialize sound. *Id.* at 100-101.

245. Thus, limitation 8[e] would have been obvious to persons of ordinary skill in the art over Finnegan in view of Begault.

246. For these reasons, each limitation in claim 8, and claim 8 altogether, would have been obvious over Finnegan, alone or in view of Begault.

C. Dependent Claims 9 and 10

247. I understand that claims 9 and 10 depend on independent claim 8.

248. Thus, it is my opinion that dependent claims 9 and 10 would have been obvious to persons of ordinary skill in the art for the same reasons as independent claim 8. *See* Section X.B.

249. In addition, claims 9 and 10 would have been obvious for the reasons I discuss in the following subsections.

1. **Claim 9: “The electronic device of claim 8 , wherein the image is an augmented reality (AR) image, and the display is included with electronic glasses worn on the head of the user.”**

250. Claim 9 would have been obvious over Finnegan, alone or in view of Begault, because persons of ordinary skill in the art would have recognized that Finnegan’s experiments and disclosures are applicable to augmented reality (“AR”) settings and presenting AR images, including with “electronic glasses.”

251. In describing the background and motivation for its experiments, Finnegan discusses how “[i]ncongruent multisensory environments” like the one it aimed to design are applicable to “augmented reality (AR) environments,” citing Zhou, which studied 3D sound in AR environments. Finnegan at 3-4, endnote 46.⁸

⁸ *See* also Section VIII.G (Overview of Zhou).

252. Finnegan further explains that its teachings are applicable to “mixed reality” settings, e.g., environments that “incorporate movement and interaction with physical and tangible props” from the “real world.” Finnegan at 11.

253. Based on these disclosures, persons of ordinary skill in the art would have known that, on the “virtuality continuum,” “mixed reality” environments include VR and AR environments:

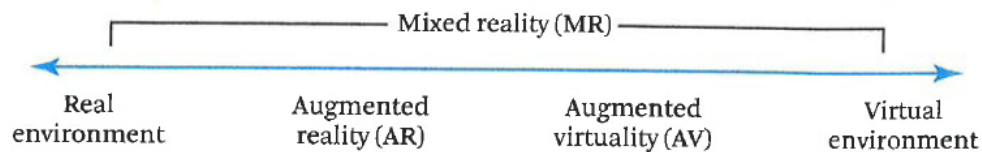


Figure 3.1 The virtuality continuum. (Adapted from Milgram and Kishino (1994))

THE VR BOOK at Fig. 3.1; *see also id.* at 29 (explaining that, on the “virtuality continuum,” “mixed reality” environments include VR and AR environments).

254. In fact, before May 2017, persons of ordinary skill in the art would have understood that AR and VR applications have “many similarities” and often use “the same hardware,” and that “the convergence of AR and VR [wa]s not new from a research perspective.” *Id.* at 484. In fact, Zhou, which Finnegan cites, describes the use of 3D audio in a game in which players “seamless[ly] transition from AR to VR.” *See* Zhou at 1004-09, Fig. 11 (depicting same).

255. Persons of ordinary skill in the art also would have known and been motivated to use “electronic glasses” in an incongruent multisensory AR environment, consistent with Finnegan’s teachings, including to present AR images.

256. “Electronic glasses” were and remain standard AR system hardware. *See* THE VR BOOK at 32-33 (describing “[o]ptical-see-through HMDs [that] enable computer-generated cues to be overlaid onto the visual field and provide the ideal augmented reality experience”). For example, Figure 3.3 of THE VR BOOK, reproduced below, shows images of an Oculus Rift next to a “CastAR” HMD, which are AR “electronic glasses.”



See also THE VR BOOK, Fig. 2.1 (depicting “head-mounted displays and viewers over time”).

257. Thus, claim 9 would have been obvious over Finnegan, alone or in view of Begault.

2. Claim 10: “The electronic device of claim 8 , wherein the image is a virtual reality (VR) image, and the display is included with a head mounted display worn on the head of the user.”

258. As I have already explained (*see* Section VIII.A), Finnegan disclosed use of an “an Oculus Rift Development Kit 2” HMD—i.e., a “head mounted display

worn the head of the user”—for “virtual reality (VR) image[s].” Finnegan at Abstract,

2. And Finnegan used that HMD to display VR images to users. *See id.* Fig. 1; Section X.B.2.

259. Thus, claim 10 would have been obvious over Finnegan, alone or in view of Begault.

D. Independent Claim 15

1. 15[pre]: “An electronic system, comprising”

260. The preamble of claim 15 would have been obvious for the same reasons as the preamble of claim 8. *See* Section X.B.1.

2. 15[a]: “a display that displays an image that is located within a near-field distance from a head of a user viewing the image”

261. Limitation 15[a] would have been obvious for the same reasons as limitation 8[a]. *See* Section X.B.2. Further, a person of ordinary skill in the art would understand that a user would be viewing the image displayed by display within a near-field distance from a head, i.e. a HMD worn a few inches from the user’s head.

3. 15[b]: “one or more processors that process sound for the image”

262. Limitation 15[b] would have been obvious for the same reasons as limitation 8[b]. *See* Section X.B.3.

4. **15[c]: “with head-related transfer functions (HRTFs) to generate binaural sound”**

263. Limitation 15[c] would have been obvious for the same reasons as limitation 8[e]. *See* Section X.B.6. In addition, a person of ordinary skill in the art would have known that the HRTFs are used for generating binaural sound for the same reasons as discussed for limitation 8[c]. *See* Section X.B.4

5. **15[d]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”**

264. Limitation 15[d] would have been obvious for the same reasons as limitation 8[c]. *See* Section X.B.4.

6. **15[e]: “wherein the coordinate location of the binaural sound continues to occur at the far-field distance located behind the image when the head of the user moves away from viewing the image and the image remains within the near-field distance from the head of the user.”**

265. Finnegan discloses or renders obvious limitation 15[e] for reasons discussed above. *See* Sections X.B (claim 8), X.D.1-X.D.5 (15[pre]-15[d]). Specifically, a coordinate location of binaural sound occurring at a far-field distance located behind an image would have been obvious to persons of ordinary skill in the art for reasons discussed. The additional recitations of limitation 15[e] would have been obvious over Finnegan in view of Begault for the following reasons.

266. Finnegan explained that its experiments were designed to trigger “incongruent design in VE development,” and identified as “an example of where [its] methods could be applied” a virtual environment where “the observer can move freely about [a] scene.” Finnegan at 11. Finnegan explained that by “[u]sing tools that automate incongruent positioning calculation, the designer” of such an environment “could design the scene as she sees fit, with the scene being generated automatically at runtime to accommodate for the observer’s distance compression.” *Id.*

267. Accordingly, based on Finnegan alone, persons of ordinary skill in the art would have known and been motivated to implement Finnegan’s incongruent positioning methods to address distance compression in virtual environments in which “observers can move freely.” Finnegan at 11. And persons of ordinary skill in the art designing such environments would have known and understood the need to control for head movement, including when “the head of [a] user moves away from viewing [an] image and the image remains within the near-field distance from the head of the user,” as claimed. As I explained in Section VII.C.1 above, in the State of the Art Section, that type of 3D audio control was well-known and routing before May 2017.

268. For example, persons of ordinary skill in the art would have known and been motivated to look to Begault for its disclosures of sound processing for virtual

reality applications, including sound processing techniques by DSPs. And in its detailed descriptions of such processing, Begault explains that “[i]n virtual reality, the key to immersivity is directly tied to the success of the interface in providing realistic information at the effectors and the degree to which sensors are able to track human interaction.” Begault at 167. Therefore, Begault further explains, “head-tracked virtual acoustic systems represent the best approach to implementing sound, so that sound sources can remain fixed in virtual space independent of head movement, as they are in natural space hearing.” *Id.*; *see also id.* at 133 (“The 3-D audio subsystem then can compensate for the relative position of source and sensor to give the illusion of a sound that remains fixed in space independent of head movement. Figure 4.38 shows the method used by a commercially available 3-D sound system for real-time placement of a virtual source to a fixed position, independent of the listener’s movement.”), Figs. 4.37-4.38.

269. Persons of ordinary skill in the art, therefore, would have known and been motivated to keep sound sources “fixed in virtual space independent of head movement” (e.g., the head of a user moving away), as Begault (and others) taught, including when implementing Finnegan’s teachings of using incongruent audio at a far-field distance and an image remaining at a near-field distance.

270. Finnegan’s near-field images also would obviously remain within “near-field distance[s] from the head of the user,” on the HMD’s lenses, until, for example, they were dropped from the field of view due to a user’s head movement.

271. For these reasons, each limitation in claim 15 of the ’836 patent, and claim 15 as a whole, would have been obvious over Finnegan, alone or in view of Begault.

E. Dependent Claims 16 and 17

272. I understand that claims 16 and 17 depend on independent claim 15.

273. Thus, it is my opinion that dependent claims 16 and 17 would have been obvious to persons of ordinary skill in the art for the same reasons as independent claim 15. *See* Section X.D.

274. In addition, claims 16 and 17 would have been obvious for the reasons I discuss in the following subsections.

1. Claim 16: “The electronic system of claim 15, wherein the display is an augmented reality (AR) display worn on the head of the user viewing the image, and the image is an AR image.”

275. Claim 16 would have been obvious for the same reasons as Claim 9. *See* Section X.C.1. Further, a person of ordinary skill in the art would understand that a display that displays an AR image would be an AR display and that a user would be viewing the AR image when the AR display, i.e. an HMD, is worn on the user’s head.

2. Claim 17: “The electronic system of claim 15, wherein the display is a virtual reality (VR) display worn on the head of the user viewing the image, and the image is an VR image.”

276. Claim 17 would have been obvious for the same reasons as Claim 10.

See Section X.C.2.

277. For these reasons, claims 16 and 17 of the '836 patent would have been obvious over Finnegan, alone or in view of Begault. Further, a person of ordinary skill in the art would understand that a display that displays an VR image would be a VR display and that a user would be viewing the VR image when the VR display, i.e. an HMD, is worn on the user's head.

XI. GROUND 2: CLAIMS 8-17 ARE UNPATENTABLE OVER NORRIS, ALONE OR IN VIEW OF BEGAULT AND ALGAZI, AND FURTHER IN VIEW OF SAUK

278. In this section, I explain why every limitation of claims 8-17, and each of claims 8-17 as a whole, would have been obvious to persons of ordinary skill in the art over Norris, alone or in view of Begault and Algazi, and further in view of Sauk.

A. Reasons Persons of Ordinary Skill in the Art Would Combine Norris and Begault

279. For the same reasons persons of ordinary skill in the art would have known and been motivated to modify a Finnegan system to use a DSP, or to combine a Finnegan system with a Begault DSP (and Begault's DSP sound processing), *See* Section X.A, persons of ordinary skill in the art would have known to modify a Norris

system to use a DSP, or to combine a Norris system with a Begault DSP (and Begault's DSP sound processing). *Id.*; *See also* Section VII.D (discussing processors used for rendering binaural sound). I discuss this further below, in the context of the challenged claims.

B. Reasons Persons of Ordinary Skill in the Art Would Combine Norris and Algazi

280. A person of ordinary skill in the art would have considered it obvious to, and would have good reason to, use HRTFs from Algazi's database with the methods, devices, and system for providing sound localization described in Norris, or a Norris-Begault system.

281. Norris discloses using HRTFs from a "database of common approximate profiles." *See, e.g.*, Norris, 21:47-50. And a person of ordinary skill in the art would have understood that using a database of HRTFs would be more efficient than acoustically measuring, or synthesizing HRTFs. *See, e.g.*, Romblom at 1 (describing the procedure of physically taking HRTF measurements as "cumbersome"), Zhou at 992 (explaining that "individual HRTFs are difficult for AR implementation due to the fact that AR systems are designed to be real time interactive systems that do not need individual user configuration.").

282. Algazi describes the public CIPIC database of HRTFs which provides a "database of high-spatial-resolution head-related transfer functions." Specifically,

Algazi discloses “a public-domain database of high-spatial-resolution head-related transfer functions” using HRTFs from “90 [human] subjects” that includes “impulse responses for 1250 directions for each ear of each subject” and “a set of anthropometric measurements that can be used for scaling.” Algazi at 99.

283. As I explained above, the CIPIC database described in Algazi was and remains a well-known and widely-used HRTF database, and a person of ordinary skill in the art would have also had good reason to use HRTFs from Algazi’s public database because developing HRTF databases is a time-consuming and expensive endeavor. A person of ordinary skill in the art also would have had good reason to use Algazi’s HRTF database since Algazi’s database is both public and because Algazi discloses that its HRTFs provide better results than other public HRTF databases such as the “public domain-dataset for the KEMAR mannequin.” Algazi at 99.

284. Also, a person of ordinary skill in the art would have had a reasonable expectation of success in combining Norris’s devices, systems, and methods (or a Norris-Begault system) with Algazi’s CIPIC HRTFs because CIPIC and other database HRTFs were frequently and routinely used by persons of ordinary skill in the art for producing binaural sound in 3D audio applications, long before May 2017. Begault at 112 (explaining that designers of 3D sound systems in most cases used HRTFs from “a library of HRTF measurements from which particular sets c[ould] be

selected on the basis of either user preference or behavioral data.”); Algazi at 99 (“several studies have employed the public-domain dataset for the KEMAR mannequin”), *id.* at 102 (noting the “public availability of the CIPIC HRTF database”); *see also* Begault at 12 (describing components of the Convolvotron and explaining a “general configuration for the interface of a sound subsystem in a virtual reality system” where the “host computer calls on databases from the reality engine that describe the acoustic world”).

285. Thus, combining CIPIC database HRTFs, as described by Algazi, with a Norris or a Norris-Begault system would have been obvious to try and would have been well within a person of ordinary skill in the art’s common sense. Indeed, Begault explains that these types of databases were used “in most cases, [where] nonindividualized (generalized) HRTFs [we]re used.” Begault at 45. Such a combination also would have required no more than combining prior art elements according to known methods in the industry. And, as mentioned above, Norris itself recommends using “stored HRTFs” “from a database of common approximate profiles,” Norris at 21:47-50, further demonstrating that it would have been obvious to use HRTF databases, including from public sources like the CIPIC HRTF database. Persons of ordinary skill in the art also would have reasonably expected to succeed in using CIPIC HRTFs with a Norris or Norris-Begault system, as such a combination would have been straightforward and routine.

C. Reasons Persons of Ordinary Skill in the Art Would Combine Norris, Algazi, and Sauk

286. A person of ordinary skill in the art would have considered it obvious to, and would have good reason to, combine the teachings of Norris, alone or with Begault and Algazi, and Sauk.

287. Sauk, like Norris, aims to provide advantageous mechanisms for improving the provision of 3D audio, including in AR, VR, and mixed-reality settings. *See, e.g.*, Sauk at 2:45-47 (“The invention concerns a method and apparatus for producing, combining, and customizing virtual sound environments to provide the user with understandable information regarding their surroundings.”); Norris at 9:12-17 (discussing “virtual realities or online games” or “play spaces”), 37:40-56 (describing how “[s]ound can be adjusted to include audial cues for an artificial 3D physical environment”), 41:48-59 (discussing “electronic eyeglasses or eye wear” for “augmented reality imagery”). Sauk describes “innovative applications for actually putting [] binaural audio information to practical use.” Sauk at 2:34-41.

288. A person of ordinary skill in the art would have recognized, long before May 2017, that Sauk’s methods of switching from binaural to stereo (or monophonic, in some instances) sound to indicate relevance of audio signals could enhance or maintain user experience and perceptual accuracy while decreasing computational processing requirements. Indeed, processing binaural sound requires significantly

more processor resources, and thus requiring faster, more expensive processors can be eliminated by switching from binaural to stereo in certain situations, such as when the sound source is directly to the left or right of a user's head. Sauk at 2:7-13 (“Considerable computational resources are required to implement accurate HRTFs because they are very complex functions of direction and frequency. The overall design of the binaural audio system is very important to reduce implementation costs, improve sound feed-back rates, and to implement practical binaural sound fields which may include many sound sources”); *see also* Begault at 100. (explaining that DSPs can be programmed with a “powerful instruction set,” and DSPs can be optimized for “specific complex operation[s]”). As Algazi 2011 explains, “two major issues in the implementation of HRTF-based rendering are computational cost and latency.” Algazi 2011 at 38. The computational requirements depend on the “complexity of the auditory scene” and “the efficiency” of the algorithm implementation, and “[t]he approach to these issues depends on the application and on the decision as to what constitutes an acceptable auditory experience, as opposed to one that is indistinguishable from actually being present.” *Id.* Thus, degrading binaural sound for less relevant sound sources would have been obvious to a person of ordinary skill in the art in order to decrease the processing power required to attain these benefits.

289. Further, a person of ordinary skill in the art would have had a reasonable expectation of success in combining Norris, alone or with Begault and Algazi, and Sauk. Like Norris, Begault, and Algazi, Sauk disclosed ways to enhance binaural sound, including by using HRTFs and through implementations on wearable electronic devices. Norris and Sauk also implement their methods using head and position tracking, as well as HRTF processing, in analogous ways (as does Begault). Sauk at 4:33-5:5 (position and head tracking), 7:24-8:33 (sound processing with HRTFs); Norris at 6:29-45 (position and head tracking), 4:20-29 (sound processing with HRTFs). In other words, Sauk provides a method of degrading less relevant sound that is readily transferable to the devices and methods that Norris discloses, or that would be used in a Norris-Begault system.

D. Ground 2(a) Claims 8-10 and 13-17 are Unpatentable Over Norris, Alone or in View of Begault and Algazi

290. In this section, I explain why every limitation of claims 8-10 and 13-17 of the '836 patent, and each of claims 8-10 and 13-17 as a whole, would have been obvious to persons of ordinary skill in the art over Norris, alone or in view of Begault and Algazi.

1. Independent Claim 8

a. 8[pre]: “An electronic device, comprising:”

291. Norris discloses the preamble of claim 8.

292. Norris's systems, devices, and methods comprise electronic devices, as Norris discloses, for example, an "electronic system" that includes "a listener wearing a wearable electronic device. Norris at 2:8-11; *see also id.* at 2:33-35 (describing an "electronic system in which electronic devices of two individuals communicate during an electronic call"), 2:40-46 (describing "a computer system in which two electronic devices communicate with each other during a telephone call between two people"), 40:10-41:19 ("a computer and an electronic device include, but are not limited to, handheld portable electronic devices (HPEDs), wearable electronic glasses, watches, wearable electronic devices....combinations of these devices, and other portable and non-portable electronic devices and systems."), Figs. 17, 25-26.

b. 8[a]: "a display that displays an image that occurs within a near-field distance from a head of a user"

293. Norris teaches, or at a minimum renders obvious, limitation 8[a].

294. To begin, Norris discloses wearable electronic devices (WEDs), including "head-mounted displays" and "electronic glasses" where "augmented reality imagery is projected through or reflected off a surface of a lens." Norris at 41:47-59 ("a 'wearable electronic device' is a portable electronic device that is worn on or attached to a person. Examples of such device include Head-mounted displays, electronic eyeglasses or eye wear (such as glasses in which augmented reality imagery is projected through or reflected off a surface of a lens)"); 31:52-

32:48 (describing a “wearable electronic device” such as “electronic glasses with a display that provides an image or video” used for an “electronic call” that displays “a virtual image of the speaker”), 37:29-39, 40:59-41:9, 41:48-59, Figs. 17, 19-20, 25-26.

295. Also, for reasons I explained in Section VII.A, persons of ordinary skill in the art would have known and understood that WEDs such as head-mounted displays and electronic glasses display images on lenses that are well within 1 m from user’s heads (e.g., inches from a user’s eyes), which is “within a near-field distance from a head of a user.” ’836 patent 27:23-25.

296. Consequently, limitation 8[a] would have been obvious to persons of skill in the art over Norris.

c. 8[b]: “one or more processors that process sound for the image”

297. Norris discloses, or at a minimum renders obvious, limitation 8[b]

298. Norris discloses a “processor unit” that includes “a processor (such as a central 20 processing unit, CPU, microprocessor, application-specific integrated circuit (ASIC), etc.) for controlling the overall operation of memory” and for “perform[ing] the operations and tasks that implement one or more blocks of [Norris’s] flow diagrams.” Norris at 41:10-60. Norris’s flow diagrams each relate to the provision and adjustment of 3D sound. *Id.* at Figs. 1-10.

299. Norris's claims also recite processors that execute instructions to, "change an interaural time difference (ITD), an interaural level difference (ILD), and a head-related transfer function (HRTF) in order to position the sound localization point of [a] binaural voice to [a] location that is away from and proximate to the first person," "position [a] sound localization point at [an] empty chair," and "change attenuation and reverberation of [] binaural sound in order to add artificial audial cues of virtual objects near [a] second person." *Id.* at cls. 15, 17-19. Each of these functions relate to processing sounds.

300. Norris also discloses that its sound processing is for accompanying virtual images. *See, e.g., id.* at 31:65-32:48 (describing an electronic system that includes a "wearable electronic device (WED)" such as "electronic glasses with a display" where "sound localization point 2050 can appear to the listener 2020 as a virtual image of the speaker 2010" where the "image represents the sound localization point 2050 so the listener 2020 can see the sound localization point 2050 as it moves in the room 2055"), Fig. 20, 40:59-41:9 (explaining the various components of an electronic device, including "one or more displays," "a processing unit," and an "imagery system" such as "a virtual image display system, virtual augmented reality system, and/or a spatial augmented reality system").

301. For example, Norris explains that the WED "can provide the listener with images and/or video that correspond to the band playing at the venue" such that

the listener “sees through a display of the wearable electronic device 1720 a real-time image of the band from a front row seat at the venue corresponding to the ticket purchased.” *Id.* at 30:37-43

302. From at least these disclosures, persons of skill in the art would have known and understood that Norris used its processors to process sound for an image. At a minimum, Norris’s disclosures would have rendered the claimed “one or more processors processing sound for [an] image” obvious.

303. Moreover, as I explained in Section VII.D above, in the State of the Art section of this declaration, the use of processors to process sound for images was routine and conventional before May 2017. *See* Section VII.D.

d. 8[c]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”

304. Norris discloses or at a minimum renders obvious, limitation 8[c].

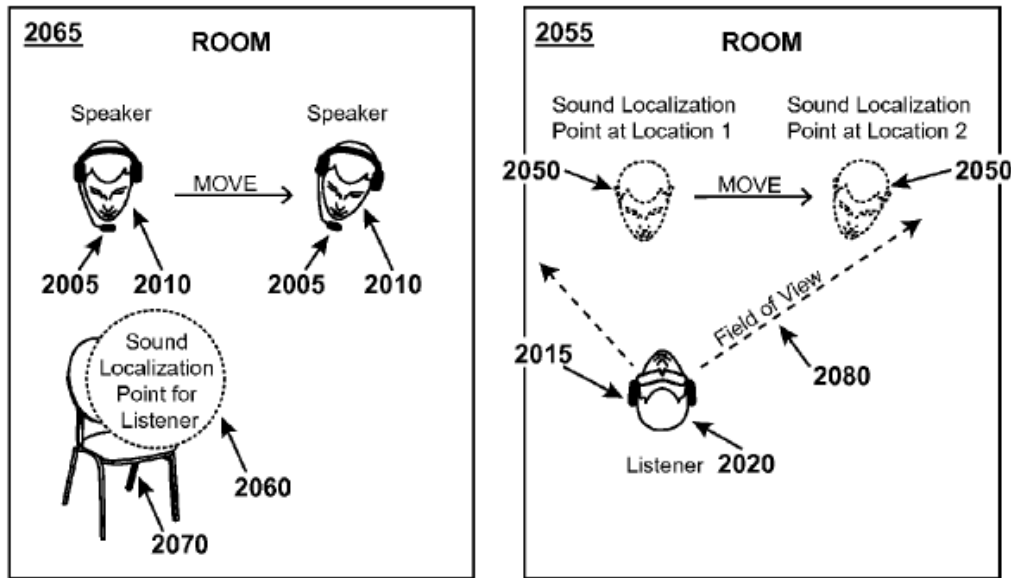
305. Norris discloses SLPs with coordinate locations that occur at far-field distances, i.e., at distances of “greater than or equal to one meter” (’836 patent at 27:23-25). *See* Norris at 20:51-21:6 (“voice of the first individual localized at a distance of ten feet”); 34:5-21 (“An electronic device of John establishes a sound localization point for the voice of Paul to be five feet in front of John”); 13:55-14:9 (“a voice of the third person localizes in a doorway that is fifteen (15) feet away”);

15:4-7 (“calls with an individual that occur in the morning are localized five feet from the individual, and calls that occur in the afternoon or evening are localized eight feet from the individual”), 15:41-42 (“a SLP defaults to a point in space that is five feet from a listener”), 15:52-54, 17:45-54 (“a sound localization point that is twenty (20) feet away from the listener”), 18:51-54, 20:56-59, 22:62-65, 34:5-21.

306. Norris further discloses that “in a Cartesian coordinate system,” “[t]he sound is adjusted to change one or more of the (X, Y, Z) coordinates” such that , “[f]or instance, when a listener is located at (0, 0, 0), an origin of the sound or the sound localization point for the listener appears to originate from (10, 17, 25).” *Id.* at 4:67-5: *see also id.* at 5:25-31 (“the binaural sound is processed to change the sound localization point from being two feet away from the listener to being thirty feet away”).

307. Additionally, persons of ordinary skill in the art would have known from Norris, and found it obvious, that such far-field SLPs would be behind their associated images, including images presented at near-field distances for reasons discussed in Section XI.D.1.b (limitation 8[a]), and along a user’s line of sight.

308. Norris’s Figure 20, reproduced below, depicts binaural sound behind an image displayed inches away from a user’s eyes and along a user’s line of sight:



Norris at Fig. 20 (excerpt).

309. Specifically, in Figure 20, a “first person speaker 2010” in room 2065 (on left) is wearing a headset with a microphone (WED 2005) and communicates with “a second person,” “listener 2020” in room 2055 (on right). *Id.* at 31:66-32:4. “The wearable electronic device 2005 of the speaker 2010 captures binaural sound and transmits sound to the listener 2020 who localizes the sound at a sound localization point 2050 located in [] room 2055.” *Id.* at 32:4-7. Conversely, “[t]he wearable electronic device 2015 of the listener 2020 captures binaural sound and transmits sound to the speaker 2010 who localizes the sound at a sound localization point 2060 located in [] room 2065.” *Id.* at 32:7-11. In room 2055, shown above on the right, listener 2020’s WED 2015 “is a pair of electronic glasses with a display

that provides an image or video of the speaker 2010 in a field of view 2080 of the listener 2020.” *Id.* at 32:40-48.

310. From these disclosures, it would have been obvious to a person of ordinary skill in the art that the “virtual image” is displayed on the lenses of listener 2020’s electronic glasses 2015, and that the SLP 2050 (where the user hears the sound coming from) occurs behind that image. And SLP 2050 occurs along (or obviously occurs along) listener 2020’s line of sight, as depicted above in the “field of view.” Norris at Fig. 20.

311. Persons of ordinary skill in the art would have known that this same configuration—an image displayed on the screen of an HMD that is located inches in front of a user’s eyes with the corresponding SLP located behind the image and along a user’s line of sight—are present in Norris’s other examples. For example, Norris describes “a music band” that “plays a live performance at a venue in New York City” such that “[b]inaural sounds of the band playing are sold to listeners that are located in other cities,” and “[l]istener[s] [] receive sound of the performance in real-time at a designated sound localization point that represents a location of a seat at the venue,” e.g., “a front row seat,” with WEDs that “provide the listener with images and/or video that correspond to the band playing at the venue.” Norris at 30:19-43; *see also id.* at 35:64-36:35 (discussing calls with an “interactive audio

input/output interfaces” and a “movie being watched on a screen” with SLPs corresponding to the images).

312. I also note that a person of ordinary skill in the art would have known that Norris’s disclosures of placing SLPs for voices at far-field locations could be applied to any of Norris’s systems or devices since Norris’s systems and devices all provide binaural sound to a user. And such person of skill in the art would have been motivated to do this, and would have expected to succeed in doing this, because, for example, doing so would have involved the same type of processing for localizing sound to other locations and would have involved no more than using similar processing in a similar configuration with only minor, straightforward differences.

e. 8[d]: “wherein the one or more processors include a digital signal processor (DSP)”

313. Norris discloses, or at a minimum renders obvious, limitation 8[d].

314. Norris, like many other references, discloses “digital signal processing (DSP) [that] provides sound to a listener through two speakers or earpieces.” Norris at 29:9-12. A person of ordinary skill in the art would have known and understood that “digital signal processing” for “provid[ing] sound to a listener” could and should be implemented with a digital signal processor (DSP) because DSPs were routinely used for processing 3D sound.

315. As I discussed in Section VII.D above, for years before May 2017, persons of ordinary skill in the art understood that DSPs can be advantageously used to process 3D sounds for virtual images. *See* Section VII.D. This was routine and straightforward for skilled artisans.

316. DSPs have been used for sound processing either in addition to or instead of other types of processors like CPUs because DSPs were known to offer benefits for such processing. *See* Begault at 100-101 (“Within a virtual reality or multimedia application, 3-D sound processing is almost always best accomplished digitally, using audio DSP chips to manipulate signal sources”; “The 3-D audio component of an integrated audio subsystem... can contain anywhere from one to as many DSPs as can be accommodated”), 101 (disclosing examples using DSPs to spatialize sound sources); Shivappa at 28:1-9, 28:21-24 (disclosing a “processor 710, such as a central processing unit (CPU) or a digital signal processor (DSP)” that is “configured to generate [an] output spatialized audio signal”); Sauk at 7:24-28 (disclosing a “sound environment manager” that “can be implemented by” “one or more digital signal processors.”); *see also supra* Section VII.D (collecting references describing DSPs and their uses).

317. Persons of ordinary skill in the art would have been motivated to implement DSPs in a Norris system because DSPs are fast. *See* Begault at 100 (explaining that DSP’s work “*very quickly*—i.e., in real-time”) (emphasis in

original); Otto at 13:50-53 (Explaining that its “DSP strategy” provides optimal performance in terms of directivity of the sound beam created and of the stability of the binaural rendering at higher frequencies”); Brown at Abstract. They excel at signal processing tasks like HRTF processing for 3D audio (i.e. binaural sound), and they are reconfigurable and thus advantageously flexible for different types of signal processing. Begault at 100 (explaining that “[o]ne can select a DSP that is most appropriate for a specific operation or type of digital filtering,” DSPs can be programmed with a “powerful instruction set,” and DSPs can be optimized for “specific complex operation[s]”); Qualcomm blog (“DSPs deliver high performance at lower clock frequencies, which saves power”). They are also appropriately sized (both in physical size and in power usage) for use in a system like Norris’s, which includes wearable electronic devices. *See* Qualcomm blog (“The DSP plays a key role in mobile heterogeneous computing. New applications offer breakthrough experiences and increase our engagement with our mobile devices. Many of these experiences—such as sound and image enhancements as well as advanced camera and sensor capabilities—include signal-processing tasks that the DSP excels in executing at low power. And saving power not only extends the battery life of the mobile device, it also keeps your phone cooler and provides a better user experience.”).

318. For these reasons, limitation 8[d] would have been obvious to persons of ordinary skill in the art based on Norris alone.

319. Additionally, limitation 8[d] would have been obvious to persons of ordinary skill in the art over Norris in view of Begault.

320. As noted, Begault describes and prescribes using DSPs for 3D audio processing. Begault at 100-101. And as discussed above, persons of ordinary skill in the art would have known that Begault's DSPs would be advantageous for processing 3D sound for images due to, among other things, their speed, reconfigurability, and size.

321. Persons of ordinary skill in the art, therefore, would have known and been motivated to implement a Norris-type system with a Begault DSP for sound processing, e.g., with a DSP replacing a Norris processor or being added to a Norris system for sound processing for images. And this would have rendered obvious both limitations 8[b] and 8[d] of the '836 patent.

322. Persons of ordinary skill in the art would have reasonably expected to succeed in using a DSP, e.g., as taught by Begault, in a Norris-type system because doing so would have required no more than combining known elements (e.g., HMDs and DSPs) and known functions (e.g., HRTF processing) through known methods. *See* Begault at 100-101 (“[w]ithin a virtual reality or multimedia application, 3-D sound processing is almost always best accomplished digitally, using audio DSP

chips” and explaining that “[t]he 3-D audio component of an integrated audio subsystem ... can contain anywhere from one to as many DSPs as can be accommodated”); Section VII.D; *see also* Section XI.A.

f. 8[e]: “that processes the sound for the image with far-field head-related transfer functions (HRTFs) while the electronic device is located at the near-field distance from the head of the user”

323. Norris (alone or with Begault) discloses, or at a minimum renders obvious, limitation 8[e].

324. Norris discloses processing sound for images with HRTFs, including for “far field distances greater than about one meter,” where a “far field HRTF” can be used. Norris at 28:24-32; *see also id.* at 5:25-31 (“binaural sound is processed to change the sound localization point from being two feet away from the listener to being thirty feet away”), 28:16-29:8 (describing HRTFs and providing an example where “[s]ound waves can [] be captured, processed, and altered with HRTFs to change sound properties before being provided to a listener,” so “the listener will believe that the sound originates from a location different than the real location”), 26:9-27 (describing “electronic glasses with speakers and microphones” where “HRIRs and HRTFs of the talking person are known” and “head orientations of the listening person are tracked such that an azimuth angle and a vertical angle can be

calculated between the mouth of the talking person (i.e. the origin of the sound impulses) and the head of the listening person.”).

325. From at least these disclosures, a person of ordinary skill would understand and find obvious that Norris’s disclosed sound processing is with “far-field” HRTFs through the many descriptions of localizing sound to distances “greater than or equal to one meter.”

326. Indeed, a person of ordinary skill in the art would have been motivated to use far-field HRTFs to process far-field and near-field sounds because doing so requires fewer stored HRTFs, and can leverage commonly available far-field HRTF databases. Moreover, a person of ordinary skill would understand that by using far-field HRTF databases, other audio cues can be changes to produce near-field effects for closer distances, e.g., by turning up the volume.

327. As explained in Section VII.B.3, not only are far-field HRTF databases more common and publicly available, but a person of ordinary skill in the art would have known that while far-field HRTFs are “essentially independent of range” the “situation is more complicated when the [sound] source [] is close to the head.” Ex. 1030 (Algazi 2011) at 34. Thus, a person of ordinary skill in the art would have known and found it obvious to use far-field HRTFs for processing sound for images displayed by Norris’s HMDs, including Norris’s electronic glasses, and including when the displays displaying images on their lenses were located inches from the

heads of users, i.e., at “near-field” distances. *See* Norris Fig. 20; THE VR BOOK Figs. 2.1 and 3.3 (showing HMDs and AR electronic glasses).

328. In addition, to the extent Patent Owner argues that Norris (alone or with Begault) does not disclose or render obvious limitation 8[e], it is my opinion that this limitation would have been obvious to a person of ordinary skill in the art over Norris (alone or with Begault) and Algazi.

329. As explained above, Norris disclosed that HRTFs from public databases can be used with its systems and methods, and Algazi describes and discloses one of the most well-known databases for such HRTFs—the CIPIC database. *See* Sections VIII.C-VIII.D. Algazi further discloses that in that CIPIC database “[a]ll HRTFs were measured with the subject seated at the center of a *1-m radius* hoop,” which is in the far-field. Algazi at 99.

330. Thus, a person of ordinary skill in the art would have known and understood that implementing Norris’s systems and methods using the CIPIC database HRTFs, consistent with Norris’s disclosures, would use “far-field” HRTFs, or HRTFs measured at distances of “greater than or equal to one meter” (’836 patent 27:23-25) as claimed.

331. A person of ordinary skill in the art would have been motivated to look at Algazi because it discloses CIPIC, one of the most well-known public HRTF databases. And as discussed above, a person of ordinary skill in the art would have

known that using the far-field CIPIC database could still be used for near-field sound sources by, for example, increasing the volume. *See* Rombloom at 1 (stating that “[w]hen binaural synthesis is chosen to simulate 3D sound sources over headphones, the Head-Related Transfer Functions (HRTF) used have most often been measured at a single distance in the far-field of the listener” and “it is possible to compensate the existing data using the well known law of ‘1/r spreading roll off’, effectively turning up the volume as you approach the head”). A person of ordinary skill in the art would also know and understand that using HRTF data from a public source and data that includes measurements from only one distance at the far field, like the CIPIC database, would require fewer stored HRTFs and would be easier than independently developing HRTFs for near-field sound sources. *Id.* (“Unfortunately, including a variety of near-field distances [for HTFT measurements] would substantially increase the required number of physical measurements.”); *see also* Section XI.B.

332. Thus, each limitation in claim 8, and claim 8 altogether, would have been obvious over Norris alone or in view of Begault and Algazi.

2. Dependent Claims 9, 10, 13, and 14

333. I understand that claims 9, 10, 13 and 14 depend on independent claim 8.

334. Thus, it is my opinion that dependent claims 9, 10, 13 and 14 would have been obvious to persons of ordinary skill in the art for the same reasons as independent claim 8. *See* Section XI.D.1.

335. In addition, claims 9, 10, 13 and 14 would have been obvious for the reasons I discuss in the following subsections.

a. Claim 9: “The electronic device of claim 8 , wherein the image is an augmented reality (AR) image, and the display is included with electronic glasses worn on the head of the user.”

336. Norris discloses, or at a minimum renders obvious claim 9.

337. As discussed above, Norris’s WEDs include “electronic eyeglasses or eye wear” “such as glasses in which augmented reality imagery” (i.e. an augmented reality (AR) image) is “projected through or reflected off a surface of a lens,” i.e. displayed. Norris at 41:48-59; *see also id.* at 32:42-48 (WED 2015 is “a pair of electronic glasses with a display that provides an image or video of the speaker 2010 in a field of view 2080 of the listener 2020.” Norris further discloses a “virtual augmented reality system” in an electronic device. *Id.* at 40:59-41:9 (describing “Electronic device 2604” that includes “one or more displays 2262,” and “an imagery system 2676 (such as an optical projection system, a virtual image display system, virtual augmented reality system, and/or a spatial augmented reality system)”), Fig. 20.

b. Claim 10: “The electronic device of claim 8 , wherein the image is a virtual reality (VR) image, and the display is included with a head mounted display worn on the head of the user.”

338. Norris discloses, or at a minimum renders obvious, claim 10.

339. To begin, Norris discloses “head-mounted displays” that are “worn on” a user’s head. Norris at 41:48-59. And as explained above, Norris discloses an “imagery system” “such as an optical projection system, a virtual image display system, virtual augmented reality system, and/or a spatial augmented reality system” for use with such HMDs. *Id.* at 40:59-41:9. Norris also describes electronic calls between “first and second individuals” in “separate virtual realities,” and teaches location and head orientation tracking that can be used for “an augmented or virtual reality application.” *Id.* at 6:29-45, 9:12-15. Thus, a person of ordinary skill in the art would have known to and would have been motivated to implement Norris’s systems and methods with HMDs displaying VR images.

c. Claim 13: “The electronic device of claim 8, wherein the image is an augmented reality (AR) image of an intelligent user assistant (IUA) that is located less than one meter away from the head of the user, and a voice of the IUA occurs at a coordinate location behind the AR image of the IUA.”

340. Norris discloses, or at a minimum renders obvious, claim 13.

341. As I discussed above with respect to claims 8 and 9, Norris discloses AR images located less than one meter from the head of a user (e.g., images on AR HMD displays or lenses) and SLPs located behind their associated AR images. *See* Sections XI.D.1 (claim 8) and XI.D.2.a (claim 9). In addition, Norris discloses that its systems and methods can utilize “an intelligent personal assistant.” Norris at 41:33-35.

342. A person of ordinary skill in the art would have known and understood that Norris’s “intelligent personal assistant” (the claimed IUA) could vocalize and be presented as an AR image. Indeed, when describing SLPs and empty space, Norris describes a “virtual object, or other manifestation or representation of a remote real or virtual character or object.” *Id.* at 4:42-47.

343. A person of ordinary skill in the art would have known and recognized that an intelligent personal assistant is one type of virtual character in an AR setting. *See* U.S. Patent No. 9,392,212 (“Ross”) (Ex. 1051) at 17:32-18:9 (explaining that a guiding character and/or avatar that “walks the user through a scene in the virtual space and points things out” is “like a commentary track on a DVD”); 14:53-15:1 (describing “content component 24” of a VR headset that allows the display “to present primary virtual reality content” where the primary virtual reality content may include “a behind the scenes tour (e.g., with the content creator physically walking a user through the scenes in the virtual space as an avatar and/or performance capture of some kind.”). Indeed, characters used as personal assistants have been around for decades, including Clippit, the paperclip default assistant used in Microsoft Word in the 90s and early 2000s. *See* Ex. 1070, Sutter, J., *Microsoft’s talking paper clip is back* (April 29, 2011), available at <http://www.cnn.com/2011/TECH/web/04/29/microsoft.paper.clip/index.html>) and video game characters such as the Lakito Bros in Super Mario, who help the main

character navigate throughout the game world. *See* Ex. 1071, *Super Mario 64*, available at <https://www.giantbomb.com/super-mario-64/3030-2931/characters/>.

Thus, a person of ordinary skill in the art would have had a reasonable expectation of success in implementing such characters in an AR environment because that would involve using the same straightforward techniques used for rendering other virtual objects in such an environment.

d. Claim 14: “The electronic device of claim 8, wherein the image is a virtual reality (VR) image of an intelligent user assistant (IUA) that is located less than one meter away from the head of the user, and a voice of the IUA occurs at a coordinate location behind the VR image of the IUA.”

344. Norris discloses, or at a minimum renders obvious, claim 14.

345. As I discussed above with respect to claims 8 and 10, Norris discloses to a person of ordinary skill in the art VR images located less than one meter from the head of a user (e.g., images on VR HMD displays or lenses) and SLPs located behind their associated VR images. *See* Sections XI.D.1 (claim 8) and XI.D.2.b (claim 10). In addition, as I explained above for claim 13, Norris discloses that its systems and methods can utilize “an intelligent personal assistant.” Norris at 41:33-35; *see* Section XI.D.2.c. A person of ordinary skill in the art would have known and understood that Norris’s “intelligent personal assistant” (the claimed IUA) could vocalize and be presented as a VR image because Norris describes a “virtual object,

or other manifestation or representation of a remote real or virtual character or object.”

Id. at 4:42-47.

346. For these reasons, each limitation in claims 9, 10, 13, and 14, and each of claims 9, 10, 13, and 14 as a whole, would have been obvious over Norris, alone or in view of Begault and Algazi.

3. Independent Claim 15

a. 15[pre]: “An electronic system, comprising”

347. The preamble of claim 15 would have been obvious for the same reasons as the preamble of claim 8, i.e., because Norris discloses an “electronic system” that includes a wearable electronic device. Norris at 2:8-11; *see also id.* at 2:33-35 (describing an “electronic system in which electronic devices of two individuals communicate during an electronic call”), 2:40-46, 40:10-41:19, Figs. 17, 25-26; Section XI.D.1.a.

b. 15[a]: “a display that displays an image that is located within a near-field distance from a head of a user viewing the image”

348. Limitation 15[a] would have been obvious for the same reasons as limitation 8[a]. *See* Section XI.D.1.b. Further, a person of ordinary skill in the art would understand that a user would be viewing the image displayed by display within a near-field distance from a head, i.e. a HMD worn a few inches from the user’s head.

c. 15[b]: “one or more processors that process sound for the image”

349. Limitation 15[b] would have been obvious for the same reasons as limitation 8[b]. *See* Section XI.D.1.c.

d. 15[c]: “with head-related transfer functions (HRTFs) to generate binaural sound”

350. Limitation 15[c] would have been obvious for the same reasons as limitation 8[c]. *See* Section XI.D.1.f. In addition, a person of ordinary skill in the art would have known that the HRTFs are used to generate binaural sound for the same reasons as discussed for limitation 8[c]. *See* Section XI.D.1.d.

e. 15[d]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”

351. This limitation would have been obvious for the same reasons as limitation 8[c]. *See* Section XI.D.1.d.

f. 15[e]: “wherein the coordinate location of the binaural sound continues to occur at the far-field distance located behind the image when the head of the user moves away from viewing the image and the image remains within the near-field distance from the head of the user.”

352. Norris discloses, or at a minimum renders obvious, limitation 15[e] for reasons discussed above. *See* Sections XI.D.1 (claim 8), XI.D.3.a-XI.D.3.e (15[pre]-15[d]).

353. Further, Norris discloses a processor that executes instructions to “change an interaural time difference, an interaural level difference, and a head-

related transfer function of [] binaural sound in order to maintain [a] sound localization point fixed at [a] location” “while the first person physically moves away from the location and changes a head orientation with respect to the location.” Norris at cl. 18, *see also id.* cl. 8 (similar). Norris also discloses that “[m]ultiple sound localization points [can] exist in [a] room, and binaural synthesis of sound dynamically occurs in real-time as the listener moves his head and/or body.” *Id.* at 8:50-53. Thus, as a “listener walks around the room, the sound localization points remain fixed or unmoved to provide virtual sound sources located in the room.” *Id.* at 8:53-55. That way, “[i]f the virtual sound source represents a fixed object, then the virtual sound source remains fixed in the room.” *Id.* at 8:55-57. “For instance, sound that emanates from a virtual water fountain is a fixed sound source since the virtual water fountain does not move.” *Id.* at 8:57-59.

354. Accordingly, based on Norris alone, persons of ordinary skill in the art would have known that Norris’s systems and methods could and should maintain sound occurring at the far-field distance for an image displayed on a near-field lens Norris’s AR or VR HMDs, e.g., the “virtual water fountain,” even while a user’s head moves away from viewing that image so that the user can continue to determine the sound source’s location. And persons of ordinary skill in the art would have known how to implement Norris’s disclosures of various sensors to detect the user’s location (*see e.g., id.* at 6:16-45) and orientation with Norris’s disclosure of designating a

sound localization point to simulate a location of a source of sound (*see, e.g., id.* at 5:45-57) such that the image, e.g., the “virtual water fountain,” will remain “within the near-field distance from the head of the user” until the user’s head rotation causes it to drop from the user’s field of view and the HMD display. Indeed, Fig. 2 of Norris describes a “method to adjust sound as a listener moves with respect to a sound localization point.” *Id.*

355. Thus, each limitation in claim 15 of the ’836 patent, and claim 15 as a whole, would have been obvious over Norris, alone or in view of Begault and Algazi.

4. Dependent Claims 16 and 17

356. I understand that claims 16 and 17 depend on independent claim 15.

357. Thus, it is my opinion that dependent claims 16 and 17 would have been obvious to persons of ordinary skill in the art for the same reasons as independent claim 15. *See* Section XI.D.3.

358. In addition, claims 16 and 17 would have been obvious for the reasons I discuss in the following subsections.

a. Claim 16: “The electronic system of claim 15, wherein the display is an augmented reality (AR) display worn on the head of the user viewing the image, and the image is an AR image.”

359. Claim 16 would have been obvious for the same reasons as Claim 9. *See* Section XI.D.2.a. Further, a person of ordinary skill in the art would understand that

a display that displays an AR image would be an AR display and that a user would be viewing the AR image when the AR display, i.e. an HMD, is worn on the user's head.

b. Claim 17: “The electronic system of claim 15, wherein the display is a virtual reality (VR) display worn on the head of the user viewing the image, and the image is an VR image.”

360. Claim 17 would have been obvious for the same reasons as Claim 10. *See* Section XI.D.2.b.

361. For these reasons, claims 16 and 17 of the '836 patent would have been obvious over Norris, alone or in view of Begault and Algazi. Further, a person of ordinary skill in the art would understand that a display that displays an VR image would be a VR display and that a user would be viewing the VR image when the VR display, i.e. an HMD, is worn on the user's head.

E. Ground 2(b): Dependent Claims 11 and 12 are Unpatentable Over Norris, Alone or in View of Begault and Algazi, and Further in View of Sauk

362. In this section, I explain why dependent claims 11 and 12 of the '836 patent would have been obvious to a person of ordinary skill in the art over Norris, alone or in view of Begault and Algazi, and further in view of Sauk. *See also* Section XI.C (Motivations to combine Sauk).

1. Claim 11: “The electronic device of claim 8 further comprising: one or more sensors that detect rotation of the

electronic device, wherein the binaural sound automatically switches to stereo sound in response to the one or more sensors detecting the rotation of the electronic device”

363. Norris alone or in view of Begault and Algazi and in further view of Sauk renders claim 11 obvious.

364. Norris discloses sensors that detect rotation, including a “head orientation tracker” that “measures one or more rotational head orientations of yaw (side-to-side or left and right), pitch (up and down), and roll (tilting) of the head.” Norris at 6:42-45, Fig. 26 (Sensors 2630), 40:4-58 (describing Figure 26 and sensors 2630); *see also id.* at 6:29-42 (explaining that measurements can be made with “a system tracking angular head velocity and/or acceleration, a camera (including a camera in a computer or an augmented or virtual reality application), a tag (such as a radio, radio frequency identification (RFID) or GPS tag system), binaural headsets or ear pieces (for example sensing interaural level differences and/or interaural time differences), a laser, a sensor (such as a sensor located in a smartphone on which a person speaking), a gyroscope, a magnetometer, and an accelerometer”).

365. Norris also taught that a user can arrange SLPs “in any order and/or move any of them to the foreground (such as bringing the associated SLP toward him to make the sound louder), moving any of them to the background (such as bringing the SLP away from him to make the sound softer), and moving any of them to

increase or decrease priority or access to resources (such as computer processing power, resolution, bandwidth, etc.).” *Id.* at 36:17-25.

366. Thus, based on Norris’s disclosures, a person of ordinary skill in the art would have known and been motivated to prioritize different sounds (such as those directly in front of the user) as well as the processing of those sounds based on SLP location, including by degrading sound to “decrease priority or access to resources” where appropriate. *Id.* Sauk discloses a useful way to do this by explaining that a user could “change the less relevant audio to a monophonic (1-D) or stereophonic (2-D) dimension,” to “signify a different level of relevancy or importance for such audio” and “remove any localization cues for” the less relevant audio. Sauk at 11:25-40; *see also* Korean Patent No. KR20110022252A (Ex. 1067) at claim 1 (discussing the removal of time delays from multichannel (spatialized) audio to produce a single channel (monophonic) audio).

367. Further, a person of ordinary skill in the art would have known and been motivated to degrade binaural audio, such as by automatically switching that sound to stereo sound, as a WED is rotated. Consistent with Sauk’s teachings, using one or more sensors to detect motion of the WED and degrading binaural sound to stereo sound for images away from which a user is rotating her view would indicate that the sound was (or was becoming) less relevant. *See* Sauk at 11:26-41. A person of ordinary skill in the art would further understand that by degrading the sound to stereo,

a user would still hear the less relevant audio outside of the field-of-view but would focus on the binaural audio that is in the field-of-view of the display.

368. In addition, a person of ordinary skill in the art would have understood that degrading sound that is out of the field-of-view from binaural to stereo could conserve processing power, and by extension the battery life of a HMD, both powerful motivations for 3D audio and system engineers. All the while, such degradation would not impact the user's ability to recognize where the sound source was located. For example, if the WED is rotated such that a sound source that was originally directly in front of the user is now directly to the user's right ear, degrading the sound source to stereo would still allow the user to recognize that the sound is coming from the user's right side.

2. Claim 12: "The electronic device of claim 8, wherein the binaural sound automatically switches to stereo sound in response to the electronic device detecting a change to a head orientation of the user that exceeds a threshold value in an azimuth direction."

369. Claim 12 would have been obvious over Norris, alone or in view of Begault and Algazi, and further in view of Sauk for the same reasons as claim 11. *See* Section XI.E.1.

370. Just as a person of ordinary skill in the art would have known and understood that degrading binaural sound to stereo sound for images away from which a user is rotating her view would helpfully indicate that such audio was (or

was becoming) less relevant and could conserve resources, a person of ordinary skill in the art would have known and understood that degrading binaural sound in the same manner when a user rotates past a threshold value in an azimuth direction, for example, such that a sound source is directly to the left or right of the user's head—which would cause some images and sound sources to rotate out of view and off a HMD display—would have the same benefits.

371. For these reasons, claims 11 and 12 of the '836 patent would have been obvious over Norris, alone or in view of Begault and Algazi, and further in view of Sauk.

XII. GROUND 3: CLAIMS 8-10 AND 15-17 ARE UNPATENTABLE OVER SHIVAPPA, ALONE OR IN VIEW OF ZHOU, AND FURTHER IN VIEW OF SAUK

A. Reasons Persons of Ordinary Skill in the Art Would Combine Shivappa and Zhou

372. A person of ordinary skill in the art would have known and been motivated to combine Shivappa's systems and methods, which process HRTFs to produce binaural sound, with Zhou's methods of using spatialized 3D audio in head-mounted display and augmented reality devices.

373. A person of ordinary skill in the art would have recognized that Zhou's disclosure of providing 3D sound by using HRTFs "improve[] the accuracy of depth judgment." Zhou at Abstract, 991. Indeed, as explained above in the State of the Art

section, it was known that 3D sound allows a user to better recognize the distance (or depth) of a visual stimulus on a screen. persons of ordinary skill in the art would have known that Zhou's 3D audio could complement user experiences, including in the AR and VR systems that Shivappa focuses on, because, as Zhou teaches, 3D audio helps with depth perception, task performance, and user collaboration (where applicable), and it enables a more realistic environment and immersive feeling. Zhou at 1014; Shivappa at 3:16-19.

374. Further, a person of ordinary skill in the art combining Shivappa's systems and methods with Zhou's teachings would have had a reasonable expectation of success. As explained above by in the State of the Art section, HRTFs were frequently and routinely used by persons of ordinary skill in the art for producing binaural sound in 3D audio applications before May 2017. *See* Section VII.B. And Zhou's use of these HRTFs in a head-mounted display would have been well within a person of ordinary skill in the art's common sense and skill and would have required no more than combining prior art elements (e.g., HMDs, HRTFs, and open source audio SDKs) using known methods. *See* X.A (explaining reasons why a person of ordinary skill in the art would combine a Finnegan that discloses HMDs with Algazi, which discloses the CIPIC HRTF database). Indeed, Zhou itself used an HMD and thus, using the HRTFs and audio SDKs described in Zhou for use in another HMD

would have been within the common knowledge of persons of ordinary skill in the art.

B. Reasons Persons of Ordinary Skill in the Art Would Combine Shivappa, Zhou, and Sauk

375. A person of ordinary skill in the art also would have known and been motivated to combine the Shivappa's teachings, alone or with Zhou. Sauk, like Shivappa, proposes helpful ways for improving 3D audio delivery to a user, including in AR, VR, and mixed-reality settings. *See, e.g.*, Shivappa at 3:16-19 ("In a particular implementation, the audio processing device may be integrated in a virtual reality (VR) system or augmented reality (AR) system"); Sauk at 2:45-47 ("The invention concerns a method and apparatus for producing, combining, and customizing virtual sound environments to provide the user with understandable information regarding their surroundings."). Indeed, Sauk describes "innovative applications for actually putting [] binaural audio information to practical use." Sauk at 2:34-41.

376. A person of ordinary skill in the art would have recognized before May 2017 that Sauk's methods of switching from binaural to stereo or monophonic sound based on the relevance of an audio signal could enhance user experience and perceptual accuracy, all the while lowering processing requirements. *See* Begault at 103 ("DSP algorithms are usually complex combinations of simple elements"); *see also id.* at 95 ("perceptual requirements must be balanced against available

resources”); Algazi 2011 (explaining that “computational requirements” of a device that delivers binaural sound “depend[s] on the “complexity of the auditory scene” and “the efficiency” of the algorithm implementation, and “[t]he approach to these issues depends on the application and on the decision as to what constitutes an acceptable auditory experience, as opposed to one that is indistinguishable from actually being present.” *Id.* Thus, degrading binaural sound for less relevant sound sources would have been obvious to a person of ordinary skill in the art in order to decrease the processing power required to attain these benefits.

377. Further, a person of ordinary skill in the art would have had a reasonable expectation of success in combining Shivappa, alone or with Zhou, and Sauk. Like Shivappa and Zhou, Sauk disclosed ways to enhance binaural sound implemented on wearable electronic devices, including through the use of HRTFs. Similarly, Shivappa and Sauk both use head and position tracking on wearable electronic devices and HRTF processing in analogous ways to achieve immersive, binaural sound for the user. Sauk at 4:33-5:5 (describing position and head tracking), 7:24-8:33 (describing sound processing with HRTFs); Shivappa at 3:60-64 (describing position and orientation tracking) 7:53-8:4 (describing sound processing with HRTFs).

C. Ground 3(a): Claims 8-10 and 15-17 are Unpatentable Over Shivappa, Alone or in View of Zhou

378. In this section, I explain why every limitation of claims 8-10 and 15-17 of the '836 patent, and each of claims 8-10 and 13-17 as a whole, would have been obvious to a person of ordinary skill in the art over Shivappa, alone or in view of Zhou.

1. Independent Claim 8

a. 8[pre]: “An electronic device, comprising:”

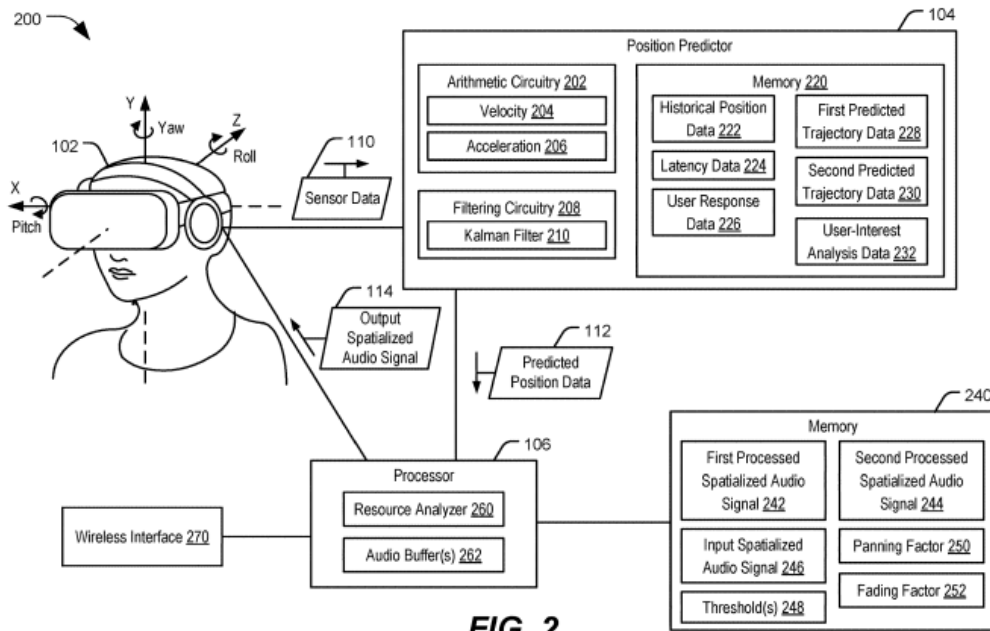
379. Shivappa teaches the preamble of claim 8. At a minimum, Shivappa renders the preamble obvious.

380. Shivappa’s systems, devices, and methods comprise electronic devices. *See, e.g.*, Shivappa at 5:37-6:3 (“the audio device 102 includes (or is integrated within) a head mounted display (HMD) of a virtual reality (VR) system or an augmented reality (AR) system”), 24:49-64 (“the audio processing device may be configured to process object-based 3D audio (e.g., audio having audio sources represented as individual objects with 3D coordinates and directivity).”), Figs. 1, 2, 7.

b. 8[a]: “a display that displays an image that occurs within a near-field distance from a head of a user”

381. Shivappa teaches, or at a minimum renders obvious, limitation 8[a].

382. Shivappa discloses multiple display devices, including “a head mounted display (HMD) of a virtual reality (VR) system or an augmented reality (AR system)” with “a display screen for displaying visual information”:



Id. Fig. 2, at 5:37-53 (“audio device 102 includes (or is integrated within) a head mounted display (HMD) or a virtual reality (VR) system or an augmented reality (AR) system”); *see also id.* at Fig. 7, 28:25-26 (“display 728”).

383. Shivappa’s devices display images within a “near-field distance from a head of a user” from the user’s head, i.e. a distance “less than one meter” from a user’s head. *Id.* at 5:42-45 (“The VR system (or the AR system) may be configured to display media content, such as movies, and to provide interactive content, such as video games, presentations, virtual meetings, etc.”); *see* ’836 patent at 27:23-25.

c. 8[b]: “one or more processors that process sound for the image”

384. Shivappa teaches, or at a minimum renders obvious, limitation 8[b].

385. Shivappa discloses an “audio processing device” that “provides spatialized audio signals to an audio device” that “may be integrated in a virtual reality (VR) system or augmented reality (AR) system.” Shivappa at 3:16-18, 3:60-4:19 (describing an “audio processing device”).

386. Shivappa further discloses a processor that “perform[s] 3D sound rendering to generate a 3D output audio signal,” including through “binauralization” and “head-related transfer function (HRTF) filtering,” to “enable a user to perceive a direction or a distance of sounds in a 3D audio space relative to a location of a user, either in a game (or other virtual reality environment) or in reality.” *Id.* at 3:19-23 (describing an “audio processing device” that “may be integrated” into VR or AR systems), 7:53-8:4 (describing a “processor” “configured to generate an output spatialized audio signal”), 5:37-43 (describing an “audio device” integrated into an HMD containing headphones “for playing audio in addition to a display screen for displaying visual information”).

387. For example, Shivappa describes a scenario where a user can “play[] a game” where a virtual image of a car “drives to the left of the user in the game,” and the “audio output enables the user to perceive a sound of a car as coming from the

user's left side.” *Id.* at 3:23-26, *see also id.* at Fig. 7 (“Processor (e.g., DSP)”), 5:45-48 (“The visuals and the audio output by the virtual reality system may change based on movement of the user (e.g., movement of the user's head).”). As another example, Shivappa's claims also recite processors that, among other things, “adjust an orientation of the 3D sound field” and “output the binaural audio signal ... for playback,” including where the audio is “associated with a virtual reality or augmented reality application” and a “display is configured to display visual information associated with the virtual reality or augmented reality application.” *Id.* at cls. 1, 6.

d. 8[c]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”

388. Shivappa teaches, or at a minimum renders obvious, limitation 8[c].

389. Shivappa discloses virtual images and associated binaural sound sources, such as doors, characters speaking in a movie, instruments at a concert, conference speakers, and vehicles that a person of ordinary skill in the art would have known to be located at far-field distances, i.e., at distances of “greater than or equal to one meter” (*see* '836 patent at 27:23-25) from a user. Shivappa at 8:42-47 (“As non-limiting examples, [] a sound source may be a character speaking in a movie, an instrument playing in a concert, a person speaking during a teleconference, a vehicle

making noise in a virtual reality video game, or any other source of sound corresponding to the input spatialized audio signal.”).

390. Further, Shivappa discloses that with 3D sound rendering, a user may perceive both direction and *distance* of a sound source. *Id.* at 5:30-36 (“Due to the 3D rendering, a user may perceive the auditory sound as being in 3D, which may enable the user to perceive direction, distance, or both of one or more sound sources corresponding to the auditory sound. For example, a user may perceive a sound of a door opening to their right ... for an auditory sound of an opening door.”),

391. A person of ordinary skill in the art also would have known and understood that Shivappa’s sound sources occur behind their associated images presented at near-field distances (e.g., on a HMD’s lenses, as discussed in Section VII.A), as well as along a user’s line of sight, because the user can look in the direction of the door, character, instrument, conference speaker, or vehicle being heard and displayed. *See, e.g.*, Shivappa at 3:14-31 (The audio outputs may be three-dimensional (3D) audio outputs that enable a user to perceive a direction or a distance of sounds in a 3D audio space relative to a location of a user, either in a game (or other virtual reality environment) or in reality. For example, if the user is playing a game and a car drives to the left of the user in the game, the audio output enables the user to perceive a sound of a car as coming from the user’s left side.’), 5:42-54 (“The visuals and audio output by the virtual reality system may change based on movement

of the user (e.g., movement of the user's head).”), 24:49-53 (“As a particular example, the audio processing device may be configured to process object-based 3D audio (e.g., audio having audio sources represented as individual objects with 3D coordinates and directivity).”). Similarly, the sound must be behind the image on the lenses in order for the user to perceive the sound's distance. *Id.* at 3:14-31 (“The audio outputs may be three-dimensional (3D) audio outputs that enable a user to perceive a direction or a distance of sounds in a 3D audio space relative to a location of a user”).

392. To the extent Patent Owner argues that Shivappa does not disclose or render obvious limitation 8[c] on its own, it is my opinion that this limitation would have been obvious to a person of ordinary skill in the art in view of Zhou.

393. Zhou discloses “generat[ing] 3D sound” using open source software development kits (SDKs) that “have already incorporated the general HRTFs database and simple model of environment to generate 3D sound” Zhou at 991, 992-993. As Zhou explains, “[a]part from its effects for the user on sound source localization, we tried to emphasize [the 3D sound generating by “applying HRTFs”]’s effect on telling the depth (distance).” *Id.* at 1013. Zhou teaches placing a virtual object containing a virtual sound in the image of a 3D scene such that the sound source is behind the screen and along a line-of-sight from the user and the screen, as shown below in Fig. 1. *Id.* at 994-995, Fig. 1.

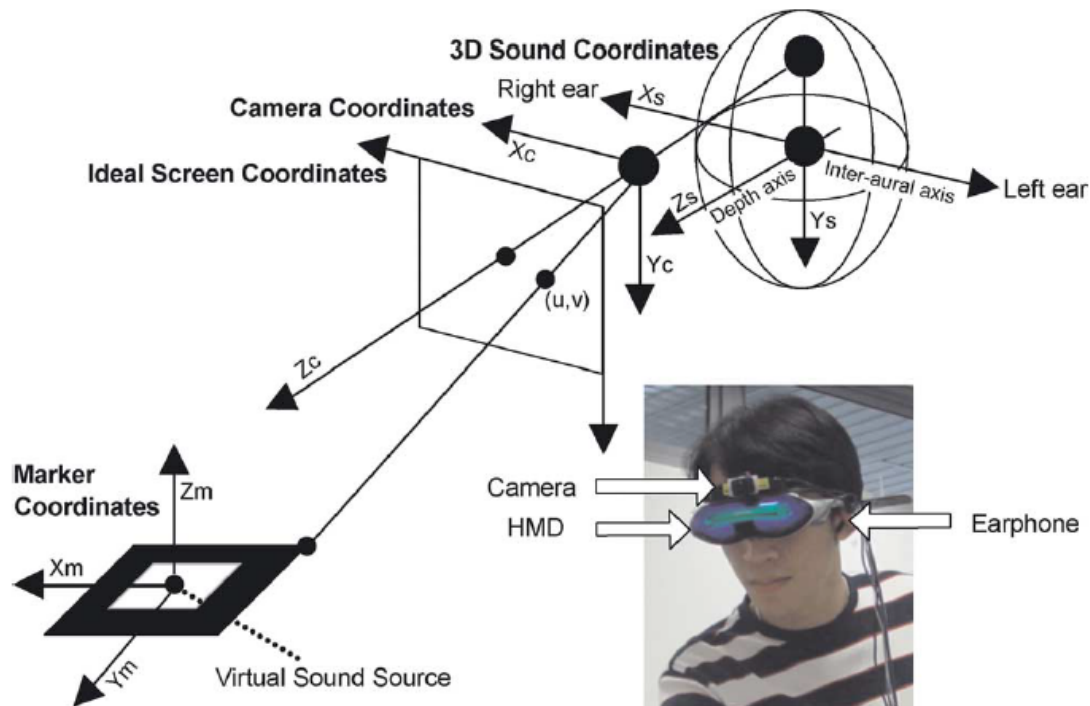


Fig. 1. In order to place a virtual object in the image of the three dimensional scene, the Euclidean transformation between the camera co-ordinates (X_c , Y_c , Z_c) and the marker co-ordinates (X_m , Y_m , Z_m) is estimated. This calculation is based on the imaged position of the corners of the square marker. The imaged position of one of these corners is depicted with pixel co-ordinates (u, v). To obtain the transformation matrix from the 3D sound co-ordinates to marker co-ordinates, translations should be taken from the camera co-ordinates to the 3D sound co-ordinates (X_s, Y_s, Z_s). The picture at the right bottom shows the physical setting of our single camera system where camera is mounted to the front of HMD and 3D sound plays back into the earphones.

394. As Fig. 1 shows, the virtual sound source is located at a distance behind the screen from the user. The location of the virtual object in the image on a 3D screen is determined using a transformation matrix from the 3D sound coordinates to the marker coordinates.

395. As an example, in its first experiment, Zhou's SDK applied HRTFs to place sound sources "around 1.5 m away" from a listener, which is a far-field distance, along the listener's line of sight and while images associated with the sound sources

were presented to the listener on the lenses of Zhou's HMD, i.e., at a near-field distance. *Id.* at 993, 1013.

396. Similarly, in a second experiment, Zhou tracked co-players "walk[ing] around in a three by three meter region," and placed sounds at the locations of those co-players' controllers (their "wands"). *Id.* at 997-98, 1006-07. Persons of ordinary skill in the art would have known that such co-players could be separated by distances of one meter or more, i.e., far-field distances, in a nine square meter play area. Thus, in this scenario, sounds (e.g., the "buzz" of a co-player's airplane) were positioned at far-field distances (on co-player wands separated by a meter or more) while images were positioned at near-field distances (e.g., the image of an airplane on the lenses of a first co-player's HMD), including while co-players were in each other's lines of sight.

e. 8[d]: "wherein the one or more processors include a digital signal processor (DSP)"

397. Shivappa teaches, or at a minimum renders obvious, limitation 8[d].

398. Shivappa, like many other references, discloses using "a digital signal processor (DSP)" for its audio processing. Shivappa at 28:1-3; *see also id.* at 28:21-24 (disclosing a "processor 710, such as a central processing unit (CPU) or a digital signal processor (DSP)" that is "configured to generate [an] output spatialized audio signal"), 28:4-5 ("The processor 710 may include or correspond to the processor 106

of FIGS. 1A-C, 2, and 3.”), Fig. 7 (“Processor (e.g., DSP)”), 7:40-44 (“The processor 106 may include one or more processors or processing units, a digital signal processor (DSP), ...”), 13:40-45 (“Each component illustrated in FIGS. 1A-C may be implemented using hardware,” e.g., “a DSP,” “software,” “or a combination thereof.”), 21:34-39 (similar), 22:39-44 (similar), 25:11-16 (similar).

f. 8[e]: “that processes the sound for the image with far-field head-related transfer functions (HRTFs) while the electronic device is located at the near-field distance from the head of the user”

399. Shivappa teaches, or at a minimum renders obvious, limitation 8[e].

400. As I explained above, Shivappa teaches using a DSP for binaural audio processing. *See* Shivappa at 28:1-5, 28:21-24 (disclosing a “processor 710, such as a central processing unit (CPU) or a digital signal processor (DSP)” that is “configured to generate [an] output spatialized audio signal”); *see also* Section XII.C.1.e (limitation 8[d]).

401. As I also explained above, Shivappa discloses processing sound for virtual sound sources that a person of ordinary skill in the art would have known would be located at far-field distances, i.e., at distances of “greater than or equal to one meter” (’836 patent 27:23-25), and thus processed with far-field HRTFs. *See* Section XII.C.1.d; *see also* Shivappa at 23:46-64. Shivappa also discloses display devices that are located at near-field distances from the heads of users when used,

e.g., an HMD worn on a user's head. *See* Section XII.C.1.b; Shivappa at Fig. 2.

402. To the extent Patent Owner argues that Shivappa does not disclose or render obvious limitation 8[e] on its own, it is my opinion that this limitation would have been obvious to a person of ordinary skill in the art in further view of Zhou.

403. Zhou disclosed advantages of 3D audio in AR and VR environments, i.e., the environments on which Shivappa focuses, and taught a system that a person of ordinary skill in the art would have known and been motivated to use to “significantly improve[] the accuracy of depth judgment and improve[] task performance,” as well as to “contribute[] significantly to the feeling of ‘human presence and collaboration’” by adding 3D sound the “the general HRTFs database.” Zhou at 989-991. A person of ordinary skill in the art would have known that the HRTFs in the general HRTFs database that Zhou describes could or would be measured at far-field distances because that is how most HRTFs are captured for the most commonly available public databases. *See, e.g.*, Romblom at 4-6 (describing “commonly available far-field databases” for HRTFs); Algazi at 99 (describing “a public-domain database” of “measured with the subject seated at the center of a 1-m radius hoop”). Thus, a person of ordinary skill in the art implementing Zhou's methods in a Shivappa's HMD-based system with the components described above, would have known and been motivated to process sound for an image using a database of far-field HRTFs with an electronic device located at a near-field distance

from the head of a user. Such an implementation would require no more than using well-known preexisting components (i.e. HMDs and far-field HRTF databases) and the combination yields no more than would be expected.

404. Thus, limitation 8[e] would have been obvious to persons of ordinary skill in the art over Shivappa in view of Zhou.

405. And for these reasons, each limitation in claim 8, and claim 8 altogether, would have been obvious over Shivappa alone or in view of Zhou.

2. Dependent Claims 9 and 10

406. I understand that claims 9 and 10 depend on independent claim 8.

407. Thus, it is my opinion that dependent claims 9 and 10 would have been obvious to persons of ordinary skill in the art for the same reasons as independent claim 8. *See* Section XII.C.1.

408. In addition, claims 9 and 10 would have been obvious for the reasons I discuss in the following subsections.

a. Claim 9: “The electronic device of claim 8 , wherein the image is an augmented reality (AR) image, and the display is included with electronic glasses worn on the head of the user.”

409. Claim 9 would have been obvious of Shivappa.

410. Shivappa discloses that its audio processing system can be used in and integrated with “a head mounted display (HMD) of a virtual reality (VR) system or

an augmented reality (AR) system” with “a display screen for displaying visual information.” Shivappa at 5:37-42. Shivappa also discloses that its HMD “may be a pair of virtual reality goggles.” *Id.* at 13:64.

411. These disclosures would have rendered claim 9 obvious to a person of ordinary skill in the art, including claim 9’s “electronic glasses,” which like “goggles,” were a well-known variant of VR and AR HMDs. *See* THE VR BOOK, Figs. 2.1, 3.3.

b. Claim 10: “The electronic device of claim 8 , wherein the image is a virtual reality (VR) image, and the display is included with a head mounted display worn on the head of the user.”

412. Claim 10 would have been obvious of Shivappa in view of Zhou.

413. As I already explained above for claim 9 (*see* Section XII.C.2.a), Shivappa discloses that its audio processing system can be used in and integrated with “a head mounted display (HMD) of a virtual reality (VR) system or an augmented reality (AR) system” with “a display screen for displaying visual information.” Shivappa at 5:37-42; *see also id.* at 3:14-23, 13:46-67 (“Referring to FIG. 2, an illustrative implementation of an audio processing device that includes or is coupled to a HMD of a virtual reality system is shown and generally designated 200.”), Fig. 2.

414. Zhou also teaches that 3D sound can be used to “increase the immersive experience in virtual reality (VR) environments and hence provide a more realistic virtual environment.” Zhou at 990.

415. Thus, claim 10 would have been obvious over Shivappa, alone or in view of Zhou.

3. Independent Claim 15

a. 15[pre]: “An electronic system, comprising”

416. The preamble of claim 15 would have been obvious for the same reasons as the preamble of claim 8. *See* Section XII.C.1.a. Shivappa discloses a system comprising of electronic devices. *See, e.g.*, Shivappa at 5:37-6:3 (“the audio device 102 includes (or is integrated within) a head mounted display (HMD) of a virtual reality (VR) system or an augmented reality (AR) system”).

b. 15[a]: “a display that displays an image that is located within a near-field distance from a head of a user viewing the image”

417. Limitation 15[a] would have been obvious for the same reasons as limitation 8[a]. *See* Section XII.C.1.b. Further, a person of ordinary skill in the art would understand that a user would be viewing the image displayed by display within a near-field distance from a head, i.e. a HMD worn a few inches from the user’s head.

c. 15[b]: “one or more processors that process sound for the image”

418. Limitation 15[b] would have been obvious for the same reasons as limitation 8[b]. *See* Section XII.C.1.c.

d. 15[c]: “with head-related transfer functions (HRTFs) to generate binaural sound”

419. Limitation 15[c] would have been obvious for the same reasons as limitation 8[e]. *See* Section XII.C.1.f. In addition, a person of ordinary skill in the art would have known that the HRTFs are used for generating binaural sound for the same reasons as discussed for limitation 8[c]. *See* Section XII.C.1.d.

e. 15[d]: “that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance”

420. Limitation 15[c] would have been obvious for the same reasons as limitation 8[c]. *See* Section XII.C.1.d.

f. 15[e]: “wherein the coordinate location of the binaural sound continues to occur at the far-field distance located behind the image when the head of the user moves away from viewing the image and the image remains within the near-field distance from the head of the user.”

421. Limitation 15[e] would have been obvious over Shivappa, alone or in view of Zhou, for reasons discussed above. *See* Sections XII.C.1 (claim 8), XII.C.3.a-XII.C.3.d (15[pre]-15[d]).

422. In addition, Shivappa discloses that its “visuals and audio output” “may change based on movement of the user (e.g., movement of the user’s head).” Shivappa at 5:45-48. Shivappa explains that, “if the user’s head moves or rotates, or if the user moves or rotates in the virtual environment, the processor 106 modifies the input spatialized audio file such that a perceived direction or distance of one or more sound sources in a 3D space is modified based on the movement.” *Id.* at 8:38-42; *see also id.* at 6:37-47 (describing ways Shivappa’s system can determine position data). Thus, “if the user is playing a game and a car drives to the left of the user in the game, the audio output enables the user to perceive a sound of a car as coming from the user’s left side. If the user turns to the right in the same such that the car is behind the user, the audio processing device processes the spatialized audio signal to cause the audio output to change such that the user experiences the sound of the car as coming from behind the user.” Shivappa at 3:23-31.

423. In view of these at least these disclosures, persons of ordinary skill in the art would have known and understood that, based on Shivappa alone, Shivappa’s systems and methods could and should maintain sound occurring at the far-field distance for an image displayed on a near-field lens of one of Shivappa’s AR or VR HMDs, e.g., the car driving by in the example above, even while a user’s head moves away from viewing that image so that the user can continue to determine the sound source’s location. That image (e.g., the car, or a character, instrument, or conference

speaker) will remain “within the near-field distance from the head of the user” as a user moves away until the user’s head rotation or causes it to drop from the user’s field of view and the HMD display.

424. For these reasons, each limitation in claim 15 of the ’836 patent, and claim 15 as a whole, would have been obvious over Shivappa, alone or in view of Zhou.

4. Dependent Claims 16 and 17

425. I understand that claims 16 and 17 depend on independent claim 15.

426. Thus, it is my opinion that dependent claims 16 and 17 would have been obvious to persons of ordinary skill in the art for the same reasons as independent claim 15. *See* Section XII.C.3.

427. In addition, claims 16 and 17 would have been obvious for the reasons I discuss in the following subsections.

a. Claim 16: “The electronic system of claim 15, wherein the display is an augmented reality (AR) display worn on the head of the user viewing the image, and the image is an AR image.”

428. Claim 16 would have been obvious for the same reasons as Claim 9. *See* Section XII.C.2.a. Further, a person of ordinary skill in the art would understand that a display that displays an AR image would be an AR display and that a user would

be viewing the AR image when the AR display, i.e. an HMD, is worn on the user's head.

b. Claim 17: “The electronic system of claim 15, wherein the display is a virtual reality (VR) display worn on the head of the user viewing the image, and the image is an VR image.”

429. Claim 17 would have been obvious for the same reasons as Claim 10. *See* Section XII.C.2.b. Further, a person of ordinary skill in the art would understand that a display that displays an VR image would be a VR display and that a user would be viewing the VR image when the VR display, i.e. an HMD, is worn on the user's head.

430. For these reasons, claims 16 and 17 of the '836 patent would have been obvious over Shivappa alone or in view of Zhou.

XIII. Conclusion

431. For these reasons, it is my opinion that claims 8-17 of the '836 patent are unpatentable under 35 U.S.C. § 103.

432. All statements made herein of my own knowledge are true, and all statements made on information and belief are believed to be true. Further, I am aware that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001. I declare under penalty of perjury that the foregoing is true and correct.

Declaration for *Inter Partes* Review of USP 11,290,836

Executed on June 7th, 2023 in New York, NY

A handwritten signature in black ink, appearing to read "A. Roginska", written over a horizontal line.

Dr. Agnieszka Roginska