

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re *Inter Partes* Review of:)
U.S. Patent No. 10,368,179)
Issued: Jul. 30, 2019)
Application No.: 16/401,242)
Filing Date: May 2, 2019)

For: **Switching Binaural Sound**

In re *Inter Partes* Review of:)
U.S. Patent No. 10,448,184)
Issued: Oct. 15, 2019)
Application No.: 16/442,509)
Filing Date: Jun. 16, 2019)

For: **Switching Binaural Sound**

**DECLARATION OF DR. AGNIESZKA ROGINSKA IN SUPPORT OF
PETITIONS FOR *INTER PARTES* REVIEW
OF U.S. PATENT NOS. 10,368,179 AND 10,448,184**

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

Table of Contents

I.	Introduction and Qualifications.....	1
II.	Summary of Materials Reviewed and Considered	4
III.	Legal Understanding.....	9
IV.	The '179 and '184 Patents	13
A.	Summary of the '179 and '184 Patents	13
B.	Priority Date	16
V.	Level of Skill in the Art.....	16
VI.	Claim Construction.....	18
VII.	Overview of the Technology and Prior Art.....	18
A.	Technology Background	18
1.	Head Mounted Displays.....	19
2.	Binaural Sound and HRTFs	27
3.	Processing Binaural Audio	35
4.	Transitioning from VR Environment.....	38
B.	Overview of the Prior Art Relied Upon in the Petitions'	
Grounds		41
1.	Clement (Ex. 1005).....	41
2.	Osman (Ex. 1006)	46
3.	Sauk (Ex. 1007).....	47
4.	Lyren (Ex. 1008)	49
5.	Reif (Ex. 1009).....	54
VIII.	Petition Grounds and Prior Art Combinations	55
IX.	The '179 Patent.....	57
A.	'179 Patent - Ground 1: Claims 1, 3-4, 6-10, and 14-19 Are	
Unpatentable Over Clement, Osman, and Sauk.....		57
1.	Reasons to Implement Clement with Osman.....	57
2.	Reasons to Implement Clement and Osman with Sauk.....	64
3.	Independent Claim 14	68

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

a.	Element 14[pre] A method executed by one or more electronic devices, the method comprising:	68
b.	Element 14[a] storing head-related transfer functions (HRTFs) in memory;	70
c.	Element 14[b] displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;	78
d.	Element 14[c] processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes in empty space from the virtual image in the environment;	83
e.	Element 14[d] determining, with the WED, when the user is leaving the environment;	93
f.	Element 14[e] alerting the user that the user is leaving the environment by reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user	101
4.	'179 Patent - Dependent Claims 15-19	104
a.	Claim 15. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by degrading the binaural sound being provided to the user as the user is leaving the environment.	104
b.	Claim 16. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by altering, by the DSP, processing of the binaural sound so the user perceives less external localization.	106
c.	Claim 17. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.	110
d.	Claim 18. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by eliminating audio cues in the binaural sound being provided to the user.....	110

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

e.	Claim 19. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to stereo sound.....	111
5.	'179 Patent - Independent Claim 7	111
a.	Elements 7[pre] to 7[d].....	111
b.	Element 7[e] alerting the user that the user is leaving the environment by altering processing of the binaural sound to degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound.....	112
6.	'179 Patent - Dependent Claims 8-10.....	112
a.	Claim 8. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by reducing a volume of the binaural sound being provided to the user.....	112
b.	Claim 9. The method of claim 7 further comprising: displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character from the binaural sound to stereo sound when the character vanishes from the environment during the game.	112
c.	Claim 10. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.	117
7.	'179 Patent - Independent Claim 1	117
a.	Elements 1[pre] to 1[d].....	117
b.	Element 1[e] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter audio cues to prevent the user from experiencing external localization of the sounds and by lowering a volume of the sounds being provided to the user.....	118
8.	'179 Patent - Dependent Claims 3-4, and 6.....	118
a.	Claim 3. The method of claim 1, wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment.	118

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

b.	Claim 4. The method of claim 1 further comprising: tracking, with head tracking in the HMD, head movements of the user; and switching the sounds from the binaural sound to stereo sound when a face of the user is directed away from the virtual image.	120
c.	Claim 6. The method of claim 1 further comprising: switching the sounds from the binaural sound to stereo sound when the virtual image vanishes in the 3D virtual environment.	123
B.	'179 Patent - Ground 2: All Claims Are Unpatentable Over Lyren and Reif.....	124
1.	The '179 Patent's claims lack written description support in its parent, the as-filed '140 Application	125
2.	Combination of Lyren and Reif.....	128
3.	'179 Patent - Independent Claim 14	129
a.	Element 14[pre] A method executed by one or more electronic devices, the method comprising:	129
b.	Element 14[a] storing head-related transfer functions (HRTFs) in memory;	130
c.	Element 14[b] displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;	130
d.	Element 14[c] processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes in empty space from the virtual image in the environment;	133
e.	Elements 14[d]-14[e] determining, with the WED, when the user is leaving the environment; and alerting the user that the user is leaving the environment by reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user.....	137
4.	'179 Patent - Dependent Claims 15-20.....	154
a.	Claim 15. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by degrading the binaural sound being provided to the user as the user is leaving the environment.	154

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

b.	Claim 16. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by altering, by the DSP, processing of the binaural sound so the user perceives less external localization.	156
c.	Claim 17. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.	156
d.	Claim 18. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by eliminating audio cues in the binaural sound being provided to the user.....	157
e.	Claim 19. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to stereo sound.....	158
f.	Claim 20. The method of claim 14 further comprising: detecting, with a microphone in the WED, a voice of a person in the environment of the user; and switching the binaural sound to stereo sound in response to detecting the voice.	158
5.	'179 Patent - Independent Claim 7	161
a.	Elements 7[pre] to 7[c]	161
b.	Element 7[d] determining, with the WED, when the user is leaving the environment; and	161
c.	Element 7[e] alerting the user that the user is leaving the environment by altering processing of the binaural sound to degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound.....	161
6.	'179 Patent - Dependent claims 8-13.....	162
a.	Claim 8. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by reducing a volume of the binaural sound being provided to the user.....	162

Declaration for *Inter Partes Review* of USP 10,368,179 and 10,448,184

b.	Claim 9. The method of claim 7 further comprising: displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character from the binaural sound to stereo sound when the character vanishes from the environment during the game.	162
c.	Claim 10. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.	163
d.	Claim 11. The method of claim 7 further comprising: switching the binaural sound to stereo sound upon activation of a switch on the WED.	164
e.	Claim 12. The method of claim 7 further comprising: wherein the WED includes electronic glasses, the environment is a physical room where the user is located, the virtual image is an augmented reality (AR) character that speaks to the user while the user plays a game, the WED includes head tracking, and a digital signal processor (DSP) in the WED reduces a volume of the binaural sound to alert the user when a face of the user is not directed to the virtual image in the environment.	164
f.	Claim 13. The method of claim 7 further comprising: sensing, with the WED, when a physical object interferes with a location of the virtual image in the environment; and switching the binaural sound to stereo sound in response to sensing the physical object interferes with the location of the virtual image in the environment.	166
7.	'179 Patent - Independent Claim 1	168
a.	Elements 1[pre] to 1[c]	168
b.	Element 1[d] sensing, with one or more sensors in the HMD, when the user leaves the 3D virtual environment that includes the virtual image; and	168

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

c.	Element 1[e] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter audio cues to prevent the user from experiencing external localization of the sounds and by lowering a volume of the sounds being provided to the user.	169
8.	'179 Patent - Dependent Claims 2-6.....	169
a.	Claim 2. The method of claim 1 further comprising: warning the user that the user is leaving the 3D virtual environment by removing, by the DSP, background sounds that occur in the 3D virtual environment as part of a game the user is playing.....	169
b.	Claim 3. The method of claim 1, wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment.	170
c.	Claim 4. The method of claim 1 further comprising: tracking, with head tracking in the HMD, head movements of the user; and switching the sounds from the binaural sound to stereo sound when a face of the user is directed away from the virtual image.	172
d.	Claim 5. The method of claim 1 further comprising: switching the sounds from the binaural sound to stereo sound when the user activates a physical switch located in the HMD.....	173
e.	Claim 6. The method of claim 1 further comprising: switching the sounds from the binaural sound to stereo sound when the virtual image vanishes in the 3D virtual environment.	173
X.	The '184 Patent.....	174
A.	'184 Patent - Ground 1: Claims 1, 3-4, 6-10, and 14-19 Are Unpatentable Over Clement, Osman, and Sauk.....	174
1.	Reasons to Implement Clement with Osman.....	174
2.	Reasons to Implement Clement and Osman with Sauk.....	175
3.	'184 Patent - Independent Claim 14	175
a.	Elements 14[pre] to 14[c]	175

Declaration for *Inter Partes Review* of USP 10,368,179 and 10,448,184

b.	Element 14[d]: alerting the user that the user is leaving the virtual image in the environment by altering processing of the binaural sound by the DSP and reducing a volume of the binaural sound that localizes to the user at the virtual image.	176
4.	'184 Patent - Dependent Claims 15-19.....	176
a.	Claim 15. The method of claim 14, wherein the DSP alters processing of the binaural sound to alert the user by changing a spatial cue that includes at least one of an interaural time difference (ITD), an interaural level difference (ILD), and the HRTFs.	177
b.	Claim 16. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by stopping, by the DSP, the processing of the sound with the HRTFs.	178
c.	Claim 17. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.	179
d.	Claim 18. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by eliminating one or more spatial cues in the binaural sound being provided to the user.	179
e.	Claim 19. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to one of stereo sound and mono sound.....	180
5.	'184 Patent - Independent Claim 7	181
a.	Elements 7[pre] to 7[c]	181
b.	Element 7[d] alerting the user that the user is leaving the environment by altering processing of the sound by the DSP to degrade a spatial cue and to lower a volume of the sound playing to the user.....	182
6.	'184 Patent - Dependent Claims 8-10.....	183
a.	Claim 8. The method of claim 7, wherein the spatial cue includes an interaural time difference (ITD).	183

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

b.	Claim 9. The method of claim 7 further comprising: displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character from the binaural sound to stereo sound when the WED no longer displays the virtual image during the game.....	183
c.	Claim 10. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.	184
7.	'184 Patent - Independent Claim 1	184
a.	Elements 1[pre] to 1[c]	184
b.	Element 1[d] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter a spectral cue and lower a volume of the sound playing to the user.....	186
8.	'184 Patent - Dependent Claims 3-4, and 6	186
a.	Claim 3. The method of claim 1, wherein the warning occurs when the user is leaving a virtual area where the virtual image is located while remaining in the 3D virtual environment.	186
b.	Claim 4. The method of claim 1 further comprising: tracking, with head tracking in the HMD, head movements of the user; and switching the sound from the binaural sound to stereo sound when a face of the user is directed away from the virtual image such that the virtual image is no longer visible to the user.....	187
c.	Claim 6. The method of claim 1 further comprising: switching the sound from the binaural sound to stereo sound when the virtual image vanishes from the 3D virtual environment.	188
B.	'184 Patent - Ground 2: All Claims Are Unpatentable Over Lyren and Reif.....	188
1.	The '184 Patent's claims are not supported by its parent, the '140 Application	188
2.	Combination of Lyren and Reif	189
3.	'184 Patent - Independent Claim 14	190

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

a.	14[pre] to 14[c]	190
b.	Element 14[d]: alerting the user that the user is leaving the virtual image in the environment by altering processing of the binaural sound by the DSP and reducing a volume of the binaural sound that localizes to the user at the virtual image.	190
4.	'184 Patent - Dependent Claims 15-20	191
a.	Claim 15. The method of claim 14, wherein the DSP alters processing of the binaural sound to alert the user by changing a spatial cue that includes at least one of an interaural time difference (ITD), an interaural level difference (ILD), and the HRTFs.	191
b.	Claim 16. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by stopping, by the DSP, the processing of the sound with the HRTFs.	193
c.	Claim 17. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.	194
d.	Claim 18. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by eliminating one or more spatial cues in the binaural sound being provided to the user.	194
e.	Claim 19. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to one of stereo sound and mono sound.	195
f.	Claim 20. The method of claim 14 further comprising: detecting, with a microphone in the WED, a voice of a person in the environment of the user; and switching the binaural sound to stereo sound in response to detecting the voice.	196
5.	'184 Patent - Independent Claim 7	196
a.	Elements 7[pre] to 7[c]	196

Declaration for *Inter Partes Review* of USP 10,368,179 and 10,448,184

b.	Element 7[d] alerting the user that the user is leaving the environment by altering processing of the sound by the DSP to degrade a spatial cue and to lower a volume of the sound playing to the user.....	197
6.	'184 Patent - Dependent claims 8-13.....	199
a.	Claim 8. The method of claim 7, wherein the spatial cue includes an interaural time difference (ITD).	199
b.	Claim 9. The method of claim 7 further comprising: displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character from the binaural sound to stereo sound when the WED no longer displays the virtual image during the game.	199
c.	Claim 10. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.	200
d.	Claim 11. The method of claim 7 further comprising: switching the binaural sound to stereo sound upon activation of a button on the WED.	200
e.	Claim 12. The method of claim 7 further comprising: wherein the WED includes electronic glasses, the environment is a physical room where the user is located, the virtual image is an augmented reality (AR) character that speaks to the user while the user plays a game, the WED includes head tracking, and a digital signal processor (DSP) in the WED reduces a volume of the binaural sound to alert the user when a face of the user is not directed to the virtual image in the environment.	201
f.	Claim 13. The method of claim 7 further comprising: sensing, with the WED, when a physical object interferes with a location of the virtual image in the environment; and switching the binaural sound to stereo sound in response to sensing the physical object interferes with the location of the virtual image in the environment.	201

Declaration for *Inter Partes Review* of USP 10,368,179 and 10,448,184

7.	'184 Patent - Independent Claim 1	202
a.	Elements 1[pre] to 1[c]	202
b.	Element 1[d] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter a spectral cue and lower a volume of the sound playing to the user.....	203
8.	'184 Patent - Dependent Claims 2-6.....	204
a.	Claim 2. The method of claim 1 further comprising: warning the user that the user is leaving the 3D virtual environment by removing, by the DSP, background sound that occurs in the 3D virtual environment as part of a game the user is playing.....	204
b.	Claim 3. The method of claim 1, wherein the warning occurs when the user is leaving a virtual area where the virtual image is located while remaining in the 3D virtual environment.	204
c.	Claim 4. The method of claim 1 further comprising: tracking, with head tracking in the HMD, head movements of the user; and switching the sound from the binaural sound to stereo sound when a face of the user is directed away from the virtual image such that the virtual image is no longer visible to the user.....	205
d.	Claim 5. The method of claim 1 further comprising: switching the sound from the binaural sound to stereo sound when the user activates a button located on the HMD.	206
e.	Claim 6. The method of claim 1 further comprising: switching the sound from the binaural sound to stereo sound when the virtual image vanishes from the 3D virtual environment.	206
XI.	Conclusion	207

I. Introduction and Qualifications

1. I, Agnieszka Roginska, have been retained by Meta Platforms, Inc., Meta Platforms Technologies, LLC, and Twisted Pixel Games, LLC (“Petitioners”) to provide my opinion concerning the validity of Claims 1-20 of U.S. Patent No. 10,368,179 (“the ’179 Patent”) in support of the Petition for *Inter Partes* Review of U.S. Patent No. 10,368,179, and the validity of Claims 1-20 of U.S. Patent No. 10,448,184 (“the ’184 Patent”) in support of the Petition for *Inter Partes* Review of U.S. Patent No. 10,448,184.

2. I am more than eighteen years of age, and I am a citizen of the United States, currently residing in the state of New York.

3. 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 to 2018. I served as Vice-Chair of the Music and Performing Arts Professions Department 2019 to 2020. From 2007 until 2022, I served as the Director of the Music Technology summer programs.

4. 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.

Declaration for *Inter Partes Review* of USP 10,368,179 and 10,448,184

5. 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".

6. 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. 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.

7. 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

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

experiences. Applications of my work include Augmented and Virtual Reality, gaming, sound recording and production, mission critical, and augmented acoustic sensing.

8. 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.

9. I have received over \$11M in funding from the NSF, NIH, Office of Naval Research, US Army, US Navy and NASA.

10. 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 the AES in various roles since 2009. I was President of the AES in 2020 and am currently on the Board of Directors of the AES.

11. 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.

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

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

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

13. 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
1001-179	U.S. Patent No. 10,368,179 (“the ’179 Patent”)
1001-184	U.S. Patent No. 10,448,184 (“the ’184 Patent”)
1004	<i>Curriculum Vitae</i> of Dr. Agnieszka Roginska
1005	U.S. Patent No. 9,690,374 (“Clement”)
1006	U.S. Patent No. 9,037,468 (“Osman”)
1007	U.S. Patent No. 7,876,903 (“Sauk”)
1008	U.S. Patent Application Publication No. 2017/0188168 (“Lyren”)
1009	U.S. Patent Application Publication No. 2018/0373412 (“Reif”)
1010	U.S. Patent Application Publication No. 2011/0206209 (“Ojala”)
1011	U.S. Patent No. 1,624,486 (“Fletcher”)
1012	U.S. Patent No. 1,855,149 (“Jones”)
1013	U.S. Patent No. 7,330,197 (“Kobayashi”)
1014	U.S. Patent No. 7,710,654 (“Ashkenazi”)
1015	U.S. Patent No. 8,520,872 (“Jang”)
1016	U.S. Patent No. 8,520,873 (“Mahabub”)
1017	U.S. Patent No. 9,041,622 (“McCulloch”)
1018	U.S. Patent No. 9,367,811 (“Ramachandran”)
1019	U.S. Patent No. 9,544,706 (“Hirst”)
1020	U.S. Patent No. 9,578,440 (“Otto”)
1021	U.S. Patent No. 9,674,507 (“Pirchheim”)
1022	RESERVED

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

1023	U.S. Patent No. 9,767,618 (“Najaf-Zadeh”)
1024	RESERVED
1025	U.S. Patent No. 10,070,239 (“Romigh”)
1026	RESERVED
1027	U.S. Patent No. 10,095,031 (“Katz”)
1028	File History for U.S. Patent Application 16/106,140
1029	U.S. Patent No. 10,509,466 (“Miller”)
1030	U.S. Patent Application Publication No. 2012/0183161 (“Agevik”)
1031	U.S. Patent Application Publication No. 2016/0124502 (“Sawyer”)
1032	U.S. Patent Application Publication No. 2016/0292924 (“Balachandreswaran”)
1033	U.S. Patent Application Publication No. 2017/0105083 (“Nair”)
1034	U.S. Patent Application Publication No. 2017/0115488 (“Ambrus”)
1035	Korean Patent Application Publication No. KR 10-2011-0022252, including English Translation and Certification (“Moon”)
1036	Christoph Anthes et al., <i>State of the Art of Virtual Reality Technology</i> , 2016 IEEE AEROSPACE CONF. 1-19 (Mar. 2016) (“Anthes”)
1037	<i>Average Cost of RAM per Gigabyte Chart</i> , HumanProgress (“HumanProgress”)
1038	DURAND R. BEGAULT, 3-D SOUND FOR VIRTUAL REALITY AND MULTIMEDIA (NASA/TM-2000-209606) (Aug. 2000) (“Begault”)
1039	Tom Emrich, <i>Virtual Reality North: A Look at the Major VR Players in Canada</i> , betakit.com (July 15, 2015), betakit.com/virtual-reality- north-a-look-at-the-major-vr-players-in-canada/ (“Emrich”)
1040	Braxton Boren, <i>History of 3D Sound</i> (Am. Univ., Wash. D.C. 2018), https://dra.american.edu/islandora/object/auislandora%3A77787/data stream/PDF/view (“Boren”)
1041	Douglas S. Brungart et al., <i>The Interaction Between Head-Tracker Latency, Source Duration, and Response Time in the Localization of Virtual Sound Sources</i> , PROC. ICAD 04- 10TH MEETING INT’L CONF. AUDITORY DISPLAY 1-7 (July 6-9, 2004) (“Brungart 2004”)
1042	Douglas S. Brungart et al, <i>The Detectability of Headtracker Latency in Virtual Audio Displays</i> , PROC. ICAD 05-11TH MEETING INT’L

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

	CONF. AUDITORY DISPLAY 37-42 (July 6-9, 2005) (“Brungart 2005”)
1043	E. Colin Cherry, <i>Some Experiments on the Recognition of Speech, with One and with Two Ears</i> , 25 J. ACOUSTICAL SOC’Y AM. 975 (1953) (“Cherry”)
1044	Christof Faller, <i>Coding of Spatial Audio Compatible with Different Playback Formats</i> , AES 117th CONVENTION 1-12 (Oct. 28-31, 2004) (“Faller”)
1045	Mohammad Dashti et al., <i>Traffic Management: A Holistic Approach to Memory Placement on NUMA Systems</i> , ASPLOS’ 1-13 (Mar. 16-20, 2013) (“Dashti”)
1046	S.S. Fisher et al., <i>Virtual Interface Environment Workstations</i> , 32 PROC. HUM. FACTORS SOC’Y 32ND ANN. MEETING 1988 91-95 (Feb. 1988) (“Fisher”)
1047	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) (“Fohl”)
1048	MARK GRIMSHAW, <i>GAME SOUND TECHNOLOGY AND PLAYER INTERACTION: CONCEPTS AND DEVELOPMENTS</i> 22 (2011) (“Grimshaw”)
1049	Eric Grownney & Agnieszka Roginska, <i>Effect of Spatial Audio Processing on Attention Redirection and Speech Intelligibility in Multi-Talker Communication</i> , AES 152ND CONVENTION 1-10 (May 2022) (“Grownney”)
1050	Maxim Liadov, <i>A New Generation of Xtreme Fidelity Processors and Sound Cards</i> , iXBT Labs (Aug. 22, 2005) (“Liadov”)
1051	JESSE SCHELL, <i>THE ART OF GAME DESIGN: A BOOK OF LENSES</i> (2008) (“Schell”)
1052	Christoph Lameter, <i>Local and Remote Memory: Memory in a Linux/NUMA System</i> , SGI (June 20, 2006) (“Lameter 2006”)
1053	Christoph Lameter, <i>NUMA (Non-Uniform Memory Access): An Overview</i> , ACMqueue 1-12 (2013) (“Lameter 2013”)
1054	Matteo Marucci et al., <i>The Impact of Multisensory Integration and Perceptual Load in Virtual Reality Settings on Performance</i> ,

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

	<i>Workload and Presence</i> , 11 NATURE SCI. REPS. 1 (2021) (“Marucci”)
1055	Joel D. Miller et al., <i>Latency Measurements of a Real-Time Virtual Acoustic Environment Rendering System</i> , PROC. 2003 INT’L CONF. AUDITORY DISPLAY 111-114 (July 6-9, 2003) (“Miller”)
1056	Francisco Cheng, <i>Mobile Processors 101: Why Smartphones are Smarter with an All-In-One Processor</i> , QUALCOMM ONQ BLOG (June 12, 2013) (“Cheng”)
1057	Masayuki Morimoto & Yoichi Ando, <i>On the Simulation of Sound Localization</i> , (E) 1 J. ACOUSTIC SOC’Y JAPAN 167 (1980) (“Morimoto”)
1058	RESERVED
1059	Stephen Paul, <i>Binaural Recording Technology: A Historical Review and Possible Future Developments</i> , 95 ACTA ACUSTICA UNITED WITH ACUSTICA 767 (2009) (“Paul”)
1060	Ben Kuchera, <i>Google Cardboard Adds ‘Inexpensive’ Spatial Audio to its SDK</i> , POLYGON (Jan. 13 2016, 12:00pm PST) (“Kuchera”)
1061	Paul Sauk & Duane Satorius, <i>Creating a Multi-Dimensional Communication Space to Improve the Effectiveness of 3-D Audio</i> , IEEE MILITARY COMM’CNS CONF. 1-7 (Oct. 29-31, 2007) (“Sauk 2007”)
1062	Mel Slater & Martin Usoh, <i>Presence in Immersive Virtual Environments</i> , IEEE VIRTUAL REALITY ANN. INT’L SYMP. 90-96 (Sept. 18-22, 1993) (“Slater”)
1063	JAKA SODNIK & SAŠO TOMAŽIČ, <i>SPATIAL AUDITORY HUMAN-COMPUTER INTERFACES</i> , SPRINGER BRIEFS COMPUT. SCI. (2015) (“Sodnik”)
1064	Paul James, <i>Sulon Cortex ‘Augmented Virtual Reality’ Headset, Interview with Don Henderson</i> , ROAD TO VR (Feb. 5, 2015) (“James”)
1065	Mona Lalwani, <i>Surrounded by Sound: How 3D Audio Hacks Your Brain</i> , THE VERGE (Feb. 12, 2015, 10:30am) https://www.theverge.com/2015/2/12/8021733/3d-audio-3dio-binaural-immersive-vr-sound-times-square-new-york (“Lalwani”)

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

1066	Dave Smith, <i>Three-Dimensional Audio Makes Virtual Reality so Much Better It's Crazy</i> , BUSINESS INSIDER (Apr. 3, 2015, 1:06pm), https://www.businessinsider.com/virtual-reality-the-importance-of-3d-positional-audio-2015-4 (“Smith”)
1067	JASON JERALD, THE VR BOOK: PERCEPTION AND INTERACTION DESIGN FOR VIRTUAL REALITY (2015) (“VR Book”)
1068	Elizabeth M. Wenzel et al., <i>Development of a Three-Dimensional Auditory Display System</i> , 20 SIGCHI BULL. 52 (Oct. 1988) (“Wenzel 1988a”)
1069	Elizabeth M. Wenzel et al., <i>A Virtual Display System for Conveying Three-Dimensional Acoustic Information</i> , PROC. HUM. FACTORS SOC’Y 32ND ANN. MEETING 86-90 (1988) (“Wenzel 1988b”)
1070	RESERVED
1071	Elizabeth M. Wenzel et al., <i>Localization with Non-Individualized Virtual Acoustic Display Cues</i> , PROC. SIGCHI CONF. HUM. FACTORS COMPUT. SYS. 351-359 (Apr. 1991) (“Wenzel 1991”)
1072	Elizabeth M. Wenzel, <i>Localization in Virtual Acoustic Displays</i> , 1 PRESENCE TELEOPERATORS VIRTUAL ENV’TS 80 (Feb. 1992) (“Wenzel 1992”)
1073	Qualcomm, <i>What’s So Special About the Digital Signal Processor?</i> , QUALCOMM ONQ BLOG (Dec. 5, 2013) (“OnQ”)
1074	Frederic L. Wightman & Doris J. Kistler, <i>Headphone Simulation of Free-Field Listening. II: Psychophysical Validation</i> , 85 J. ACOUSTIC SOC’Y AM. 868 (1989) (“Wightman”)
1075	Jiann-Rong Wu et al., <i>Head Motion and Latency Compensation on Localization of 3D Sound in Virtual Reality</i> , PROC. ACM SYMP. VIRTUAL REALITY SOFTWARE TECH. 15-20 (Sept. 1997) (“Wu”)
1076	Dmitry N. Zotkin et al., <i>Customizable Auditory Displays</i> , PROC. 2002 INT’L CONF. AUDITORY DISPLAY 1-10 (July 2-5, 2002) (“Zotkin”)
1077	Gr. Brit. Provisional Patent Specification No. 394,325 (“Blumlein”)
1078	Linley Gwennap, <i>Qualcomm Extends Hexagon DSP</i> , MICROPROCESSOR REP. 1-4 (Aug. 26, 2012) (“Gwennap”)
1079	Stanley P. Lipshitz, <i>Stereo Microphone Techniques.....Are the Purists Wrong?</i> 34 J. AUDIO ENG’G SOC’Y 716 (1986) (“Lipshitz”)

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

1080	U.S. Patent No. 9,851,786 (“Pedrotti”)
1081	U.S. Patent No. 9,787,846 (“Srinivasan”)
1082	U.S. Patent Application Publication No. 2015/0094142 (“Stafford”)
1083	U.S. Patent Application Publication No. 2019/0043259 (“Wang”)
1085	Oculus Connect 5 Spatial Audio for Oculus Quest and Beyond video clips (May 12, 2023)
1086	Oculus Connect 5 Spatial Audio for Oculus Quest and Beyond video clip transcriptions (May 12, 2023)
1087	Oculus Connect 5 Keynote Day 01 video clips (May 16, 2023)
1088	Oculus Connect 5 Keynote Day 01 video clip transcriptions (May 16, 2023)
1093	U.S. Patent No. 5,173,944 (“Begault ‘944”)
1094	U.S. Patent No. 5,438,623 (“Begault ‘623”)

III. Legal Understanding

14. I am not an attorney. However, 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 (“POSA”) at the time of the priority date 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

Declaration for *Inter Partes Review* of USP 10,368,179 and 10,448,184

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 prior art reference can render a patent claim obvious to a POSA 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. In analyzing obviousness, I understand that it is important to consider the scope of the claims, the level of skill of a POSA in the relevant art, the scope and content of the prior art, the differences between the prior art and the claims, and so-called 'secondary considerations' of non-obviousness.

18. I understand that the claimed subject matter is obvious when a POSA would have been motivated to combine pre-existing elements and that combination yields no more than what a POSA would have been expected. 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 POSA would employ.

19. 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 POSA can fit the teachings of multiple pieces of prior art together.

20. I understand that certain issues relating to validity must be judged from the perspective of a POSA. I further understand that factors that may be considered in determining the level of skill of a POSA 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.

21. I have been informed and understand that a POSA 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.

22. I understand that, when there is a reason to solve a problem and there are a finite number of identified, predictable solutions, a POSA has good reason to pursue the known options within his or her technical grasp.

23. 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.

24. I understand that, to have sufficient written description support, the disclosure (including the as-filed claims) must reasonably convey to a POSA that the inventor was in possession of the claimed subject matter as of the filing date.

IV. The '179 and '184 Patents

A. Summary of the '179 and '184 Patents

25. The '179 and '184 Patents share a common specification. The '179 and '184 Patents are entitled “Switching Binaural Sound.” *See, e.g.*, '179 Patent, Title.¹

26. The '179 and '184 Patents do not purport to have invented an electronic device or Head Mounted Display (HMD) capable of using head related transfer functions (HRTFs) to generate localized audio, a method for filtering or processing binaural audio, nor the concept of alerting a user when he/she leaves a virtual environment, *see, e.g.*, '179 Patent, 2:38-4:2 (describing that sound localization was known in the prior art), and as described further in the Technology Background, these concepts were well-known by 2015.

27. Instead, the claimed inventions of the '179 and '184 Patents relates to “provid[ing] binaural sound to a person through electronic earphones ... which localizes to a sound localization point (SLP) in empty space that is away from but proximate to the person” such that when a triggering “event” occurs “the binaural sound switches or changes to stereo sound, to mono sound, or to altered binaural sound.” *See, e.g.*, '179 Patent, Abstract.

¹ Exemplary cites are made to just the '179 Patent in this section. The disclosures of the '184 Patent are identical.

28. Specifically, the '179 and '184 Patents state that under various conditions (e.g., when an “event” occurs), it may be desirable to alter the processing of binaural sound, including switching from binaural to stereo or mono sound, reducing the volume of the sound, or altering the binaural sound. *See, e.g.,* '179 Patent, 4:3-18 (“In an example embodiment, switching from binaural sound or altering binaural sound and/or a SLP occurs so that the user is unable to perceive externalization of one or more SLPs or audio cues. This prevents, inhibits, reduces, or encumbers the user from externally localizing sound or a portion thereof.”), 14:8-26, (“In FIGS. 4-16, example actions include, but are not limited to, one or more of switch the sound from binaural sound to stereo sound, switch the sound from binaural sound to mono sound, switch the sound from stereo sound to binaural sound, switch the sound from mono sound to binaural sound, maintain binaural sound but alter the binaural sound, stop binaural sound, discontinue playing sound, mute the sound, lower a volume of the sound ...”), 27:62-63 (“FIG. 16 is a method to provide binaural sound to a person and to take an action when an event occurs.”).

29. The '179 and '184 Patents include laundry lists of possible triggering “events.” One such condition is when the user moves into a “restricted area” where providing binaural sound is dangerous or prohibited. *See, e.g.,* '179 Patent, 16:26-48 (“Examples of restricted areas include, but are not limited to, an area, a location, or a point that prohibits SLPs or sound localization, an area in which it is dangerous

to localize sound, a vehicle, or other location.”). For example, when a participant in a binaural telephone conversation enters a vehicle and begins driving, the “earphones immediately stop localizing the voice of [the first participant]” and switch from binaural sound to mono sound. *See, e.g.*, ’179 Patent, 16:43-48.

30. Other events include the “position or movement or orientation of a person or head of the person (such as direction the person walks or head orientation of the person) [or] distance from the person to an object (such as distance from the person to a wall or an obstruction),” *See, e.g.*, ’179 Patent, 9:5-42 (“Examples of events and things that can trigger events include ... position or movement or orientation of a person or head of the person (such as direction the person walks or head orientation of the person), distance from the person to an object (such as distance from the person to a wall or an obstruction), available space (such as how much physical 3D space is available to receive and/or localize a sound or voice), safety (such as not localizing sound when the person is driving a vehicle), proximity to or being at a restricted area (such as state, local, or United States Federal regulations prohibiting externally localizing sound while in a certain building or on an airplane)”), 28:9-29:3 (“Events are also triggered according to one or more audio cues detected in a user's or a remote user's physical or virtual environment such as cues indicating a location, position, or orientation, or a change in them, cues indicating a reference frame of a user or a remote user, a lateral or vertical motion,

a change in distance, a change in a physical or a virtual environment such as its shape, acoustic properties, noise level, or placement of objects or structures in the environment.”). However, none of these is the claimed “warning the user that the user is leaving the 3D virtual environment.”

B. Priority Date

31. My opinions are additionally guided by my appreciation of how a POSA would have understood the claims of the ’179 Patent or the ’184 Patent at the time of their alleged invention,

32. For parts of my analysis, as described below in Ground 1 (Clement-based ground) for each patent, I have been asked to assume the time of the alleged invention (sometimes called the priority date) for the ’179 and ’184 Patents is the earliest priority date on the face of these patents: December 27, 2015.

33. For other parts of my analysis, as described below in the Ground 2 (Lyren-based ground) for each patent, I have been asked to assume the priority date for ’179 Patent is its actual filing date: May 2, 2019, and the priority date for ’184 Patent is its filing date: June 16, 2019.

V. Level of Skill in the Art

34. A POSA at the time of the priority date for the ’179 and ’184 Patents, either December 27, 2015 or their filing dates: May 2, 2019 and June 16, 2019, would have had a bachelor’s degree in computer science/engineering, music

technology, and at least two years of industry experience or academic research experience in virtual/augmented reality, audio engineering, or an equivalent. Additional education can compensate for less experience, and vice-versa.

35. At the time of the earliest claimed priority date, and at the time of the alleged invention, I met the qualifications to be a POSA. Although my experience and education at that time exceeded that of a POSA, I would have been personally familiar with the knowledge and capabilities of a POSA based on my experiences as a professor in the Music Technology program, and a former audio software specialist, and audio research scientist. In particular, I would have been exposed to the knowledge and capabilities of a POSA based on my responsibilities as a professor, which included educating, overseeing program requirements, and advising on the research of undergraduate and graduate students, and based on my long-standing experience and involvement with professional organizations and conferences, through which I am exposed to the experience and expertise of those working in the industry.

36. I have applied the perspective of a POSA in forming and expressing my opinions. Unless otherwise stated, my testimony below refers to the knowledge of one of ordinary skill in the art at the relevant time.

VI. Claim Construction

37. In my analysis, I have applied the plain and ordinary meanings of the claim terms in the '179 and '184 Patents, as a POSA would have understood them in the context of that patent at the time of their priority dates.

38. In particular, I have been asked to assume that the term “processing, by a digital signal processor (DSP) ...” has its plain and ordinary meaning.

39. However, 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 '179 and '184 Patents.

40. The opinions that I provide in this declaration would remain true under any reasonable construction of the claim terms in the '179 and '184 Patents.

VII. Overview of the Technology and Prior Art

A. Technology Background

41. In this section, I provide overviews of technologies and prior art pertinent to the '179 and '184 Patents and to the state of the art as of the '179 and '184 Patents earliest possible priority date. 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.

42. 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.

1. Head Mounted Displays

43. HMDs for display of virtual reality (VR) environments were well-known before December 2015, the earliest possible priority date for the alleged inventions of the '179 and '184 Patents.

44. 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 (Ex. 1046), 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.” Fisher (Ex. 1046), 92-93.

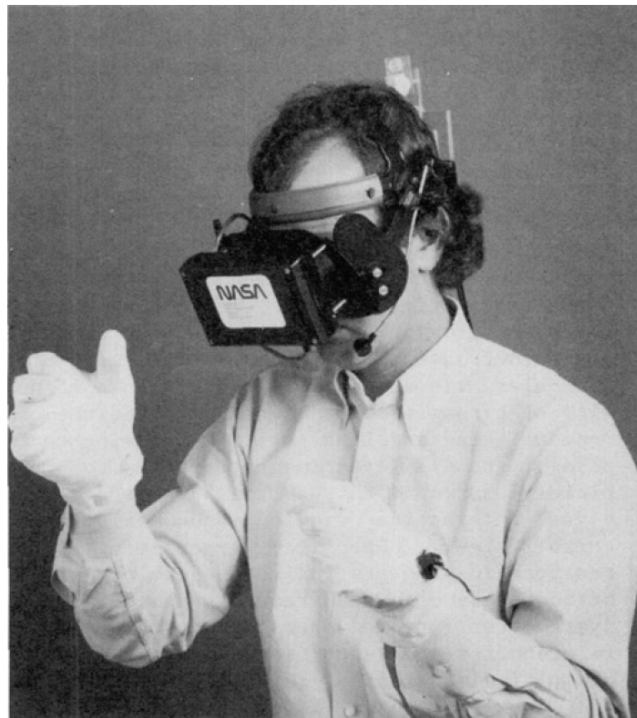


Figure 1. Head-coupled, stereoscopic display system with DataGlove tactile input technology for virtual object manipulation.

Fisher (Ex. 1046), 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).” Wenzel 1988b (Ex. 1069), 87. The following figure, from VR Book (Ex. 1067) depicts some of the many different types of HMDs that were known to persons of ordinary skill in the art before December 2015:

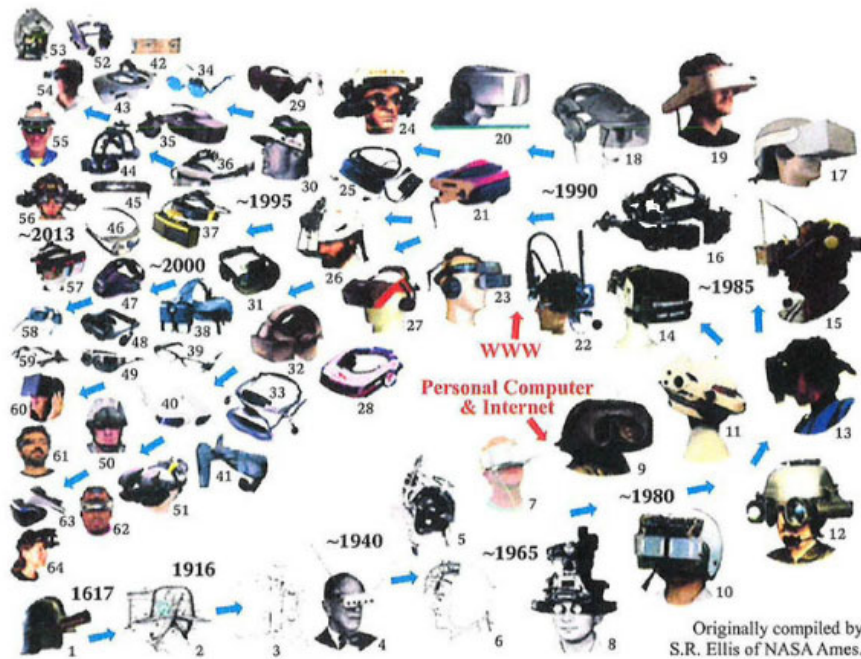


Figure 2.1 Some head-mounted displays and viewers over time. (Originally compiled from Ellis (2014))

VR Book (Ex. 1067), Fig. 2.1.

45. In fact, “head tracking” is a “core element[] for VR experiences.” VR Book (Ex. 1067), 263. Persons of ordinary skill in the art knew before December 2015 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* VR Book, (Ex. 1067) 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”). As Begault (Ex. 1038) 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 (Ex. 1038) at 133. Figure 4.37 in Begault (Ex. 1038) 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”:

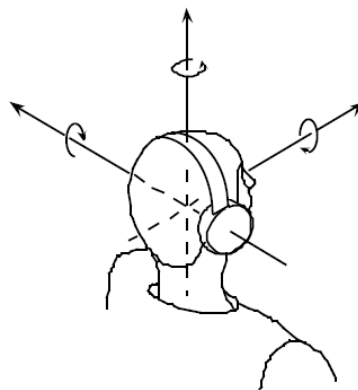


Figure 4.37. Coordinates for pitch-yaw-roll and x , y , z in a head-tracked system. *Courtesy of Crystal River Engineering.*

Begault (Ex. 1038) at 133, Fig. 4.37. Figure 4.38 in Begault (Ex. 1038) “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":

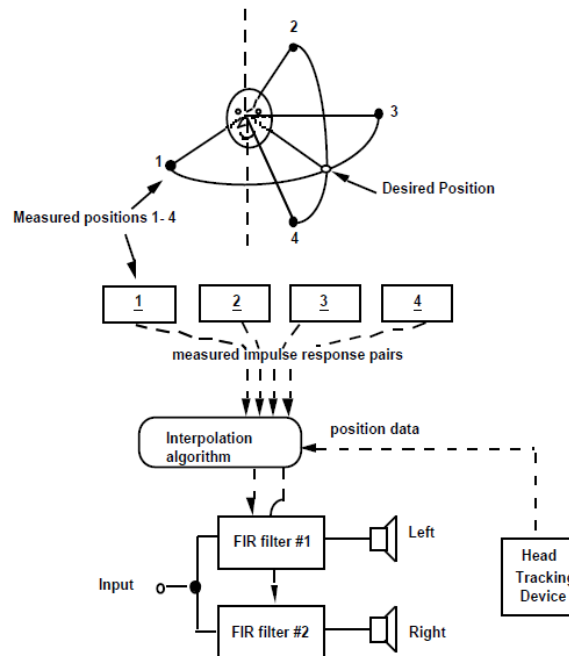


Figure 4.38. Interpolation procedure used to compensate for head movement of the listener. To simulate moving sound sources, the position data is obtained algorithmically instead of from a head tracker.

Begault (Ex. 1038), 133, Fig. 4.38. As Begault (Ex. 1038) 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.” Begault (Ex. 1038), 133.

46. 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.

47. For example, U.S. Patent Publication No. 2017/0115488 (Ambrus, Ex. 1034), 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 (Ex. 1034), [0074]-[0078], [0080]-[0082] (tracking head movements to determine head orientation and gaze direction).

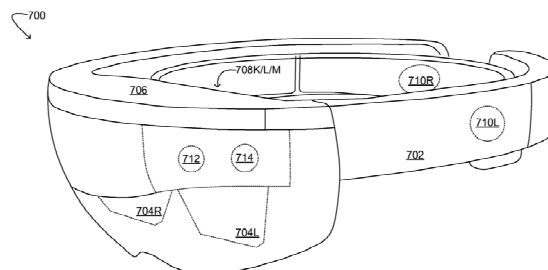
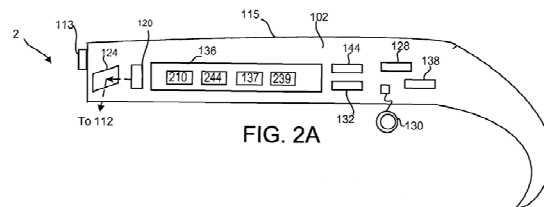


FIG. 7

Ambrus (Ex. 1034), Fig. 7; *see also* Ambrus (Ex. 1034) 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.”). Likewise, U.S. Patent No. 9,041,622 (McCulloch, Ex. 1017), also assigned to Microsoft, describes “inertial sensors . . . for sensing position, orientation, and

sudden accelerations of head mounted display device 2.” McCulloch (Ex. 1017), 8:17-39.



McCulloch (Ex. 1017), Fig. 2A.

48. As another example, U.S. Patent Pub. No. 2016/0292924 (Balachandreswaran, Ex. 1032) describes that an “HMD 12 may comprise: a display system 121 having a display 122, such as, for example, a flat panel display; a camera system 123 which may include one or more cameras; an audio system 124 with audio input and output to provide the user with audio interaction; [and] an inertial measurement unit 127, such as, for example, a 3-axis accelerometer, 3-axis compass or 3-axis gyroscope; and/or a 2D/3D wireless local position system 128 provided by ultrasonic, RF, other wireless or magnetic tracking technologies or other suitable local positioning technologies.” Balachandreswaran (Ex. 1032) at [0063],

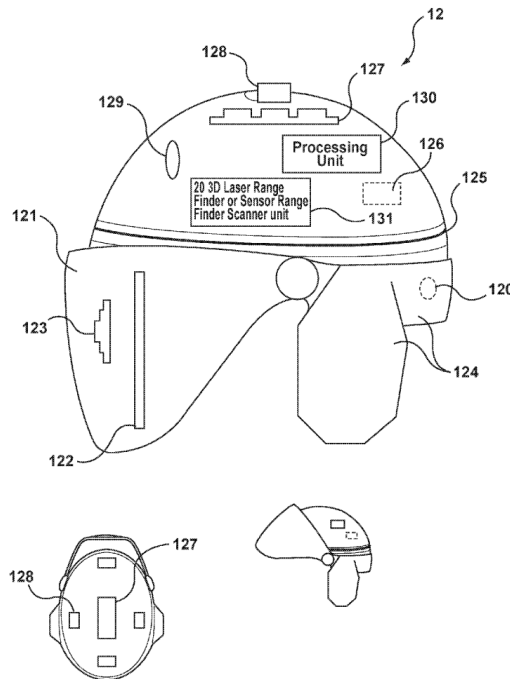


FIG. 5

Balachandreswaran (Ex. 1032), Fig. 5; *see also* U.S. Patent No. 9,367,811 (Ramachandran, Ex. 1018), 2:29-3:2 (describing devices, including head mounted displays, with positioning sensors), 12:10-23 (describing visual mapping, localization, and tracking); U.S. Patent No. 9,674,507 (Pirchhelm, Ex. 1021), 1:18-29 (describing systems for tracking the position and orientation), 3:18-52 (HMD with 6DoF head tracking); U.S. Patent No. 10,095,031 (Katz, Ex. 1027), 5:16-23 (describing a VR headset with a mapping module to map the movements of the VR headset), Fig. 2.

2. Binaural Sound and HRTFs

49. 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 December 2015.

50. 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.” Sauk (Ex. 1007), 1:15-16. This type of audio allows a listener to be immersed in an environment and allows for a more realistic experience.

51. 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 ’179 and ’184 Patent calls sound localization points (SLPs), allow a user to “perceive various virtual audio sources as located in various x,y,z locations in world space.” Miller (Ex. 1029), 23:53-59.

52. Binaural sound involves techniques “where the outer ears (the pinnae) are either directly implemented or modeled as digital filters.” Begault (Ex. 1038), x. These digital filters can be designed to contain the directionally-dependent cues used by humans to locate sound. Indeed, by “filtering a digitized sound source with

these filters, one can potentially place sounds anywhere in the virtual space about a headphone listener.” Begault (Ex. 1038), x; Sauk (Ex. 1007) 1:15-21 (“Binaural audio is sound that is processed to provide the listener with a three dimensional virtual audio environment. This type of audio allows the listener to be virtually immersed into any environment to simulate a more realistic experience. Having binaural Sound emanating from different spatial locations outside the listener's head is different from stereophonic sound and it is different from monophonic audio.”); ’179 Patent, 2:18-23 (“The binaural sound localizes to a sound localization point (SLP) in empty space that is away from but proximate to the person.”).

53. Binaural audio may be simulated by use of HRTFs. HRTFs are spatial filters that “describe[] how sound waves interact with the listener’s body” including the head and the outer ear or pinna, from a specific location. VR Book (Ex. 1067), 240; Lalwani (Ex. 1065) (“In addition, sound waves interact with the physical constitution of the listener—the pinna (or outer ear), the head, and the torso—and the surrounding space, creating listener-specific variations otherwise known as head-related transfer function.”); Boren (Ex. 1040), 18-21; Begault (Ex. 1038), 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. ... [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.”) (bold in original); U.S. Patent No. 7,710,654 (Ashkenazi, Ex. 1014), 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.”).

54. Specifically, HRTFs describe the filtering of a sound source before it reaches the eardrum. This filtering includes the spectral modifications caused by sound interacting with the human head, torso, and pinnae, as well as cues of interaural time differences (“ITDs”) and interaural intensity differences (“IIDs,” also known as interaural level differences (“ILDs”)). Begault (Ex. 1038), 41; Fohl (Ex. 1047), 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.”); ’179 Patent, 3:4-17 (“Binaural sound typically accounts for two types of localization cues: temporal cues and spectral cues. Tempora[l] cues arise

from an interaural time difference (ITD) due to spacing between the ears. Spectral cues arise from an interaural level difference (ILD) due to shadowing of sound around the head. Spatial cues are ITDs and ILDs or head-related-transfer-functions (HRTFs).”).

55. The relationship between the physical characteristics of the human head, torso, and pinnae, and the acoustic characteristics known as HRTFs have been studied by POSAs for decades. For example, in 1907, Lord Rayleigh laid the foundation of human sound localization with his Duplex Theory, which included the two fundamental cues of localization: ITDs and ILDs. Begault (Ex. 1038), 31; Lalwani (Ex. 1065) (“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.”).

56. 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.

57. For example, in 1927 Harvey Fletcher and Leon 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. 1011), 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. 1012), 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.”).

58. And in 1931, Fletcher developed a more sophisticated binaural recording device, nicknamed “Oscar,” at Bell Labs. *See* Paul (Ex. 1059), 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.”).

59. And HRTFs themselves were described at least by 1980. Morimoto and Ando (1980). Morimoto (Ex. 1057), 167-174. And by 1989, researchers at the University of Wisconsin found that after measuring HRTFs at a large number of positions around the subject, these HRTFs were used to create digital filters to simulate sound localized to positions around the subject. Wightman (Ex. 1074), 868, 876, 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.”).

60. In the 1980’s and 90’s, scientists at NASA did significant research on HRTFs and spatialized sound. *See, e.g.*, Wenzel 1988a (Ex. 1068); Wenzel 1988b (Ex. 1069); Wenzel 1991 (Ex. 1071). NASA developed the Convolvotron, which was one of the first head-virtual audio systems for its VIEW system, discussed above, which used HRTFs to deliver spatialized sound in a virtual reality

environment. Begault (Ex. 1038),² 165-166. By the 1980s, processors such as the Convolvotron were capable of processing multiple sound sources in real-time using HRTFs. *See* Begault (Ex. 1038), 165-70.

61. 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 (Ex. 1046), 93; *see also* Begault (Ex. 1038), 170-174, (discussing and describing the 3D audio (i.e., binaural sound) processes and capabilities of NASA’s VIEW system).

62. Also, in addition to describing the 3D audio capabilities of the VIEW system, Begault (Ex. 1038) described several other convolution devices that generated and used binaural sound for military aerospace research. *See* Begault (Ex. 1038), 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).

63. Indeed, by 2015, it was well-understood that a pair of HRTFs (one for each ear) may be used to synthesize binaural sound that appeared to come from a particular point in space. VR Book (Ex. 1067), 240 (“Multiple HRTFs from

² Begault (Ex. 1038) is entitled “3D Sound for Virtual Reality and Multimedia,” which describes the signal processing, and system considerations for implementing real-time 3D sound using HRTFs. *See* Begault (Ex. 1038), 105-153.

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 (Ex. 1038), 41.

64. And by 2015, numerous patent publications described VR HMDs that were capable of generating spatial audio with HRTFs. *See, e.g.*, U.S. Patent No. 9,787,846 (Srinivasan, Ex. 1081), 2:3-23, 5:34-67 (headset user device), 6:31-60 (mixed reality applications), 11:48-12:5 (“For example in some embodiments the audio position processor may be configured to apply a head related transfer function (HRTF) using the relative position/orientation information as an input parameter to generate multiple audio output channels from a mono audio input channel.”), Figs. 3, 7; U.S. Patent No. 9,767,618 (Najaf-Zadeh, Ex. 1023), 4:41-65 (HMD with “circuitry [] and applications for providing 3D audio for a HMD” using HRTFs), Fig. 4; U.S. Patent No. 10,509,466 (Miller, Ex. 1029), 23:48-59 (HMD “spatially render[s]” audio “to make the user perceive various virtual audio sources as located in various x,y,z locations in world space—this is using 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. 1030), [0044] (HMD “may apply the selected HRTF to an audio signal (e.g., from an audio player, radio, etc.) to generate an output signal”).

65. It was also well-known that binaural sound enhances the realism and the sense of presence in virtual reality environments. *See, e.g.*, VR Book (Ex. 1067) at 239-240 (“Auditory cues play a crucial role in everyday life as well as VR, including adding awareness of surroundings, adding emotional impact, cuing visual attention, conveying a variety of complex information without taxing the visual system, and providing unique cues that cannot be perceived through other sensory systems. ... For more realistic audio and where auditory cues are especially important, auralization—the rendering of sound to simulate reflections and binaural differences between the two ears—can be used. The result of auralization is spatialized audio—sound that is perceived to come from some location in 3D space”); *see also* Begault (Ex. 1038) at 13-14 (“3-D sound can contribute to the sense of immersivity in a 3-D environment. One might be able to work more effectively within a VR environment if actions were accompanied by appropriate sounds that seemingly emit from their proper locations.”).

3. Processing Binaural Audio

66. Numerous methods of processing binaural audio using HRTFs would have been known to persons of ordinary skill in the art before December 2015, with which a POSA would have understood and known how to generate and process binaural audio.

67. An HRTF is 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 (1038), Chpt. 2.

68. 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 time-domain Head-Related Impulse Response 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, a POSA would understand that one would 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.³ 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 waveforms. Stated mathematically, $X(z) \cdot H(z)$ 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 (Ex. 1038), 106.

69. 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 (Ex. 1076), 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”).
- Srinivasan (Ex. 1081), 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 would 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).
- Najaf-Zadeh (Ex. 1023), 4:41-65 (describing HMD with “circuitry [] and applications for providing 3D audio” processed using HRTFs), Fig. 4.

³ Convolution in the time domain is equivalent to multiplication in the frequency domain.

- Miller (Ex. 1029), 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”);
- Agevik (Ex. 1030), [0044] (describing an HMD that would “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. Transitioning from VR Environment

70. It was well-known to track the position of a user and warn the user when the user is approaching the boundary of a play area. For example, it was known that “[f]or a user approaching the edge of tracking . . . a warning grid (Figure 18.3) . . . can fade in to provide a cue to the user that he should back off.”

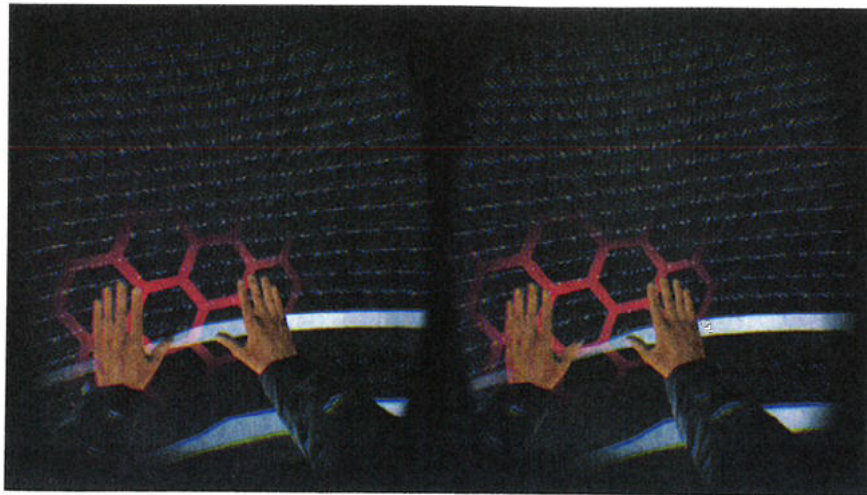


Figure 18.3 A warning grid can be used to inform the user he is nearing the edge of tracked space.
(Image courtesy of Sixense)

VR Book (Ex. 1067), 213; *see also* U.S. Patent No. 7,330,197 (Kobayashi, Ex. 1013), 13:40-60 (“the plane or the boundary line of the warning area 50 may be

rendered by CG . . . to notify the user of this apparatus that the user is near the boundary of the movable area 60.”); U.S. Patent No. 9,851,786 (Pedrotti, Ex. 1080), 4:49-5:24, 12:60-13:11 (“this visual warning is in the form of a grid of lines or a ‘virtual fence’ located at the boundary”), Fig. 11; U.S. Patent Pub. No. 2016/0124502 (Sawyer, Ex. 1031), [0068] (“rendering a grid pattern on the walls of a virtual space . . . to visually alert a user as to the location of physical boundaries”).

71. In addition to displaying the boundary, it was also well-known to transition from the 3D VR environments to other environments or back to a view of the real world. For example, the prior art taught alerting users that they were leaving a VR environment using a gradual fade between the two different environments. *See, e.g.*, Clement (Ex. 1005), 2:43-63. Such transitions did not just involve a visual fade, but also an audio fade. *See* Clement (Ex. 1005), 9:20-31 (“[T]he transition may include a fade of sound content.”); U.S. Patent No. 8,520,872 (Jang, Ex. 1015), 1:57-2:11 (describing shifting sounds from one virtual environment to another in the game *Second Life*); Pedrotti (Ex. 1080), 4:49-5:24, 11:66-12:50, 13:34-45, 15:25-65, 16:20-32.

72. In particular, it was well-known that these transitions involved altering sound processing as the user leaves the first environment. Jang (Ex. 1015), 4:56-5:10 (The virtual reality apparatus may include ... a sound processing unit to *process* and produce audio signals corresponding to the spaces in the virtual environment,

and a second control unit to receive sound source data from the virtual reality server, to detect a space shift of the character in the virtual environment from the first space to the second space, and to control the sound processing unit, such that when the space shift is detected, the sound processing unit decreases a volume corresponding to the first space while increasing a volume corresponding to the second space.”); 7:64-8:15 (“The control unit 114 may further detect virtual space shift of the character and control the operation of the sound processing unit 116 so as to cause the sound processing unit 116 to, for example, gradually decrease the volume of the sound of a first space while gradually increasing the volume of the sound of a second space when the character shifts from the first space to the second space.”); Clement (Ex. 1005), 9:20-31 (“In some embodiments, the transition may include a fade of sound content. For example, when transitioning from the virtual world to the real world, audio content associated with the virtual world may fade out, accompanied by a fade in of ambient sound.”).

73. It was also well-known that altering the processing of sound, for instance switching “from binaural to monophonic or stereophonic” would “remove[] any localization cues for that audio.” Sauk (Ex. 1007), 11:26-41. And there were well-known methods for altering sound processing involved changing from one audio processing method to another. *See, e.g.*, U.S. Patent Pub. No. 2017/0105083 (Nair, Ex. 1033), [0195] (describing that certain devices may have limited

computational resources for binaural processing), [0209] (“At stage 204, for each audio source, the model controller determines which of the binaural synthesis methods will be used for performing binaural synthesis of the audio source. ... The model controller 204 may decide between binaural synthesis methods depending on the results of the resource computation 200 and/or depending on the information associated with the audio source in input 202.”), [0194]-[0224] (generally describing selecting different audio processing methods based on computational resources).

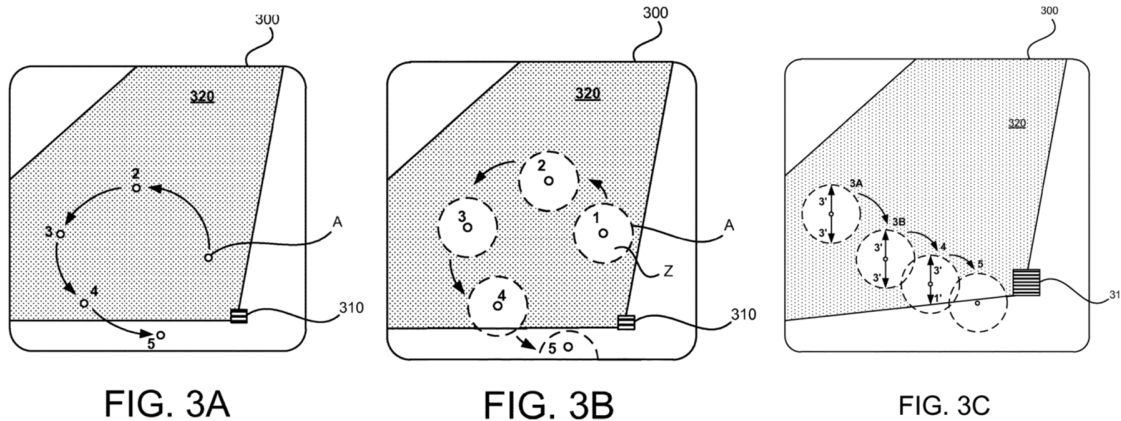
B. Overview of the Prior Art Relied Upon in the Petitions’ Grounds

1. Clement (Ex. 1005)

74. U.S. Patent No. 9,690,374 (“Clement”) is entitled “Virtual/augmented reality transition system and method” and discloses an “[a] system and method of operating an audio visual system generating a virtual immersive experience may include an electronic user device in communication with a tracking device that may track a user’s physical movement in a real world space and translate the tracked physical movement into corresponding movement in the virtual world generated by the user device.” Clement (Ex. 1005), Title, Abstract, Figs. 1A-1C, 3A-3C, 4:6-49 (describing tracking a user’s movements in the physical space). Clement explains that “[i]n an immersive experience, such as an experience generated by a Virtual Reality (VR) system or an Augmented Reality (AR) system, a relatively clear boundary may exist between the immersive experience generated by the VR/AR

system, or the ‘virtual’ world, and the environment outside of the virtual world, or the ‘real’ world.” Clement (Ex. 1005), 1:11-20. To address this, Clement teaches using “[a] smooth, or graceful, transition, for example, from the virtual world to the real world as the user encounters a boundary of the virtual world,” which “may avoid the disorientation and motion sickness which may occur as a user continues to move in the real world, while motion appears to have stopped upon reaching the boundary of the virtual world.” Clement (Ex. 1005), 2:43-63.

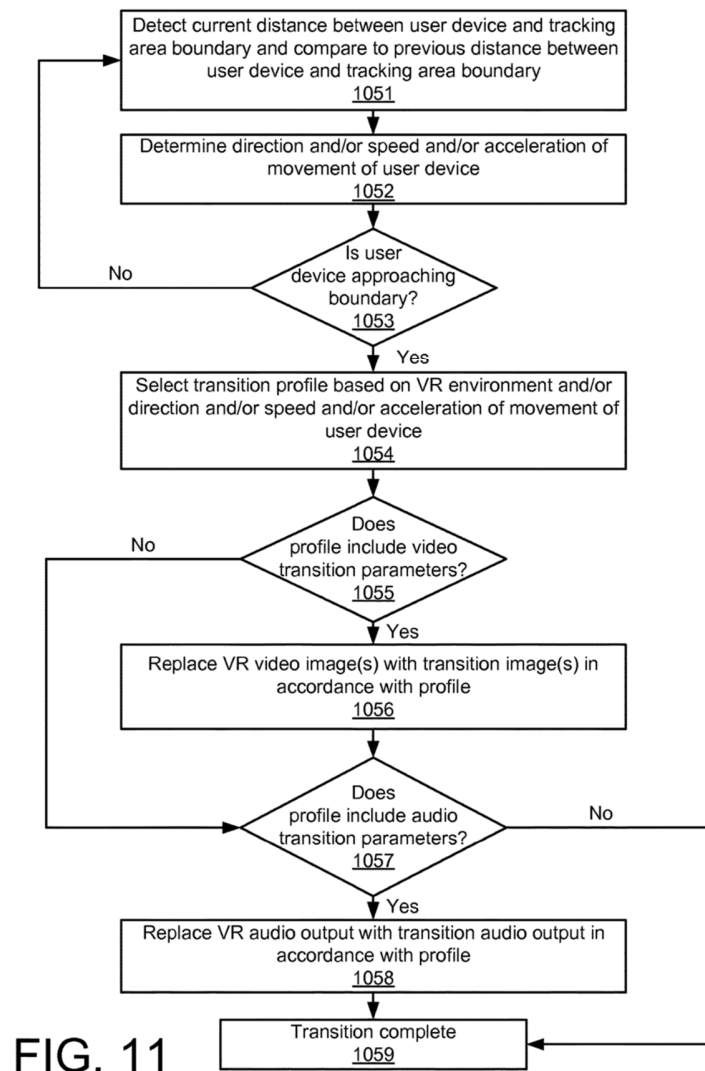
75. Clement describes that “the HMD 100 may include a sensor that detects that the user user's proximity to a boundary of the tracking area 320, or that the user is within a threshold distance of a boundary of the tracking area 320.” Clement (Ex. 1005), 7:4-16; *see also id.* 4:15-7:3, Figs. 3A-3C. “[T]he tracking device 310 may not just track the physical position and movement of the user A in the space 300, but may also detect when the user A is within a predetermined distance, or threshold, of a boundary or edge of the tracking area 320” where “the system may issue a warning or indicator, and/or may automatically initiate a fluid or graceful transition out of the virtual world” when the user approaches the boundary (ex. at position 4 of Fig. 3B). Clement (Ex. 1005), 5:55-6:6.



Clement (Ex. 1005), Figs. 3A-3C. “As the HMD 100 and/or the tracking device 310, or other component of the system, determines that the user is positioned within the preset threshold (in this particular example, 3 feet of the boundary of the tracking area 320, as at position 3B), the system 100 may initiate a gradual transition out of the virtual world. At position 4, the user has moved closer to the boundary of the tracking area 320 (for example, within 1 foot of the boundary as shown in the example illustrated in FIG. 3C). At position 4, the transition out of the virtual world may be at least partially complete, with the transition process fully complete by the time the user reaches position 5 and is fully outside of the tracking area 320.” Clement (Ex. 1005), 6:23-39.

76. “If it is confirmed at block 1053 that the user is approaching/nearing the boundary (and is about to enter an area where the tracking device 310 may not be able to continue to track the user's physical movement in the space 300), a transition profile may be implemented, at block 1054, based on various factors, such as, for example, the virtual environment, the real world space 300 in which the

virtual environment is implemented, direction and/or speed and/or acceleration associated with the movement of the user, and other such factors.” Clement (Ex. 1005), 12:30-46, *see also id.*, 12:20-13:10.



Clement (Ex. 1005), Fig. 11.

77. Clement discloses that “[t]he transition process, to transition the user from a fully immersed virtual reality state at position 3A to a real world state at position 5, may follow a variety of different profiles.” Clement (Ex. 1005), 6:40-57.

The transition profile can include both video and audio transitions. Clement explains that “[i]f, at block 1055, it is determined that the transition profile includes video parameters, then, at block 1056, the virtual image(s) associated with the virtual world displayed on the display 140 of the HMD 100 may be replaced with transition image(s) displayed on the display 140 in accordance with the profile” and “may include, for example, a symmetric or asymmetric fade out of the virtual image(s) accompanied by a fade in of the transition image(s).” Clement (Ex. 1005), 12:47-65. In addition, Clement explains that “[i]f, at block 1057, it is determined that the transition profile includes audio parameters, then, at block 1058, the virtual audio output associated with the virtual world and output to the user via the audio output device 130 of the HMD 100 may be replaced with transition audio output in accordance with the profile” and “this replacement of the virtual audio output with transition audio output may include, for example, a serial fade out of the virtual audio followed by a fade in of the transition audio output.” Clement (Ex. 1005), 12:47-65. Clement discloses that “[t]his transition may also include a transparent display 140 transitioning to a transparent state so that the real world is visible to the user through the transparent display, or a movement of a portion of the HMD 100 out of the field of view of the user so that the real world is directly visible to the user.” Clement (Ex. 1005), 12:47-65, *see also id.*, 12:20-13:10, Fig. 11.

2. Osman (Ex. 1006)

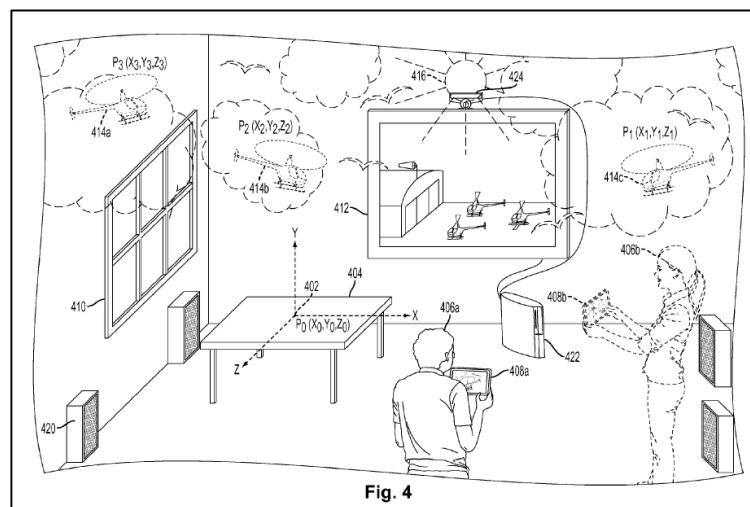
78. U.S. Patent No. 9,037,468 (“Osman”) is entitled “Sound localization for user in motion” and describes systems and methods for “simulating the source of the sound in a three-dimensional space” in order to provide an “augmented-reality environment with realistic delivery of sound.” Osman (Ex. 1006), Title, 1:28-32, 8:65-66.

79. Specifically, Osman teaches the use of HRTFs to create binaural sound. Osman describes that HRTFs “characterize[] how an ear receives a sound from a point in space” and are “the modifications to a sound from a direction in free air to the sound arriving at the eardrum” including “the shape of the listener's outer ear, the shape of the listener's head and body, the acoustical characteristics of the space in which the sound is played, and so on.” Osman (Ex. 1006), 4:48-67. Osman teaches that “[a] pair of HRTFs for the ears may be utilized to synthesize a binaural sound that seems to come from a particular point in space.” Osman (Ex. 1006), 4:48-67. “Sound projection provides a compelling experience for the user” because the user feels “that the sound is coming from a particular point in space.” Osman (Ex. 1006), 5:63-6:5.

80. To create binaural audio, Osman teaches that the system “track[s] the location where the sound is to be originated, as well of the current location of the user, in order to project the sound as if the sound were coming from the simulated

sound origin, which is within a three-dimensional space.” Osman (Ex. 1006), 1:58-67. “For example, the sound may appear to originate at a portable device, an object within a game, a virtual object virtually embedded into and augmented reality space, a player, etc.” Osman (Ex. 1006), 1:58-67.

81. For example, Figure 4 of Osman shows “[p]layer 406 b is playing the same game as player 406 a but in a remote location, and player 406 b is represented as a virtual element in the game for player 406 a.” Osman (Ex. 1006), 8:65-9:11. For each player, the virtual scene is tied to a reference point and “the coordinates of the virtual objects in the virtual scene may be determined with respect to the reference point.” Osman (Ex. 1006), 8:65-9:11.



Osman (Ex. 1006), Fig. 4.

3. Sauk (Ex. 1007)

82. U.S. Patent No. 7,876,903 (“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.

83. Specifically, Sauk (Ex. 1007) teaches that the audio format (or “dimension”) of a sound (i.e., binaural, stereo, or mono) may be used to signify the importance of that sound. Sauk (Ex. 1007), 11:26-40. Sauk describes a binaural audio environment operating in “Selective Relevance Mode” with “a plurality of audio dimensions such as 3-D (binaural), 2-D (stereophonic), and 1-D (monophonic).” Sauk (Ex. 1007), 11:26-40. Because users can perceive audible differences between these different formats, “changing an audio format for sounds from binaural to monophonic or stereophonic audio signifies a different level of relevancy or importance for such audio.” Sauk (Ex. 1007), 11:26-40. For example, “rather than keeping the less relevant audio in the same audio dimension with their desired binaural sound sources, a user could change the less relevant audio to a monophonic (1-D) or stereophonic (2-D) dimension.” Sauk (Ex. 1007), 11:26-40. And as audio processing changes from binaural to monophonic or stereophonic audio, “[t]his process removes any localization cues for that audio.” Sauk (Ex. 1007), 11:26-40. In addition, “[t]he decibel level of the 1-D, 2-D or 3-D audio can

be adjusted to whatever the listener desires for that dimension.” Sauk (Ex. 1007), 11:26-40.

4. Lyren (Ex. 1008)

84. Lyren is the published version of the earliest application in the ’179 and ’184 Patents’ priority chain.

85. As described previously for the ’179 and ’184 Patents, Lyren describes that under various conditions (e.g., when an “event” occurs), it may be desirable to alter the processing of binaural sound, including switching from binaural to stereo or mono sound, reducing the volume of the sound, or altering the binaural sound.⁴ Lyren (Ex. 1008), [0025] (“Example embodiments include systems, apparatus, and methods that change binaural sound in response to an event. When the event occurs, binaural sound changes, such as switching to stereo sound, switching mono sound, switching to altered binaural sound, removing or changing a sound localization point (SLP), moving a SLP (such as moving the SLP from being externally localized to being internally localized), or taking another action in accordance with an example embodiment.”). For example, Lyren teaches the following:

The sound changes from being provided in binaural sound to being provided in stereo sound, mono sound, or altered binaural sound. Alternatively, the sound changes from being provided in stereo sound

⁴ As I will explain with respect to Ground 2 of the ’179 Patent, Lyren does not describe the claimed subject matter of the ’179 Patent because it lacks any disclosure of the claimed detecting/sensing of the user leaving the environment or the claimed alerting/warning the user when that happens. The same is true for the ’184 patent.

or mono sound to being provided in binaural sound. Sound can switch back and forth from being provided in binaural, stereo, and mono sounds (including switching between different variations of binaural sound, such as binaural sounds having different SLPs, different volumes at SLPs, etc.).

Lyren (Ex. 1008), [0067];

The event triggers or initiates a switch from binaural sound to stereo sound, from stereo sound to binaural sound, from binaural sound to mono sound, from mono sound to binaural sound, or from binaural sound to altered binaural sound. For example, a person receives binaural sound with a first codec, and a switch occurs such that the person receives binaural sound with a second codec. As another example, a person receives binaural sound rendered with a first set of HRTFs, and a switch occurs such that the person receives a second binaural sound rendered from a second set of HRTFs. As another example, a person receives binaural sound with a first set of SLPs, and a switch occurs such that the person receives binaural sound with a second set of SLPs. As yet another example, a person receives binaural sound with a first set of background sound, and a switch occurs such that the person receives binaural sound with a second set of background sound. A person can receive binaural sound corresponding to one virtual or real or augmented space, and a switch occurs such that the person receives binaural sound from a second virtual or real or augmented space. As yet another example, after an event is detected, a change to the original binaural sound occurs while still providing the listener with altered or changed binaural sound (such as changing one or more SLPs, ITDs, ILDs, HRTFs, etc. in the original binaural sound while still maintaining binaural sound).

Lyren (Ex. 1008), [0089];

In FIGS. 4-16, example actions include, but are not limited to, one or more of switch the sound from binaural sound to stereo sound, switch the sound from binaural sound to mono sound, switch the sound from stereo sound to binaural sound, switch the sound from mono sound to binaural sound, maintain binaural sound but alter the binaural sound, stop binaural sound, discontinue playing sound, mute the sound, lower a volume of the sound, raise a volume of the sound, “cancel-out” or

quiet a sound or part of a sound by processing it with Active Noise Control (ANC), provide a sound or audio alert, provide a visual alert, move one or more SLPs, adjust or alter a SLP, cancel a SLP, replace a SLP with a different SLP, replace a binaural environment with a different binaural environment, switch one or more codecs, cancel a command, execute a command or instruction, alter a HRTF of a person, change or alter an ITD or an ILD, end a computer program or process, start a computer program or process, provide a notification to a computer program or a person, and other actions discussed herein.

Lyren (Ex. 1008), [0103]; *see also* Lyren (Ex. 1008), [0063]-[0096] (providing additional examples), [0097] (“Figures 4-16 provide examples of events for changing sound from binaural sound to stereo sound, mono sound, or altered binaural sound” and “other types of actions”).

86. Lyren further teaches a wide variety of triggering events. Lyren (Ex. 1008), [0064] (“Example embodiments can switch in response to or based on a wide variety of different types of events.”). For example Lyren teaches:

Example embodiments are not limited to a specific type of an event or a specific time or duration of an event. As noted, such events can be dynamic or static and selected by one or more of a user, a person, apparatus or machine, method, etc. Examples of events and things that can trigger events include, but are not limited to, one or more of a time of day, a calendar day (such as a specific day of the week or day in a month), a location (such as a location of an electronic device or of a person listening to the sound), actions of a third person (such as a person walking into a room), a command or request from a person (such as a person interacting with a user interface to switch the sound), a command or request from a machine (such as a process, software program, intelligent user agent, or intelligent personal assistant commanding, requesting, initiating, or executing the switch), processing power (such as available processing power of an electronic device during a voice exchange or sound localization), bandwidth (such as available transmission and receiving wireless bandwidth of an

electronic device during a voice exchange or sound localization), memory (such as available memory of an electronic device during a voice exchange or sound localization), position or movement or orientation of a person or head of the person (such as direction the person walks or head orientation of the person), distance from the person to an object (such as distance from the person to a wall or an obstruction), available space (such as how much physical 3D space is available to receive and/or localize a sound or voice), safety (such as not localizing sound when the person is driving a vehicle), proximity to or being at a restricted area (such as state, local, or United States Federal regulations prohibiting externally localizing sound while in a certain building or on an airplane), time (such as to switch the sound after a predetermined or given amount of time), a person's identity in a communication (such as to switch a call from binaural sound to stereo when a certain person calls using voice-over internet protocol, VoIP), and other examples provided herein.

Lyren (Ex. 1008), [0065]; *see also* Lyren (Ex. 1008), [0063]-[0096] (providing additional examples); [0097] (“Figures 4-16 provide examples of events for changing sound from binaural sound to stereo sound, mono sound, or altered binaural sound” and “other types of actions”).

87. For example, Lyren teaches triggering an alteration of the sound provided to the user that localizes from the image (e.g., reducing volume and/or shifting from binaural to stereo or mono) based on events such as other people or characters entering or leaving a room or the environment. Lyren (Ex. 1008), [0065] (“Examples of events and things that can trigger events include, but are not limited to, one or more of ... actions of a third person (such as a person walking into a room)”), [0225] (“Events can be triggered by a change in a user’s ability to associate visual cues or images with the associated audio cues, such as a visually rendered

character vanishing from an augmented or virtual environment, the presence or absence of a physical object”), [0232] (“As another example, if there are several people/SLPs in a physical or virtual space, any SLP that travels “off stage” by leaving the room can still be included in the conversation, but the sound can switch to mono or stereo.”). Lyren also teaches altering the sound based on determining “a location (such as the location of an electronic device or of a person listening to the sound).” Lyren (Ex. 1008), [0065]; *see also* Lyren (Ex. 1008), [0123] (“Examples of restricted areas include, but are not limited to, an area, a location, or a point that prohibits SLPs or sound localization, an area in which it is dangerous to localize sound, a vehicle, or other location. Examples of such locations include at or near a construction zone or other inherently dangerous or hazardous area, inside an automobile, on a motorcycle or other motorized vehicle, in a library or a hospital or a sports arena or an elevator or a school or classroom, on a public transport (such as a bus, train, or airplane).”), [0240] (“The intelligent user agent surmises that externally localizing Alice's voice to Bob now might be dangerous for Bob since he is on a steep incline. The HPED receives the call, and the intelligent user agent adjusts the call so Alice's voice internally localizes to Bob through his earphones. In spite of the settings to externally localize Alice's voice, the intelligent user agent made a determination to trump or override the settings and have her voice internally

localize to Bob. This decision was made as being in the best interest of Bob's safety.”).

88. Lyren also states that its techniques “are not limited to a specific type of an event or a specific time or duration of an event,” and that “events can be dynamic or static and selected by one or more of a user, a person, apparatus or machine, method, etc.” Lyren (Ex. 1008), [0065].

89. But as I will explain with respect to Ground 2 of the ’179 and ’184 Patents, Lyren lacks any disclosure of the claimed detecting/sensing of the user leaving the environment or the claimed alerting/warning the user when that happens.

5. Reif (Ex. 1009)

90. U.S. Patent No. 10,416,837 (“Reif”) contains that disclosure that is missing from Lyren. Reif is entitled “Virtual Reality Safety Bounding Box” and describes identifying a safe region in VR with a bounding box. Reif (Ex. 1009), Abstract. The bounding box “encloses an area that allows a user wearing a head mounted display (HMD) to safely view content of a virtual world.” Reif (Ex. 1009), Abstract. The VR system tracks the user’s location and “[w]hen it is determined that the user is located outside the safe region of the bounding box, the VR system can modify what is presented to the user via the HMD, can warn the user, and/or can suggest that the user move back to the safe region.” Reif (Ex. 1009), Abstract.

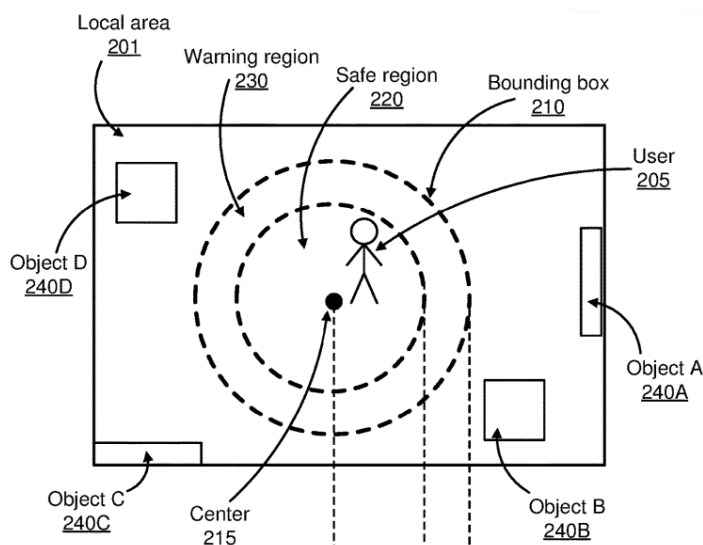


FIG. 2A

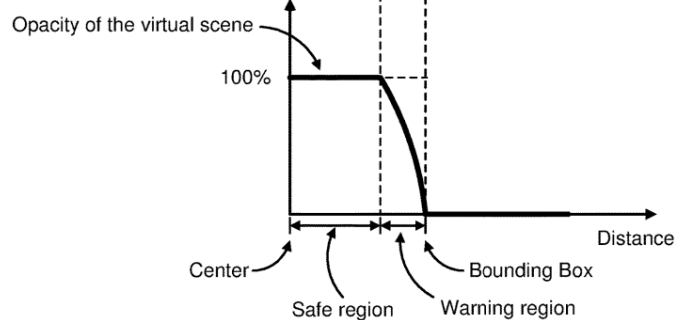


FIG. 2B

Reif (Ex. 1009), Fig. 2B.

VIII. Petition Grounds and Prior Art Combinations

91. Below is a summary of the grounds and related prior art that I will explain in the following sections:

<u>Patent</u>	<u>Ground</u>	<u>Relevant Prior Art</u>	<u>Claims</u>
'179 Patent	Ground 1	Clement with Osman and Sauk	Independent Claims 1, 7, 14

Declaration for *Inter Partes* Review of USP 10,368,179 and 10,448,184

			Dependent Claims 3-4, 6, 8-10, 15-19
'179 Patent	Ground 2	Lyren with Reif	All Claims
'184 Patent	Ground 1	Clement with Osman and Sauk	Independent Claims 1, 7, 14 Dependent Claims 3-4, 6, 8-10, 15-19
'184 Patent	Ground 2	Lyren with Reif	All Claims

92. For clarity, I have annotated the claim elements of the '179 and '184 Patents using numbered and lettered bracketed labels. I have included a full listing of the claims, with labels, in the Appendix at the end of this declaration. I will use these labels to refer to each element through this declaration.

93. I will discuss each element in further detail below, as well as explaining how the prior art discloses each element, or how a POSA would consider the element obvious in view of what the prior art discloses.

94. In some sections of this declaration, I refer to my explanation of a different past or future claim element. I intend that reference fully incorporate my explanation of that past or future element, as if it was copied and pasted in full to where I reference it.

IX. The '179 Patent

A. '179 Patent - Ground 1: Claims 1, 3-4, 6-10, and 14-19 Are Unpatentable Over Clement, Osman, and Sauk

95. I explain below, on an element-by-element basis, how every element of Claims 1, 3-4, 6-10, and 14-19 of the '179 Patent is either disclosed in or obvious over the combination of Clement with Osman and Sauk, as well as why a POSA would have considered it obvious to implement Clement with Osman and Sauk.

1. Reasons to Implement Clement with Osman

96. A POSA would have considered it obvious to, and have good reason to, implement the system described in Clement using 3D or binaural sound, as taught in Osman.

97. Clement discloses “[a] system and method of operating an audio visual system generating a virtual immersive experience may include an electronic user device in communication with a tracking device that may track a user’s physical movement in a real world space and translate the tracked physical movement into corresponding movement in the virtual world generated by the user device.” Clement (Ex. 1005), Title, Abstract, Figs. 1A-1C, 3A-3C, 4:6-49. Clement discloses “an audio visual system configured to generate a virtual immersive experience,” but provides little detail about the audio processing. Clement (Ex. 1005), 1:1-67.

98. Osman discloses details for implementing a binaural audio for virtual objects in an augmented reality space. Osman teaches that “[a] pair of HRTFs for

the ears may be utilized to synthesize a binaural sound that seems to come from a particular point in space.” Osman (Ex. 1006), 4:48-67. “Sound projection provides a compelling experience for the user” because the user feels “that the sound is coming from a particular point in space.” Osman (Ex. 1006), 5:63-6:5.

99. To provide realistic sound in the VR environment, a POSA would have been motivated to use binaural sound, as taught by Osman. For example, a POSA would have understood that “3-D sound can contribute to the sense of immersivity in a 3-D environment.” Begault (Ex. 1038), 13-14. “3D audio or binaural synthesis may be used in applications such as games, virtual reality or augmented reality to enhance the realism of computer-generated sound effects supplied to the user.” Nair (Ex. 1033), [0003]. Further, using binaural sound in VR enables users to “work more effectively” because “appropriate sounds [] seemingly emit from their proper locations.” Begault (Ex. 1038), 13-14.

100. A POSA understood that “[a]uditory cues play a crucial role in everyday life as well as VR, including adding awareness of surroundings, adding emotional impact, cuing visual attention, conveying a variety of complex information without taxing the visual system, and providing unique cues that cannot be perceived through other sensory systems,” that “[s]ounds work well for creating situational awareness and can obtain attention independent of visual properties and

where one is looking,” and that “[s]patialized audio can be useful to serve as a wayfinding aid.” VR Book (Ex. 1067), 239-40.

101. A POSA would have been motivated to deliver audio in a way that improved the immersive VR experience. Smith (Ex. 1066) (“This kind of positional sound is what makes virtual reality experiences so convincing and immersive. ... It’s extremely important. It’s crucial to the experience actually.”). Further, a POSA would have understood that presence and awareness are correlated with the number of modalities used in a virtual simulations (auditory, visual, haptic, and tactile). Slater (Ex. 1062) (describing the importance of multiple sensory modalities (visual, auditory, and kinesthetic) for presence); Marucci (Ex. 1054) (finding that a combination of visual, auditory, and tactile stimulation increases presence more than visual alone).

102. Adding binaural sound would enhance users’ perception of space, distance, and size within the virtual environment.

One of the most important features of VR environments is 3D perception, which includes video and sound. Spatial auditory interfaces in VR and mechanisms for the production of realistic acoustic perception are a big challenge and the main point of interest for many researchers. An accurate spatial reproduction of sound can significantly enhance 3D perception, which is important for achieving the sound localization of visual objects.

Sodnik (Ex. 1063), 54. That enhanced sense of space and distance, in turn, enhances performance, enabling users to better identify and localize objects. Osman (Ex. 1006), 6:42-53.

103. Finally, a POSA would understand that binaural sound was a computationally efficient way of achieving enhanced presence, spatial perception, and performance in virtual environments, while requiring relatively little additional processing power. Begault (Ex. 1038), 11 (“Compared to the synthesis of virtual visual experiences, the transformation of experience via acoustic storage and simulation of spatial audio is not only much easier to accomplish, but is also usually more convincing.”); Kuchera (Ex. 1060) (“If you want to create an immersive experience that runs on your phone, the best way to do that, the most cost-effective way to do that computationally is actually through audio.”); Schell (Ex. 1051), 4 (“[S]ound is what truly convinces the mind is in a place; in other words, ‘hearing is believing.’”); Grimshaw (Ex. 1048), 34 (“3D-positioned audio is a powerful technique to bridge the gap between the virtual game reality and the physical world of the player’s senses.”); Zotkin (Ex. 1076), Abstract (“High-quality virtual audio scene rendering is a must for emerging virtual and augmented reality applications.”).

104. A POSA understood and appreciated that when the sound in VR does not match the imagery, the illusion begins to break down, creating disbelief and detracting from the users’ sense of presence. Smith (Ex. 1066); Lalwani (Ex. 1065).

This sense of presence is all-important in VR as it makes the VR experience compelling and effective. A POSA would have thus been motivated to implement realistic delivery of sound into the 3D VR experience.

105. Thus, a POSA understood that adding binaural audio, as taught by Osman, to Clement would enhance users' (1) perception of space, distance, and size, within the virtual environment, (2) presence, i.e., the sense that the user is actually in the virtual environment, and (3) performance, enabling the user to better identify and localize objects. Osman (Ex. 1006), 6:42-53.

106. A POSA would have considered Clement and Osman together. Both Clement and Osman are directed to presenting VR environments where players interact with virtual objects in a physical operating space. Specifically, Clement relates to "a virtual or augmented reality system." Clement (Ex. 1005), 1:6-7, *see also id.*, Abstract ("A system and method of operating an audio visual system generating a virtual immersive experience may include an electronic user device in communication with a tracking device that may track a user's physical movement in a real world space and translate the tracked physical movement into corresponding movement in the virtual world generated by the user device."), 1:11-20 ("In an immersive experience, such as an experience generated by a Virtual Reality (VR) system or an Augmented Reality (AR) system, a relatively clear boundary may exist between the immersive experience generated by the VR/AR system, or the 'virtual'

world, and the environment outside of the virtual world, or the ‘real’ world.”). Likewise, Osman is directed to “a multi-player augmented-reality environment” where users interact in a “a shared virtual space.” Osman (Ex. 1006), 6:6-35; *see also id.*, 8:65-9:11 (“FIG. 4 illustrates an augmented-reality environment with realistic delivery of sound, according to one embodiment. ... Although player 406 a is inside a room, the virtual reality, also referred to herein as the virtual scene, may extend beyond the physical boundaries of the room.”), 9:12-11:3, Fig. 4.

107. As discussed, Clement discloses “an audio visual system configured to generate a virtual immersive experience,” but provides little detail about the audio processing. Clement (Ex. 1005), 1:1-67. Thus, a POSA would look to Osman’s description of binaural sound in “a multi-player augmented-reality environment” and would understand that Osman’s binaural sound is directly pertinent to Clement’s use of sound in VR environments.

108. Further, a POSA would understand that adding Osman’s binaural sound to Clement’s VR system amounts to no more than applying a known technique (HRTF processing) to a known device ready for improvement to yield predictable results (spatialization of sound). As described in the Technology Background, binaural sound in VR systems including processors using HRTFs to spatialize audio was well-known.

109. Further, a POSA would have had a reasonable expectation of success in combining Clement and Osman because Clement and Osman use similar hardware that performs similar functions. Both use a wearable electronic device. Specifically, Osman discloses headphones which provide localized sound and which are tracked as the user moves. Osman (Ex. 1006), 10:17-26, 11:23-42, 12:1-14. Likewise, Clement describes a head mounted display device (HMD) that is “worn ... on a head of the user, to generate an immersive VR environment” which contains a tracking device to “track the user’s movement in the physical space.” Clement (Ex. 1005), 2:43-47, 4:6-48; *see also id.*, 2:3-8, 2:63-4:5, 7:3-16, Figs. 1A, 1B, 1C, 2. Clement’s HMD delivers both audio and visual content and includes a display, optical components, lenses, and an audio output device, which may include “speakers mounted in headphones and coupled on the frame 120.” Clement (Ex. 1005), 2:63-4:5, Figs. 1A-1C. POSAs understood the headphones of Clement’s HMD would have been configured to deliver binaural sound, as taught by Osman. And a POSA would have known how, as well as had reason, to implement the headphones of Osman in the HMD of Clement. For example, a POSA would have known that such a combination would provide enhanced perception of space, distance, and size, enhanced presence, and enhanced performance within the virtual environment, as described previously. McCulloch (Ex. 1017), 21:21-38 (describing

an HMD that includes a binaural sound engine); Clement (Ex. 1005), 3:30-4:48, 14:3-24; Fig. 2.

110. In addition, Clement describes “a processor 190 to control operation of the components of the control system 170,” which includes both a video control device and an audio control device. Clement (Ex. 1005), 3:20-4:5, Fig. 2. Likewise, Osman discloses using a processor, which can be a general or special purpose processor, for executing a sound localization computer program and/or controlling a sound projection module that modifies sound signals in order to perform sound localization. Osman (Ex. 1006), 12:15-67, Fig. 7. A POSA would have understood that the processor of Clement would have been the special purpose processor taught by Osman.

2. Reasons to Implement Clement and Osman with Sauk

111. Clement teaches a system which “may issue a warning or indicator, and/or may automatically initiate a fluid or graceful transition out of the virtual world” when the user approaches the boundary of a play area. Clement (Ex. 1005), 5:55-6:6, Abstract. A POSA would have combined Clement and Osman with Sauk to provide additional teachings regarding binaural sound, including signaling the importance of things or events to provide flexibility and additional options for different conditions, that enhance the system’s ability to alert or warn the user when the user approaches the boundary or when transitioning from a virtual environment

to the real world. For example, Clement discloses an “audible indicator,” but a POSA would have known that multiple options for audible indications are available and different ones may be preferable in different situations.

112. Sauk teaches that binaural audio may also be used as a cue to “provide the user with understandable information regarding their surroundings.” Sauk (Ex. 1007), 2:45-64. Specifically, Sauk teaches that the audio format (or “dimension”) of a sound (i.e., binaural, stereo, or mono) may be used to signify the importance of that sound. Sauk (Ex. 1007), 11:26-40. Because users can perceive audible differences between these different formats, “changing an audio format for sounds from binaural to monophonic or stereophonic audio signifies a different level of relevancy or importance for such audio.” Sauk (Ex. 1007), 11:26-40. A POSA would understand that because users are able to perceive changes in audio dimension, incorporating a change in audio dimension to the binaural audio warning of Clement and Osman would enhance the user’s ability of the user to perceive the warning. A POSA would have been motivated to make this change in order to enhance the effectiveness of the warning to ensure that the warning would be heeded by the user.

113. In addition, a POSA would have been further motivated to provide more effective auditory displays that serve as audible warnings, as described in Sauk, for the safety of the user and to prevent injury to the user. As discussed in Sauk,

audio warnings may be provided to the user based on proximity to a safety hazard. Sauk (Ex. 1007), 12:33-51 (providing an audible warning when the user is close to a safety hazard); *see also* Pedrotti (Ex. 1080), 15:25-65 (providing an audio warning to “assist the user in avoiding collision with physical objects in the area of the user, and any possible resulting damage to such objects or injury to the user.”).

114. Further, a POSA would have expected to succeed in combining the warning of Sauk with Clement and Osman because like Osman and Clement, Sauk describes a binaural audio system with a “head tracking apparatus 102 ... mounted on a headphone frame 105” with “[s]peakers 104 and 106 ... attached to the headphone frame 105.” Sauk (Ex. 1007), Abstract, 4:33-5:5. A POSA would have implemented Sauk’s binaural audio warning in the Osman binaural audio system, as both devices perform head and position tracking as well as HRTF processing, in analogous ways to localize sound. Sauk (Ex. 1007), 4:33-5:5 (“The head-tracking apparatus 102 housed in the headphone 108 can be any means that provides information regarding the yaw, pitch, roll (orientation) and position of the listener's head 110 to the sound processor. For example, a three-axis gyroscope can be used for determining orientation, and a GPS unit can be used for determining position. The information obtained is provided to a binaural audio processing system.”), 7:24-8:33 (“The audio generator 496 processes enunciated data as necessary to implement the various audio environments selected by a user. In this regard, the audio generator

496 includes digital signal processing circuitry for audio generation of enunciated data. For example, each word or sound specified by the enunciated data can require a specific set of HRTF filters 408, a set of binaural room impulse response (BRIR) filters 410, and a set of reverberation filters 412.”); Osman (Ex. 1006), 11:4-12:14 (describing position tracking with radio-frequency identification (RFID), GPS, or triangulation), 12:36-54 “Head Tracking module 738 determines the position of one or two of the user's ears (which may be determined indirectly by determining a location of the headphones). Head Tracking module 738 may determine the position of the user’s ears using one or more different technologies such as image recognition, RFID, ultrasound, infrared, triangulation, etc.”), 12:15-13:17 (describing sound processing with HRTFs). As described in the Technology Background, a POSA understood that it was well-known how to configure Clement’s HMD to deliver binaural sound, as taught by Osman, and would have been motivated to do so.

115. In addition, Sauk is directed to the same field of endeavor as Clement and Osman: VR environments. Sauk (Ex. 1007), 1:9-13 (“The inventive arrangements relate to the field of audio processing and presentation and, in particular, to combining and customizing multiple audio environments to give the user a preferred illusion of sound (or sounds) located in a three dimensional space surrounding the listener.”), 2:45-64. Both Clement and Osman describe providing spatial information about the environment with audio signals in these VR

environments, and as described above, a POSA would look to Sauk when enhancing these audio signals, as Sauk describes “innovative applications for actually putting such binaural audio information to practical use.” Sauk (Ex. 1007), 2:34-41; Clement (Ex. 1005), Abstract, 1:5-20; Osman (Ex. 1006), 8:65-11:3, Fig. 4.

3. Independent Claim 14⁵

a. Element 14[pre]⁶ A method executed by one or more electronic devices, the method comprising:

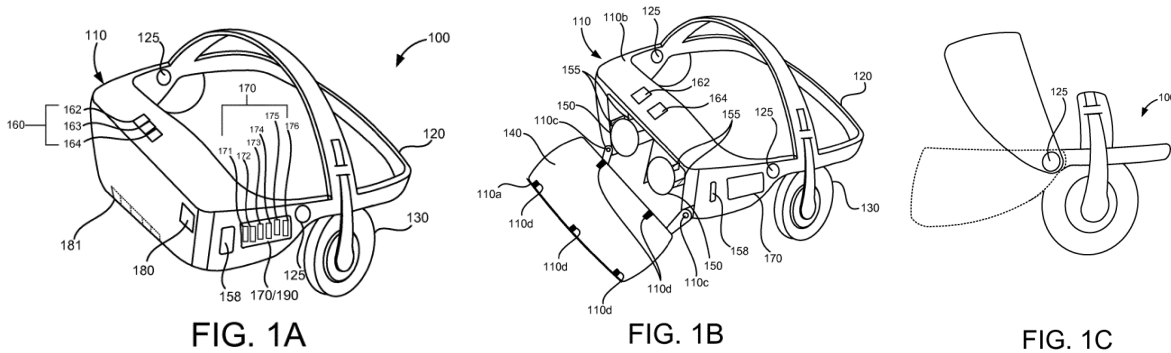
116. Counsel instructed me to assume this preamble was limiting. It is my opinion that Clement discloses the recited “method executed by one or more electronic devices.”

117. Clement discloses “[a] system and method of operating an audio visual system generating a virtual immersive experience may include an electronic user device in communication with a tracking device that may track a user's physical movement in a real world space and translate the tracked physical movement into corresponding movement in the virtual world generated by the user device.” Clement (Ex. 1005), Abstract. Clement further discloses “[a] Virtual Reality (VR) system and/or an Augmented Reality (AR) system may include, for example, a head mounted display device (HMD) or similar device worn by a user, for example, on a

⁵ Claim 14 is the broadest claim and will be addressed first.

⁶ See Claim Listing (Appendix) for assigned reference letters.

head of the user, to generate an immersive VR environment to be experienced by the user.” Clement (Ex. 1005), 2:43-62;



Clement (Ex. 1005), Figs. 1A-1C; *see also* Clement (Ex. 1005), 2:63-4:14.

118. Likewise, Osman teaches “[m]ethods, apparatus, and computer programs for simulating the source of sound are provided,” using a headset and a portable device, as shown in Fig. 6.

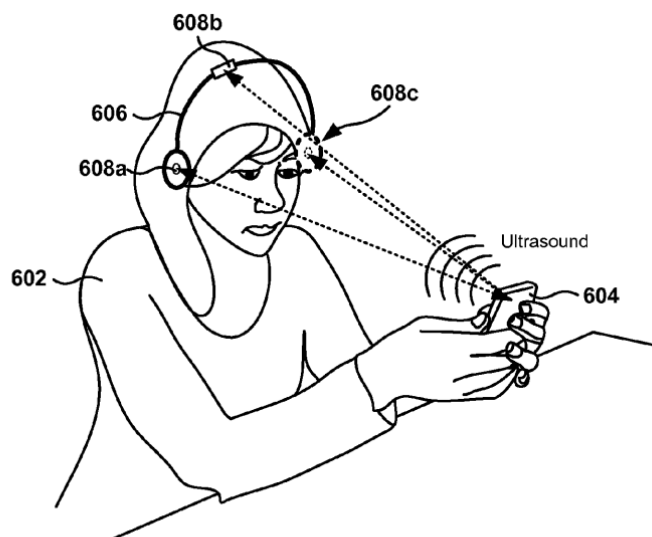
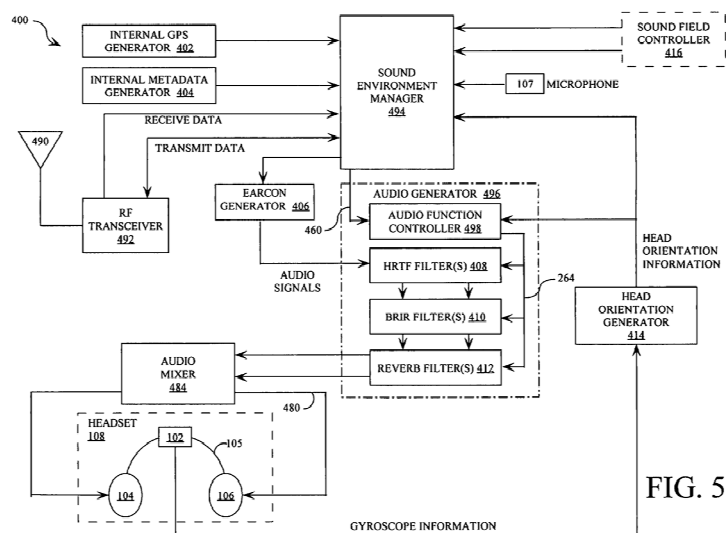


Fig. 6

Osman (Ex. 1006), Abstract, 3:60-4:5, Fig. 6; *see also* Osman (Ex. 1006), Figs. 1-5, 7.

119. Sauk also teaches a “[m]ethod and apparatus for producing, combining, and customizing virtual sound environments” using a headset with a binaural sound system, as shown in Fig. 5.



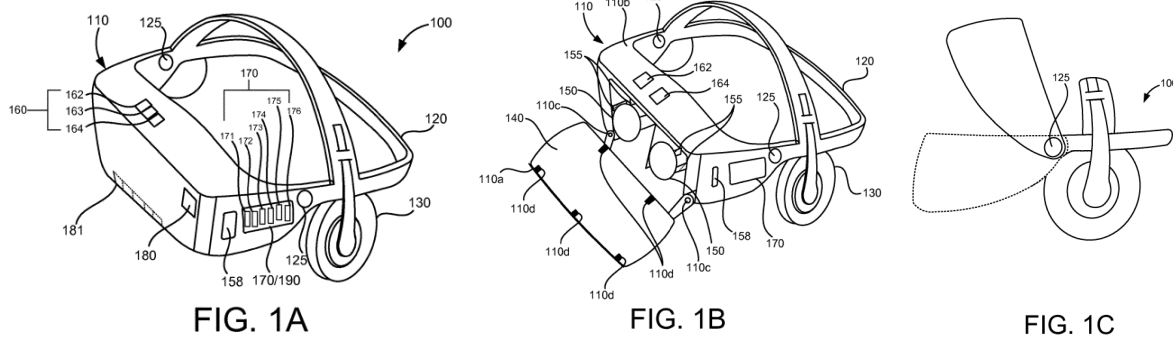
Sauk (Ex. 1007), Abstract, 5:41-57, Figs. 4-5.

b. Element 14[a] storing head-related transfer functions (HRTFs) in memory;

120. Clement and Osman discloses and/or renders obvious “storing head-related transfer functions (HRTFs) in memory.”

121. Specifically, Clement describes a head mounted display device (HMD) that is “worn ... on a head of the user, to generate an immersive VR environment.” Clement (Ex. 1005), 2:43-47. Clement’s HMD delivers both audio and visual

content and includes a display, optical components, lenses, and an audio output device, which may include “speakers mounted in headphones and coupled on the frame 120.” Clement (Ex. 1005), 2:63-4:5,



Id., Figs. 1A-1C; *see also* Clement (Ex. 1005), 2:3-5 (“FIGS. 1A and 1B are perspective views and FIG. 1C is a side view of a head mounted display device, in accordance with an embodiment broadly described herein.”), 2:6-8 (“FIG. 2 is a block diagram of the head mounted display device, in accordance with an embodiment broadly described herein”), Fig. 2.

122. Clement describes “a processor 190 to control operation of the components of the control system 170” of the HMD, which includes both a video control device and an audio control device as well as “one or more memory devices for storing instructions and data.” Clement (Ex. 1005), 3:20-4:5, 14:3-23,

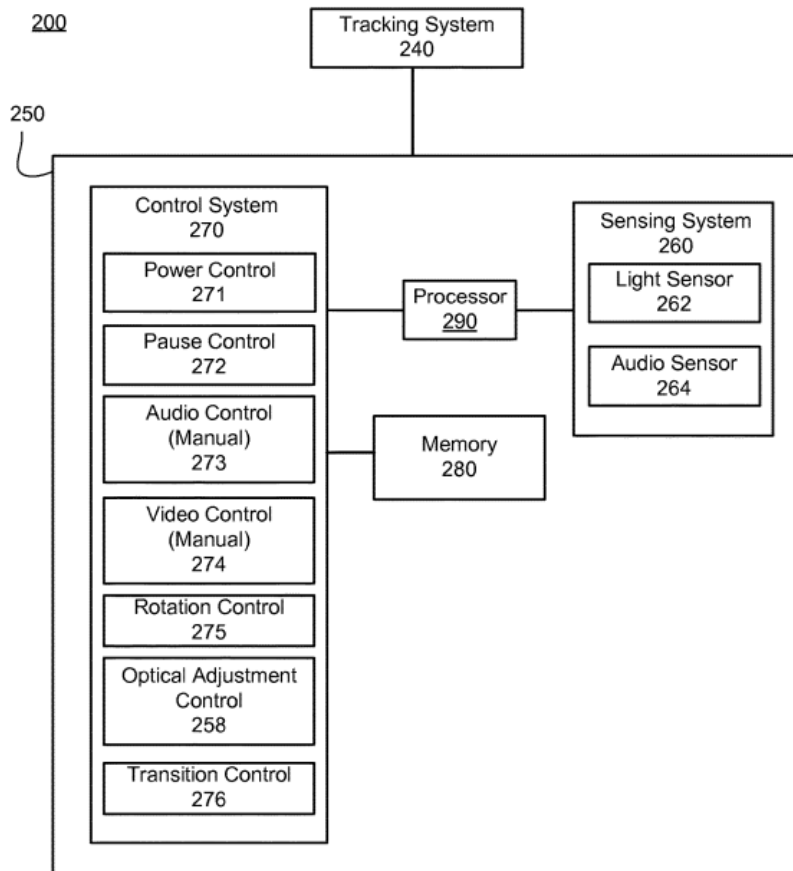


FIG. 2

Clement (Ex. 1005), Fig. 2. A POSA would understand that the memory of Clement would have been used to store HRTFs for processing binaural audio. A POSA would understand that it was well-known for HRTFs for processing binaural audio to be stored in memory.

123. For example, Osman discloses headphones which provide localized sound. Osman (Ex. 1006), 10:17-26, 11:23-42, 12:1-14. Fig. 2. Osman teaches using either standard HRTFs or individualized HRTFs calibrated to the user to generate this localized sound. Osman (Ex. 1006), 5:18-28. Osman explains that its

system includes “processor 732, which is coupled to memory 734” and “[s]ound localization computer program 736 [which] resides in memory 734, but can also reside in permanent storage device 758.” Osman (Ex. 1006), 12:15-67, Fig. 7. When stored in permanent storage, such as on a hard drive, the computer programs must first be moved to computer memory before processing by the computer system. *See, e.g.,* Osman (Ex. 1006), 22:43-62.

124. Osman also teaches storing the HRTFs in a database. Specifically, Osman teaches that calibration methods that may be used to customize the sound localization experience for the particular user or to develop an HRTF for the user. Osman (Ex. 1006), 5:18-28. For example, Osman describes that HRTFs are stored in memory at database 710, which “includes a plurality of sound localization functions associated with a plurality of different users.” Osman (Ex. 1006), 13:10-18,

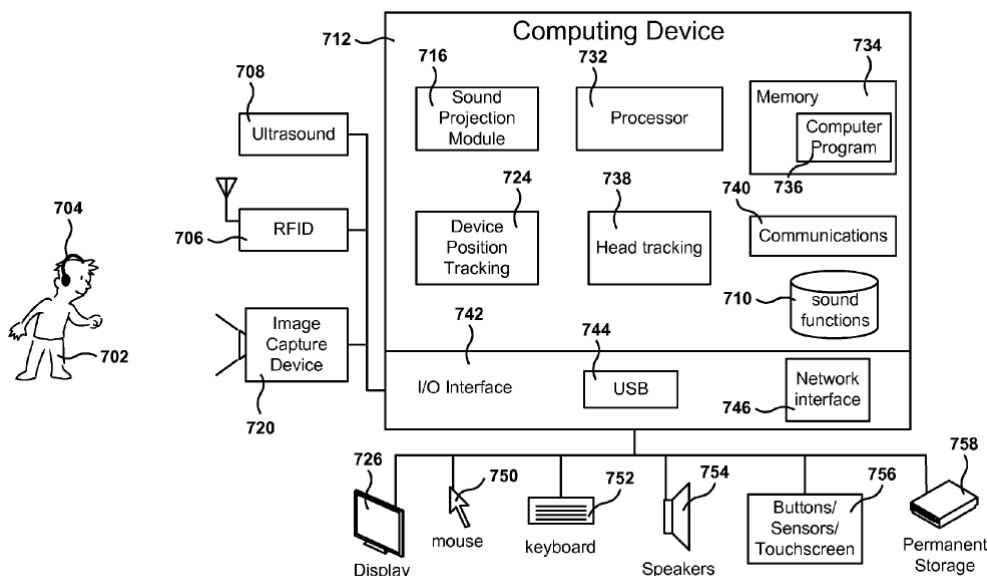


Fig. 7

Osman (Ex. 1006), Fig. 7. From these HRTF databases, Osman describes calibration methods that “create a sound localization function for the user that is based on one or more HRTF’s from the database” by finding HRTFs that “work” for the user. Osman (Ex. 1006), 14:29-45; *see also* Osman (Ex. 1006), 14:51-19:36, Figs. 9A-9B, 10A-10B, 11-12.

125. A POSA understood that storing HRTFs in memory was well-known and used in instances, such as Osman, where a user’s customized HRTF is retrieved from a larger database containing customized HRTFs for different users. *See, e.g.*, U.S. Patent No. 9,544,706 (Hirst, Ex. 1019) at 8:13-23 (“The applications 560, 536, which may be executed on client devices 516 a and 516 n, respectively, may send a communication to the computing device(s) 502 via the communications network 514 providing information identifying the user and requesting the customized HRTF

associated with the user's profile. The computing device(s) 502 may then send a communication with the customized HRTF data to the client devices 516 a and 516 n via the communications network 514. The client devices 516 a and 516 n may store the customized HRTF in an HRTF cache 550, 554.”), Fig. 5; U.S. Patent No. 10,070,239 (Romigh, Ex. 1025) at cl. 1 (claiming “[a] system for generating a personalized Head-Related Transfer Function”), 6:52-56 (“the memory 36 may include one or more levels of data, instruction, and/or combination caches, with caches serving the individual processing unit or multiple processing units (not shown) as is well known in the art.”).

126. Moreover, a POSA would have stored the HRTFs in the HMD’s memory to reduce latency, facilitate signal processing, and enhance the user experience. Storing HRTFs in memory of the HMD would avoid the need to connect to a remote memory device, which would have introduced delay or latency. *See, e.g.,* Fohl (Ex. 1047), 312-313 (“The active FIR filter coefficients are stored in a dualported RAM (DPRAM), so the coefficient update and the filtering may be executed asynchronously.”).

127. In particular, when storing HRTFs in memory, it was well understood that “[i]t is good to avoid remote memory accesses” and “[p]roper placement of data will increase the overall bandwidth and improve the latency to memory.” Lameter 2013 (Ex. 1053), 1. Specifically, it was known that “[t]oday, processors are so fast

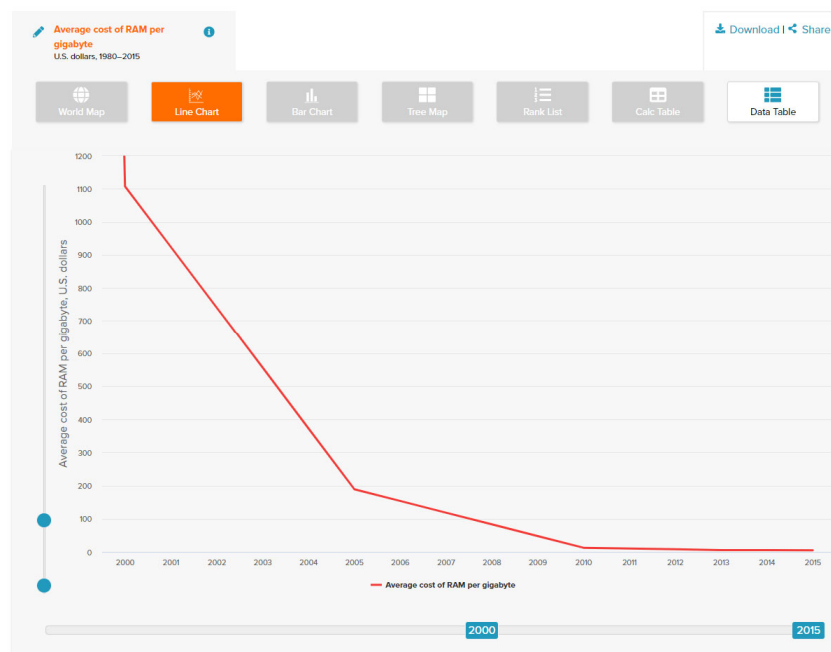
that they usually require memory to be directly attached to the socket that they are on” as opposed to accessing data stored in remote memory because “[a] memory access from one socket to memory from another has additional latency overhead to accessing local memory—it requires the traversal of the memory interconnect first,” while “[o]n the other hand, accesses from a single processor to local memory not only have lower latency compared to remote memory accesses but do not cause contention on the interconnect and the remote memory controllers.” Lameter 2013 (Ex. 1053), 1; *see also* Lameter 2006 (Ex. 1052), (“The optimal placement of memory for a process is through node local allocations. In that case the memory access has the smallest latency possible and traffic on the NUMA interlink (a precious centralized resource) does not have to occur at all.”); Dashti (Ex. 1045), 381-394 (“Fortunately, high latencies can be avoided or substantially reduced if we carefully place memory pages on nodes so as to avoid traffic congestion.”). As such, a POSA would understand that HRTFs should be stored in memory local to the processors, in the HMD, to avoid latency.

128. Further, a POSA understood that latency would detract from the realism of the environment, reduce localization and situational awareness, and cause motion sickness in users. Therefore, a POSA would be motivated try to avoid and/or compensate for latency in the audio and visuals for the virtual reality environment.

Declaration for *Inter Partes Review* of USP 10,368,179 and 10,448,184

Wu (Ex. 1075), 1; VR Book (Ex. 1067), 183-185 (“VR latency is a major contributor to motion sickness ...”).

129. Further, between 2000 and 2015, memory cost fell by a ratio of 224:1, facilitating the developing of smaller, but more powerful devices.



HumanProgress (Ex. 1037) (showing cost per gigabyte over time). By 2015 memory was small enough to fit into an untethered HMD suitable for VR. *See* James (Ex. 1064) (showing an “entirely self contained” HMD including processing and memory). Thus, by 2015, a POSA would have known that memory would have been compressed into sizes small enough to be incorporated into an HMD. For example, by 2015, untethered HMDs with binaural sound had already begun to appear, and state-of-the-art HMDs included stand-alone HMDs without the use of an additional PC. *See, e.g.*, Anthes (Ex. 1036), 3, 5 (describing the state of the art in virtual reality,

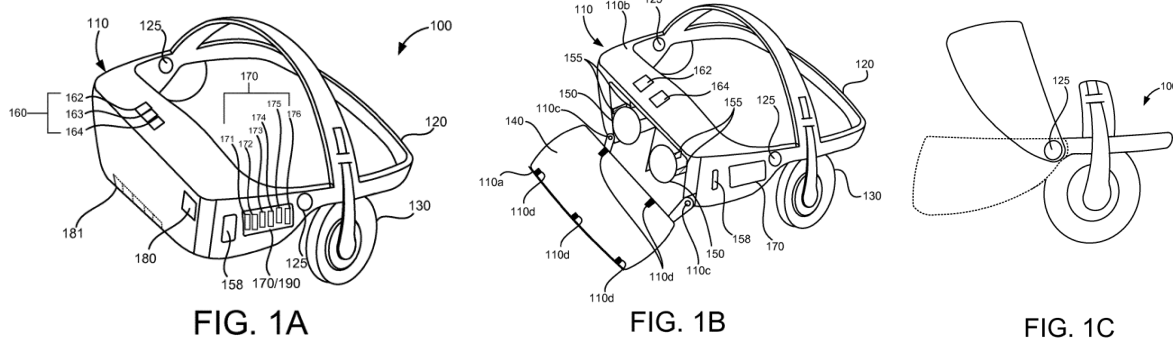
including “Mobile HMDs,” such as Gameface and Auravisor, that are “able to be used without an additional PC”); Emrich (Ex. 1039) (describing Sulon Technologies’ Cortex, an untethered HMD). Further, Clement teaches that its HMD includes processors and memory, showing that a POSA would have been able to use memory in the HMD to store HRTFs. Clement (Ex. 1005), 4:15-60, 13:40-14:24.

130. As discussed, a POSA would have understood that HRTFs would have been stored in memory in the HMD itself, and been motivated to do so to allow for faster and more efficient processing. Clement teaches that its HMD includes a memory and processor, and as discussed in the Technology Background, the prior art taught HMDs that spatialized sound can be created using HRTFs. Clement (Ex. 1005), 3:20-4:5, 14:3-23, Fig. 2. Because all these elements were well-understood, a POSA would have known how to incorporate the cheap, small, and large capacity memory available at the time in an HMD, and how to use this memory to store HRTFs. Thus, a POSA would have had a reasonable expectation of success in incorporating these teachings. Clement (Ex. 1005), 3:30-4:48, 14:3-24; Fig. 2.

c. Element 14[b] displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

131. Clement with Osman teaches “displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment.”

132. Specifically, Clement teaches “a head mounted display device (HMD) or similar device worn by a user, for example, on a head of the user, to generate an immersive VR environment to be experienced by the user” in which “[m]ovement of the user in the real world may be translated into corresponding movement in the environment generated in the virtual world.” Clement (Ex. 1005) 2:3-62. To do this, Clement teaches that the HMD includes “[a] display 140, such as, for example, a smartphone or other display device, may be mounted on the front face 110 a of the housing 110.” Clement (Ex. 1005), 2:63-3:19,



Clement (Ex. 1005), Figs. 1A-1C,

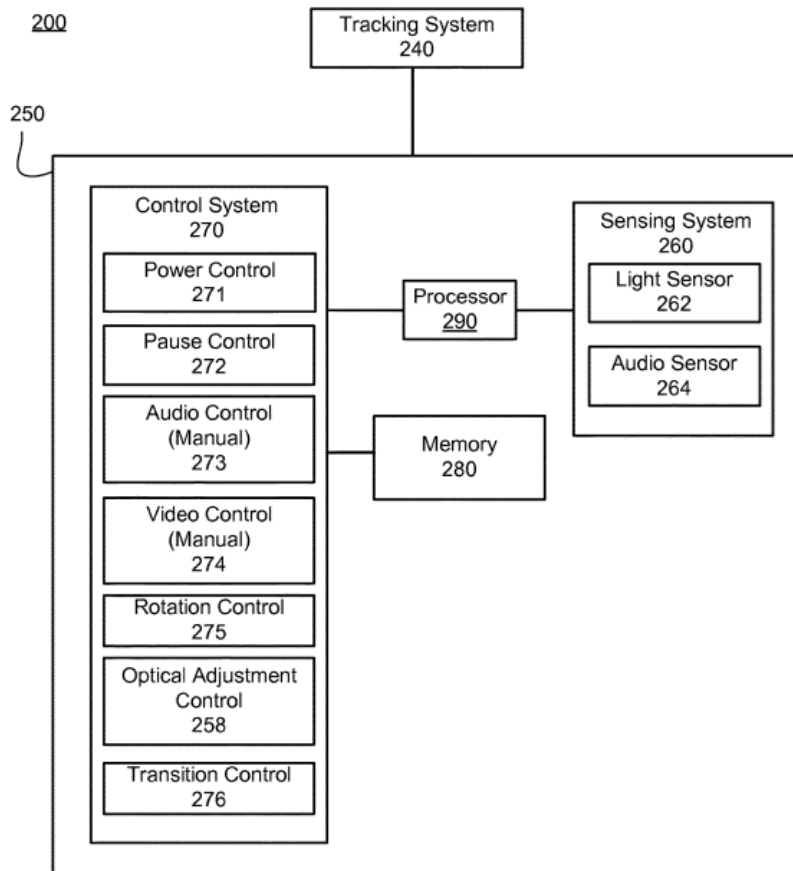


FIG. 2

Clement (Ex. 1005), Fig. 2.

133. Clement describes a “virtual world” in which the user is able to move freely. Specifically, Clement describes that “[t]he system may track the user’s movement in the physical space, or the ‘real’ world, and cause the virtual world to ‘move’ in coordination with the user’s movement in the real world.” Clement (Ex. 1005), Abstract, 4:6-14, 6:7-25 (describing tracking the user). A POSA would understand the virtual world in Clement is “a three-dimensional (3D) virtual environment.”

134. Clement further describes that “virtual images [are] displayed on the display 140 of the HMD 100.” Clement (Ex. 1005), 8:36-9:6. Figures 7 to 9 of Clement illustrate the display of a virtual environment by the HMD to the user, which includes virtual images.

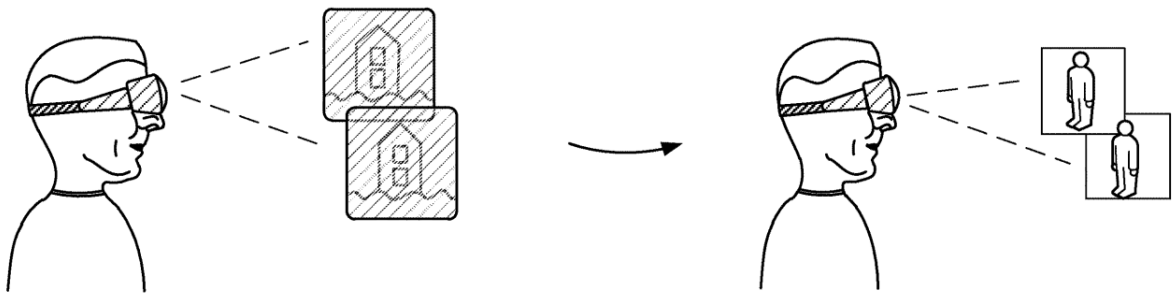


FIG. 7

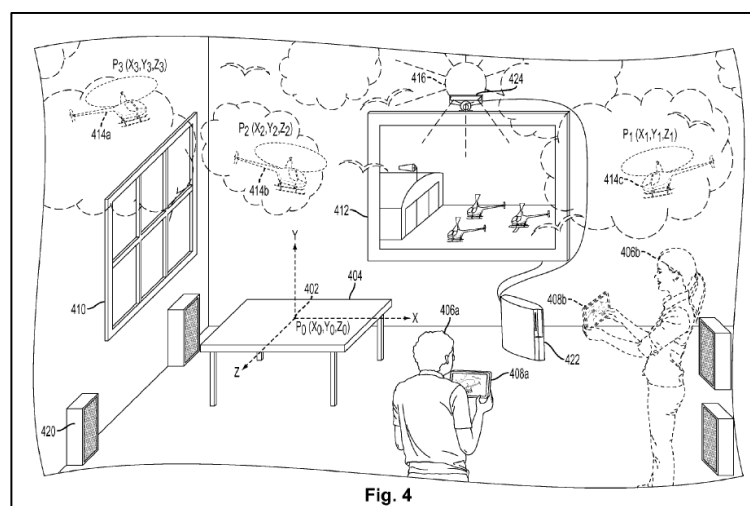
Clement (Ex. 1005), Fig. 7; *see also* Clement (Ex. 1005), at Figs. 8-9. In particular, Clement describes that as the user transitions from the virtual world to the real world, the HMD displays “the fade out of the virtual images displayed on the display 140 of the HMD 100 and the fade in of ... the images of the real world.” Clement (Ex. 1005), 9:20-31, *see also* Clement (Ex. 1005), cls. 4, .13, 17, 18, 21.

135. A POSA would understand HMDs for displaying “a three-dimensional (3D) virtual environment that includes a virtual image” were well-known as discussed in the Technology Background, and included components such a display, lenses, or other optical components. *See, e.g.*, Fisher (Ex. 1046), 92-93. Because

the HMD of Clement includes a display, a POSA would understand that it would have been used to display “a three-dimensional (3D) virtual environment that includes a virtual image.” Clement (Ex. 1005), 2:63-3:19.

136. Like Clement, Osman also discloses that the user is immersed in a “virtual reality” or a “virtual scene” (i.e., the claimed “environment”) with “virtual objects” including “helicopters 414a-414c, clouds, birds, sun 416, etc.” (i.e., the claimed “virtual image”). Osman (Ex. 1006), 8:65-9:56. Using the locations of the user’s ears and the perceived sound origin of the virtual object, “[t]he HRTF of the user is used to convert the sound into a localized sound that appears to the user as coming from the sound origin” and delivered to the user through headphones. Osman (Ex. 1006), 10:3-32, *see also* Osman (Ex. 1006), 9:57-10:32, Fig. 4.

137. Osman teaches that its virtual images are displayed on “a portable device.” Osman (Ex. 1006), 8:65-9:56,



Osman (Ex. 1006), Fig. 4. And as discussed in the Technology Background, POSA would have understood that displaying virtual and augmented reality images in the display of an HMD was well-known. *See, e.g.*, Fisher (Ex. 1046), 92-93. A POSA would have understood that the virtual images of Osman would have been displayed on displays such as those of Clement's HMD, instead of on a portable device. In fact, many HMDs capable of doing so existed, as discussed in the Technology Background. And a POSA would have been motivated to do so in order to create more realistic virtual and augmented reality experiences.

d. Element 14[c] processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes in empty space from the virtual image in the environment;

138. Clement with Osman renders obvious “processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes in empty space from the virtual image in the environment.”

139. Clement with Osman teaches or renders obvious “processing, by a digital signal processor (DSP).” Osman teaches that processing spatialized audio “may be performed with a digital processing system, such as a conventional, general-purpose computer system” and that “[s]pecial purpose computers, which are designed or programmed to perform only one function, may be used in the alternative.” Osman (Ex. 1006), 12:15-27, 12:55-67,

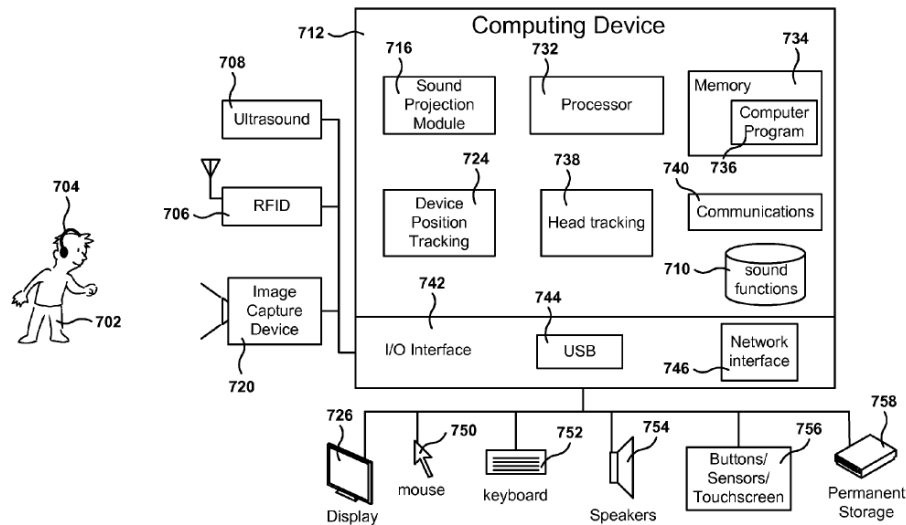


Fig. 7

Id., Fig. 7. A POSA would have understood that the “digital processing system” described in Osman would have been, would have included, or would have been obvious to include, a digital signal processor (DSP). In particular, a POSA would have understood that Osman’s “special purpose processor” designed to perform digital signal processing functions is a DSP. *See Technology Background.*

140. A POSA would have also looked to Sauk, which describes the use of “one or more digital signal processors” to implement processing of spatial audio. Sauk (Ex. 1007) at 7:24-45.

141. A POSA would have been motivated to use a DSP for processing spatialized audio. Using DSPs for this purpose was well-known. For example, a POSA understood that “[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.” Begault (Ex. 1038), 100. Begault describes a

“rudimentary, mostly nonmathematical overview of the theory behind DSP algorithms used for implementation of 3-D sound, particularly for implementation within an integrated system” which “should enable the reader to conceptualize an overall system design for a particular application.” Begault (Ex. 1038), 101; *see also* U.S. Patent No. 5,173,944 (Begault '944, Ex. 1093), 6:51-56 (“The pseudo-stereophony effect of the present invention can be implemented into hardware using off-the-shelf integrated circuit chips, such as a Motorola 56001 DSP chip. All of the processing described with respect to FIG. 2 can be accomplished with these or equivalent DSP chips.”); U.S. Patent No. 5,438,623 (Begault '623, Ex. 1094), 5:6-26 (“It is important to note that both interaural time delay and interaural magnitude of the audio signals function as primary perceptual cues to the location of sounds in space, when convolved, for example, with monaural speech or audio signal sound sources. Accordingly, the digital filter 161 as well as the other digital filters 162, 163 and 164 are comprised of digital signal processing chips, e.g. Motorola type 56001 DSPs that access interchangeable PROMs, such as type 27C64-150 EPROMs manufactured by National Semiconductor Corp.”), 8:4-15 (“Accordingly, thus what has been shown and described is a system of digital filters implemented using selectively interchangeable PROM-DSP chip combinations which generate synthetic head related transfer functions that impose natural cues to spatial hearing on the incoming signals, with a different set of cues being generated for each

incoming signal such that each incoming stream is heard at a different location around the head of a user and more particularly one wearing headphones.”); Begault (Ex. 1038), 20 (“Changing the character of an input signal in the digital domain to create virtual spatial hearing experiences has been greatly facilitated by the development of DSP chips, particularly since they can be configured as filters that modify sound in the same way as in natural spatial hearing.”), 109-116 (describing using a DSP for HRTF processing), 165-170 (same), 185-187 (same); U.S. Patent No. 8,520,873 (Mahabub, Ex. 1016), 1:25-33 (“The psychoacoustic effects from GenAudio’s AstoundSound™ technology may be achieved through the application of digital signal processing (DSP) for head related transfer functions (HRTFs).”); U.S. Patent No. 9,578,440 (Otto, Ex. 1020), 5:37-39 (“The binaural processing system may further comprise a binaural processor which computes the left HRTF and right HRTF in real-time.”); Wenzel 1992 (Ex. 1072), 88-89, 92-93 (describing the Convolvotron); *see also* Technology Background.

142. A POSA would have been motivated to use a DSP for HRTF processing for several reasons. First, “[a]udio DSP chips typically are designed to perform a limited set of digital DSP instructions *very quickly*—i.e., in real time” allowing for fast and efficient processing for spatialization of audio. Begault (Ex. 1038), 100. Moreover, DSPs may be optimized to perform complex operations such as accepting real-time signals and turning them into computationally actionable data. POSAs

would have been motivated to provide HRTF spatialized audio as efficiently and quickly as possible to enhance realism in VR, and would have been motivated to use a DSP to do so, given its known advantages.

143. Second, DSPs excel at executing signal processing tasks, like using HRTFs to spatialize audio, efficiently. A POSA would understand that DSPs allow processing to be done quickly and using lower power. Further, DSPs' efficiency meant sound would have been spatialized using low power, which meant little impact on battery life and temperature of the HMD. OnQ (Ex. 1073), 1-3 ("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."); Gwennap (Ex. 1078), 3.

144. Third, a POSA would have understood that providing that task-specific hardware would guarantee the availability of the processing power necessary for audio processing. *See, e.g.*, Begault (Ex. 1038), 20. Instead of sharing a processor with other processing functions, use of having dedicated hardware such as a DSP ensures that the audio processing is always being done. A POSA would understand that using a DSP would ensure fast and efficient audio processing was dedicated to this task regardless of other processing demands. In addition, a POSA would have

understood that another advantage of DSPs is “ease of reconfiguration [such that] one can write different software specifications to the signal processor, rather than resoldering hardware components.” Begault (Ex. 1038), 20. A POSA understood that DSPs would have been configured as filters that modify sound in the same way as in natural spatial hearing, thus ensuring that the sound provided was realistic. Begault (Ex. 1038), 20.

145. Finally, a POSA would have known that DSPs would have been integrated into the HMD. For example, because of the small hardware size of DSPs, DSPs had been incorporated into smart phones to provide fast and efficient processing of sound, including HRTF processing. Cheng (Ex. 1056), 1-3 (describing the DSP in Qualcomm’s Snapdragon processor); Gwennap (Ex. 1078), 1. Indeed, dedicated DSPs existed in the early 2000’s to process spatial audio - mostly for gaming. For example, the Creative Soundblaster X-Fi had EAX 5.0 technology and by 2015 would have processed 128 binaural sound sources using HRTFs. Liadov (Ex. 1050).

146. A POSA would have been motivated to include the DSP in the HMD as described for Element 14[a]. A POSA understood that performing processing of spatialized sound in an untethered HMD (as opposed to remotely at a separate device) would reduce communication latency. *See, e.g.*, Miller (Ex. 1055), 113 (describing “Client/Server Communication Latency” in VR systems contributing to

as much as 3 ms latency); Brungart 2004 (Ex. 1041), 1 (describing latency arising from “the communications delay between that device and the audio display”). A POSA would have been motivated to reduce latency because it was well known that latency detracts from the user’s experience in interactive virtual environments. Fohl (Ex. 1047), 312 (describing spatial audio rendering requirements, including “Audio latency ≤ 10 ms to avoid perceivable delay”); Brungart 2005 (Ex. 1042), 37 (describing loss of realism and increase in fatigue or annoyance from latency). And as described previously, the efficiency of the DSP allowed processing of spatialized sound to be done using low power, which meant little impact on battery life and temperature of the HMD. OnQ (Ex. 1073), 1-3.

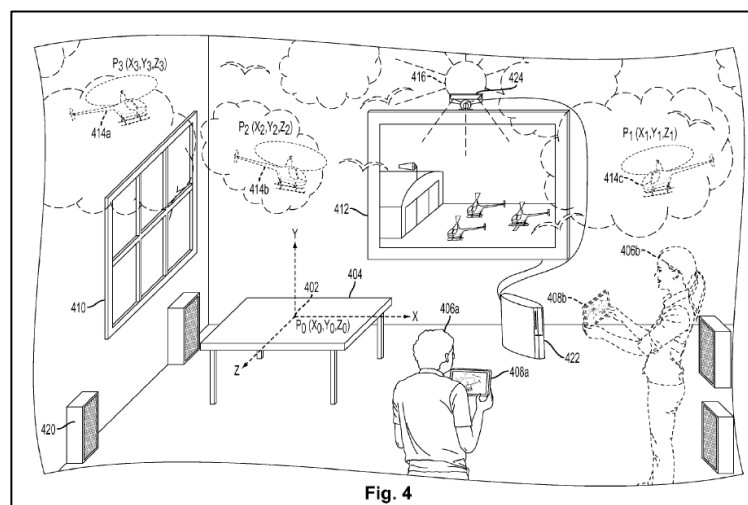
147. A POSA would have understood by 2015, DSPs had numerous advantages and in particular were small and would have been incorporated into an HMD easily. Begault (Ex. 1038), 20 (“Other advantages of DSP include relatively smaller hardware size, greater predictability of noise, and ease of reconfiguration—one can write different software specifications to the signal processor, rather than resoldering hardware components.”); McCulloch (Ex. 1017), 8:17-39, Fig. 2A (describing a HMD with a digital processing unit). The advantages of a DSP described above would have encouraged a POSA to use a DSP for processing spatialized sound and, given the well-known and long use of DSPs for this purpose, a POSA would have had a reasonable expectation of success in doing so.

148. A POSA would have understood that using DSPs to process HRTFs amounts to no more than combining prior art elements according to known methods, or applying a known technique (HRTF processing using DSPs) to a known device, to yield predictable results. DSPs were, at the very least, one of a finite number of options for HRTF processing and would have been obvious to try. *See, e.g.*, Begault (Ex. 1038), 20 (“Simultaneously, the development of DSP chips, which are capable of millions of computational instructions per second, have over time been increasingly used in everyday audio and graphics applications”); Osman (Ex. 1006), 12:55-67 (“It should be appreciated that processor 732 may be embodied in a general-purpose processor, a special purpose processor, or a specially programmed logic device.”).

149. Further, Clement with Osman teaches or renders obvious “processing ... sound with the HRTFs to generate binaural sound that externally localizes in empty space from the virtual image in the environment.” Osman teaches that “[o]ne example of a sound localization function is a Head-Related Transfer Function (HRTF), which is a response that characterizes how an ear receives a sound from a point in space” and “[a] pair of HRTFs for the ears may be utilized to synthesize a binaural sound that seems to come from a particular point in space.” Osman (Ex. 1006), 4:37-67. A POSA would have understood that HRTFs are used to generate binaural sound that is provided to the user. Begault (Ex. 1038), 40-41, 49-50; ’179

Patent, 3:52-55 (“As a general rule, binaural sound accommodates for or is derived from one or more ITDs, ILDs, HRTFs, natural ear spacing, and head shadow.”); 3:55-58.

150. As described for Element 14[b], Osman discloses that the user is immersed in a “virtual reality” or a “virtual scene” with “virtual objects” including “helicopters 414a-414c, clouds, birds, sun 416, etc.” Osman (Ex. 1006), 8:65-10:2. Using the locations of the user’s ears and the perceived sound origin of the virtual object, “[t]he HRTF of the user is used to convert the sound into a localized sound that appears to the user as coming from the sound origin” and delivered to the user through headphones. Osman (Ex. 1006), 10:3-32, *see also* Osman (Ex. 1006), 9:57-10:32, Fig. 4. Specifically, Osman describes that “[a]s seen in FIG. 4, virtual objects can be located anywhere in space.” Osman (Ex. 1006), 9:49-65. Osman describes that such virtual objects can be located in empty space, i.e., where no physical, real objects exist, as shown in Figure 4:



Osman (Ex. 1006), Fig. 4.

151. Osman describes that for a virtual game cannon, “[a] sound effect simulates the fire of the cannon from the actual location of user 406 b” where “[t]he firing shot is heard first, and is followed by the sound of the cannonball as the cannonball travels through the air from the location of player 406 b to the location of player 406 a” and “[a]s the cannonball travels through the air, the sound increases in intensity, as in real life.” Osman (Ex. 1006), 10:54-11:3. “Finally, the explosion is heard when the cannonball hits the target, and if the target is near the user, the sound will be delivered with high volume.” Osman (Ex. 1006), 10:54-11:3.

152. For generating spatialized sound, Osman describes that “game console 422 tracks the location of the ears of the player 406a” and “[w]hen sounds are generated in the game (e.g., the sound of a helicopter flying) game console 422 determines the coordinates in the virtual space of the sound origin.” Osman (Ex. 1006), 10:3-16. Osman describes that the game console then “determines the relative location between the sound origin and the ears perceiving the sound” and uses the user’s HRTF “to convert the sound into a localized sound that appears to the user as coming from the sound origin.” Osman (Ex. 1006), 10:3-16. Osman explains that “[t]he acoustic signals for the localized sound, which are played by the earphones, provide an acoustic clue to the user regarding the virtual location in space of the virtual object.” Osman (Ex. 1006), 10:3-33. A POSA would understand that

the generating virtual images with spatialized sound created using HRTFs as described in Osman is “processing ... sounds with the HRTFs to generate binaural sound ... that externally localizes in empty space to the virtual image in the 3D virtual environment.”

153. In addition, as described in the Motivation to Combine, a POSA would have been motivated to combine and would have had a reasonable expectation of success in combining the teachings of Osman for processing sounds with the HRTFs to generate binaural sound that externally localizes in empty space to the virtual image in the 3D virtual environment with the virtual reality system of Clement to create a more realistic virtual experience.

e. Element 14[d] determining, with the WED, when the user is leaving the environment;

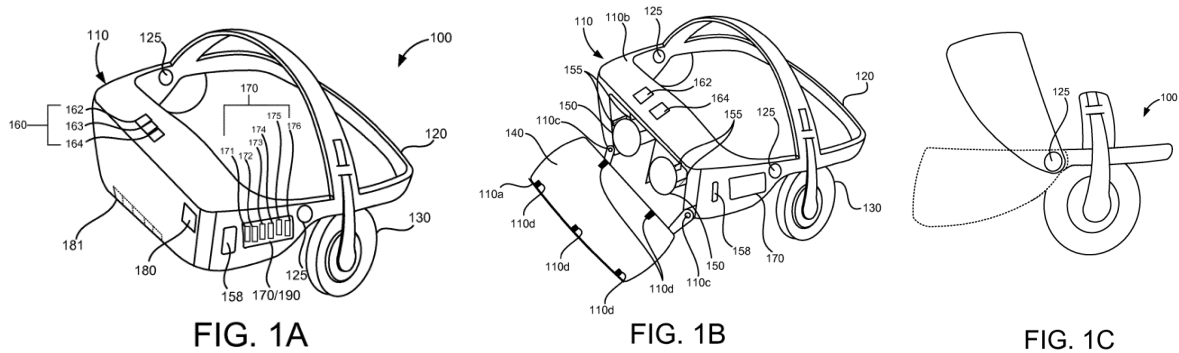
154. Clement teaches “determining, with the WED, when the user is leaving the environment.”

155. Clement teaches that “[i]n a VR system, a user may physically move in a prescribed physical space in which the system is received and operated” and that “[t]he system may track the user’s movement in the physical space, or the ‘real’ world, and cause the virtual world to ‘move’ in coordination with the user’s movement in the real world.” Clement (Ex. 1005), 4:6-14. A POSA would understand that the “tracking area” or “tracking zone” described in Clement is a 3D virtual environment. Clement (Ex. 1005), 4:15-48.

156. Clement describes that “[i]f the user A moves out of the tracking area 320, or for some other reason cannot be ‘seen’ by the tracking device 310 (such as, for example, being blocked by an obstacle), then continued movement of the user A in the real world space 3200 will not be translated into corresponding movement in the virtual world.” Clement (Ex. 1005), 4:15-48. If this happens, “the virtual world may remain still, or appear stuck, while the user A continues to move, which may cause disorientation, motion sickness, dissatisfaction with the virtual immersive experience and the like.” Clement (Ex. 1005), 4:49-5:16. Clement teaches that “[a] fluid, or graceful, transition from the virtual world to the real world, before actually reaching this still or stuck point in the virtual world, may alleviate the negative effects of disorientation, motion sickness and the like.” Clement (Ex. 1005), 4:61-5:16; *see also* Clement (Ex. 1005), 3:42-5:54, 6:7-56, Figs. 3A-3C.

157. Clement describes that “the HMD 100 may include a sensor that detects that the user user's proximity to a boundary of the tracking area 320, or that the user is within a threshold distance of a boundary of the tracking area 320.” Clement (Ex. 1005), 7:4-16; *see also* Clement (Ex. 1005), 3:20-41 (“The HMD 100 may also include a sensing system 160 including various sensing system devices 162-164 and a control system 170 including various control system devices 171-176 to facilitate manual user control and automated control of the HMD 100.”), 3:42-4:5 (block

diagram of a system 200 for transitioning from a virtual/augmented reality environment to an ambient environment), 4:15-7:3,



Clement (Ex. 1005), Figs. 1A-1C,

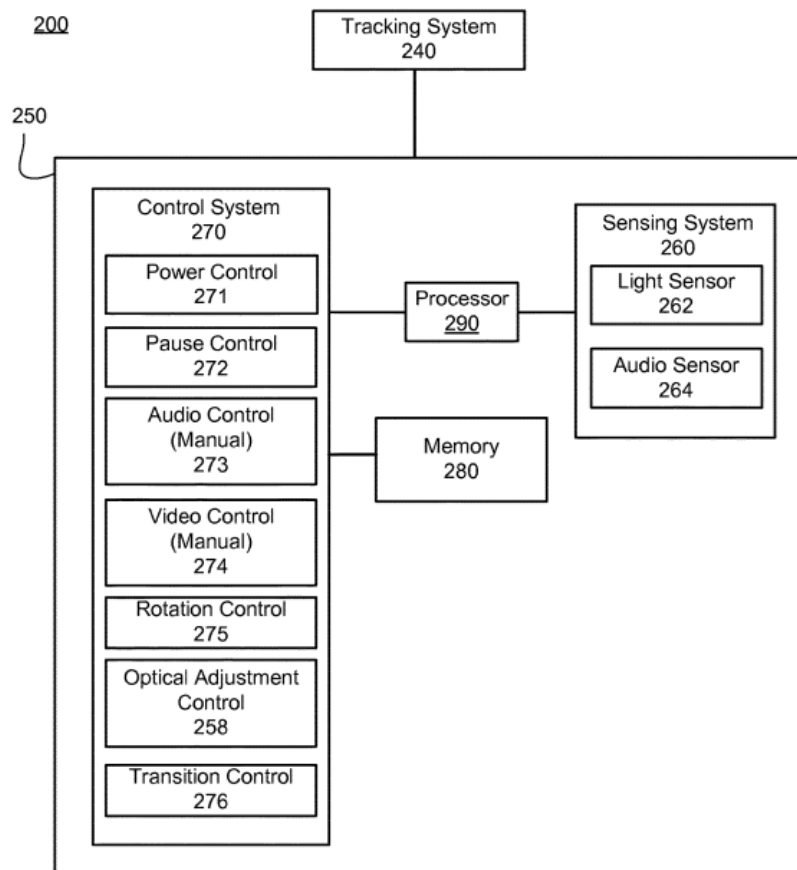
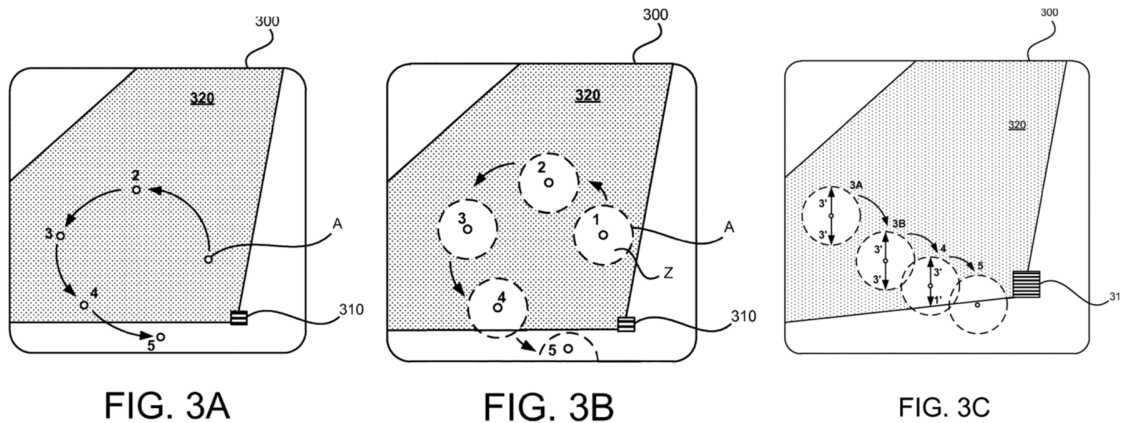


FIG. 2

Clement (Ex. 1005), Fig. 2.

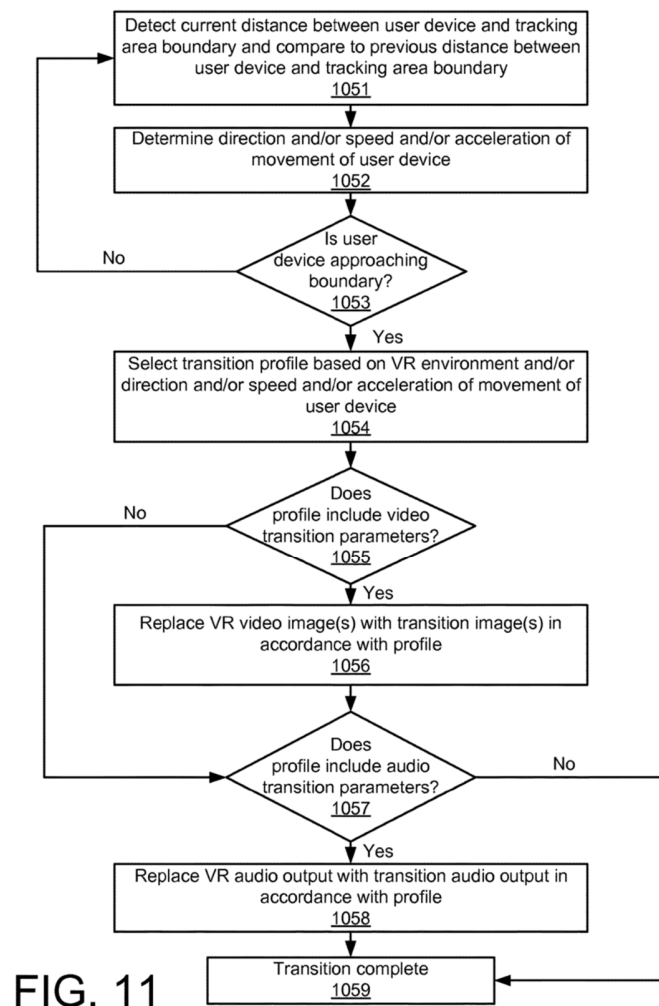
158. Clement teaches that “the tracking device 310 may not just track the physical position and movement of the user A in the space 300, but may also detect when the user A is within a predetermined distance, or threshold, of a boundary or edge of the tracking area 320” where “the system may issue a warning or indicator, and/or may automatically initiate a fluid or graceful transition out of the virtual world” when the user approaches the boundary (ex. at position 4 of Fig. 3B). Clement (Ex. 1005), 5:55-6:6. When the HMD detects that the user is nearing the boundary, a transition process is initiated. The HMD detects the current distance between the user and the boundary of the tracking area and compares that to a previously detected distance.



Clement (Ex. 1005), Figs. 3A-3C. For example, Clement describes that “[a]s the HMD 100 and/or the tracking device 310, or other component of the system, determines that the user is positioned within the preset threshold (in this particular example, 3 feet of the boundary of the tracking area 320, as at position 3B), the

system 100 may initiate a gradual transition out of the virtual world,” that “[a]t position 4, the user has moved closer to the boundary of the tracking area 320 (for example, within 1 foot of the boundary as shown in the example illustrated in FIG. 3C),” and that “[a]t position 4, the transition out of the virtual world may be at least partially complete, with the transition process fully complete by the time the user reaches position 5 and is fully outside of the tracking area 320.” Clement (Ex. 1005), 6:23-39.

159. Clement further discloses Fig. 4 which describes that “[i]f it is determined, at block 1040, that the user is within a set threshold or range of the boundary of the tracking area 320, at which loss of tracking by the tracking device 310 may be imminent, the transition process may be initiated at block 1050. Clement (Ex. 1005), 12:20-29,



Clement (Ex. 1005), Fig. 11. “As shown in FIG. 11, a current distance between the user and the boundary of the tracking area 320 may be detected, and may be compared to a previously detected distance between the user and the boundary, at block 1051.” Clement (Ex. 1005), 12:30-46. Clement describes that “[b]ased on this comparison, a direction and/or speed and/or acceleration of the movement of the user may be determined, at block 1052, to confirm the user is approaching the

boundary at block 1053” and “[i]f it is confirmed at block 1053 that the user is approaching/nearing the boundary (and is about to enter an area where the tracking device 310 may not be able to continue to track the user's physical movement in the space 300), a transition profile may be implemented, at block 1054, based on various factors, such as, for example, the virtual environment, the real world space 300 in which the virtual environment is implemented, direction and/or speed and/or acceleration associated with the movement of the user, and other such factors.” Clement (Ex. 1005), 12:30-46, *see also* Clement (Ex. 1005), 12:47-13:10.

160. The transition profile can include both video and audio transitions. Clement explains that “[i]f, at block 1055, it is determined that the transition profile includes video parameters, then, at block 1056, the virtual image(s) associated with the virtual world displayed on the display 140 of the HMD 100 may be replaced with transition image(s) displayed on the display 140 in accordance with the profile” and “may include, for example, a symmetric or asymmetric fade out of the virtual image(s) accompanied by a fade in of the transition image(s).” Clement (Ex. 1005), 12:47-65. In addition, Clement explains that “[i]f, at block 1057, it is determined that the transition profile includes audio parameters, then, at block 1058, the virtual audio output associated with the virtual world and output to the user via the audio output device 130 of the HMD 100 may be replaced with transition audio output in accordance with the profile” and “this replacement of the virtual audio output with

transition audio output may include, for example, a serial fade out of the virtual audio followed by a fade in of the transition audio output.” Clement (Ex. 1005), 12:47-65. In particular, “The transition audio output may include, for example, the actual sound generated in the real world space 300, a preset tone, no sound, and the like.” Clement (Ex. 1005), 12:66-13:10. Clement discloses that “[t]his transition may also include a transparent display 140 transitioning to a transparent state so that the real world is visible to the user through the transparent display, or a movement of a portion of the HMD 100 out of the field of view of the user so that the real world is directly visible to the user.” Clement (Ex. 1005), 12:47-65, *see also* Clement (Ex. 1005), 12:20-13:10, Fig. 11.

161. Moreover, tracking users in VR in the context of a boundary was well known. For example, Pedrotti (Ex. 1080) teaches that an HMD that includes sensors that are used to determine its position and orientation in space. Pedrotti (Ex. 1080), 4:49-5:24. For instance, Pedrotti (Ex. 1080) discloses that “[a]ttached to or incorporated in the HMD 106 is a tracker 108, which can be used to determine the position and orientation of the headset 104, and thus of the head of user 102, by, for example, sensing an electromagnetic field generated from a local base station, video game console or other apparatus (not shown).” Pedrotti (Ex. 1080), 4:49-5:24. Pedrotti (Ex. 1080) teaches that “[t]he tracker 108 provides information about the sensed magnetic field to a processor (also not shown), which derives the position

and orientation of the tracker 108, and thus HMD 106, relative to a source of the electromagnetic field from such information.” Pedrotti (Ex. 1080), 4:49-5:24. Pedrotti (Ex. 1080) explains that “[s]uch trackers, and the processors which determine the position and orientation of such trackers, are well known to those of skill in the art.” Pedrotti (Ex. 1080), 4:49-5:24. Pedrotti (Ex. 1080) teaches that the HMD may include “[o]ther forms of trackers, including those incorporating inertial sensors (e.g. accelerometers and/or gyroscopes)” and “essentially any type of tracker (e.g., magnetic, optical, electrical, sonar, radar, ultrasonic, or any combination thereof) which can provide information regarding its position (and, in some cases, orientation, if needed or desired) can be used in the present approach.” Pedrotti (Ex. 1080), 4:49-5:24. Regardless of the type of sensor used, the tracker provides position and/or orientation information to the processor. Pedrotti (Ex. 1080), 4:49-5:24.

f. Element 14[e] alerting the user that the user is leaving the environment by reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user

162. Clement and Osman with Sauk teaches or renders obvious “alerting the user that the user is leaving the environment by reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user.”

163. First, Clement and Osman with Sauk teaches or renders obvious “reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user.”

164. As discussed for Element 14[b], Clement with Osman discloses a 3D virtual environment that includes binaural audio that localizes from a virtual image. Further, as discussed for Element 14[d], Clement teaches that “[a]s the HMD 100 and/or the tracking device 310, or other component of the system, determines that the user is positioned within the preset threshold (in this particular example, 3 feet of the boundary of the tracking area 320, as at position 3B), the system 100 may initiate a gradual transition out of the virtual world.” Clement (Ex. 1005), 6:23-39.

165. Clement further describes that the transition profile can include both video and audio transitions. Clement explains that “[i]f, at block 1055, it is determined that the transition profile includes video parameters, then, at block 1056, the virtual image(s) associated with the virtual world displayed on the display 140 of the HMD 100 may be replaced with transition image(s) displayed on the display 140 in accordance with the profile” and “may include, for example, a symmetric or asymmetric fade out of the virtual image(s) accompanied by a fade in of the transition image(s).” Clement (Ex. 1005), 12:47-65. In addition, Clement explains that “[i]f, at block 1057, it is determined that the transition profile includes audio parameters, then, at block 1058, the virtual audio output associated with the virtual

world and output to the user via the audio output device 130 of the HMD 100 may be replaced with transition audio output in accordance with the profile” and “this replacement of the virtual audio output with transition audio output may include, for example, a serial fade out of the virtual audio followed by a fade in of the transition audio output.” Clement (Ex. 1005), 12:47-65. In particular, “[t]he transition audio output may include, for example, the actual sound generated in the real world space 300, a preset tone, no sound, and the like.” Clement (Ex. 1005), 12:66-13:10, Fig. 11.

166. A POSA would have understood that the transition of Clement is lowering of the volume of the binaural sound, and this would warn the user that the user is leaving the virtual environment. In particular, this transition would have reduced the user’s sense of immersion and presence in the environment and thus alert the user that the user is leaving the virtual environment.

167. In addition, a POSA would have also looked to Sauk, which likewise teaches that “[t]he decibel level of the 1-D, 2-D or 3-D audio can be adjusted to whatever the listener desires for that dimension.” Sauk (Ex. 1007), 11:40-41.

168. As described for Element 14[c], a POSA would understand that in the combination of Clement and Osman, the audio that fades is binaural sound and as such the combination describes “reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user.”

169. Second, the combination of Clement and Osman with Sauk teaches “alerting the user that the user is leaving the environment” by lowering the volume.

170. Clement teaches that “the tracking device 310 may not just track the physical position and movement of the user A in the space 300, but may also detect when the user A is within a predetermined distance, or threshold, of a boundary or edge of the tracking area 320” where “the system may issue a warning or indicator, and/or may automatically initiate a fluid or graceful transition out of the virtual world” when the user approaches the boundary (ex. at position 4 of Fig. 3B). Clement (Ex. 1005), 5:55-6:6. As described above, Clement teaches lowering the volume of the binaural sound as part of the transition out of the virtual world, which warns the user that the user is leaving the VR environment. Clement (Ex. 1005), 12:30-13:10.

4. '179 Patent - Dependent Claims 15-19

- a. Claim 15. The method of claim 14 further comprising:
alerting the user that the user is leaving the
environment by degrading the binaural sound being
provided to the user as the user is leaving the
environment.**

171. Clement and Osman with Sauk teaches “alerting the user that the user is leaving the environment by degrading the binaural sound being provided to the user as the user is leaving the environment.”

172. First, as discussed for Element 14[e], Clement teaches degrading the sound by reducing its volume.

173. Second, Sauk teaches that “less relevant audio” would have been “change[d]” from binaural to “a monophonic (1-D) or stereophonic (2-D) dimension.” Sauk (Ex. 1007), 11:26-41. Switching from binaural audio to monophonic or stereo would have been used to “signify a different level of relevancy or importance for such audio.” Sauk (Ex. 1007), 11:26-41; Sauk 2007 (Ex. 1061), 2. Sauk explains that this process “removes any localization cues for that [binaural] audio.” Sauk (Ex. 1007), 11:26-41; *see also* Sauk (Ex. 1007), 7:1-8:33, Fig. 5. When localization cues are removed, the binaural sound is degraded.

174. A POSA understood that the localization cues for binaural audio were removed by altering the filters used in the filtering process of binaural audio, and/or by altering the binaurally processed audio to reduce or eliminate the perception of spatial sound by removing the interaural (ITD and ILD) cues. Ex. 1038, 41; Sauk (Ex. 1007) 7:1-8:33, 11:26-41, Fig. 5; Sauk 2007 (Ex. 1061), 2-3.

175. As discussed in the Motivation to Combine, a POSA would have known that other types of alerts may be more effective in some situations to address the relevancy or immediacy of the transition. Clement teaches warning the user when the user is leaving and transitioning out of the environment because the user is heading towards a real wall in the real world, as opposed to a less important

transition when the user is not headed towards an obstacle or other danger. Clement (Ex. 1005), 7:28-8:35. One way to inform the user of the immediacy or relevance of the transition is to provide an additional audible indication, such as degrading the “binaural [audio] to monophonic or stereophonic audio [to] signif[y] a different level of relevancy.” Sauk (Ex. 1007) at 11:25-41; Sauk 2007 (Ex. 1061), 2. Further, as described in the Motivation to Combine, a POSA would have implemented the techniques of Sauk to more effectively communicate warnings to the user in order to ensure user safety.

176. Finally, a POSA would have expected to succeed in modifying the combination such that the binaural sound provided is switched to stereo or monophonic sound. As discussed above, these types of changes in sound processing were known in the art. Sauk (Ex. 1007), 11:26-41; Sauk 2007 (Ex. 1061), 2. A POSA would have made this modification.

b. Claim 16. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by altering, by the DSP, processing of the binaural sound so the user perceives less external localization.

177. Clement and Osman with Sauk teaches “alerting the user that the user is leaving the environment by altering, by the DSP, processing of the binaural sound so the user perceives less external localization” for the same reasons as Claim 15.

178. Specifically, Sauk describes switching from binaural audio to monophonic or stereo would have been used to “signify a different level of relevancy or importance for such audio.” Sauk (Ex. 1007), 11:26-41; Sauk 2007 (Ex. 1061), 2. Sauk explains that this process “removes any localization cues for that [binaural] audio.” Sauk (Ex. 1007), 11:26-41; *see also* Sauk (Ex. 1007), 7:1-8:33, Fig. 5.

179. A POSA understood that the localization cues for binaural audio were removed by altering the filters used in the filtering process of binaural audio, and/or by altering the binaurally processed audio to reduce or eliminate the perception of spatial sound by removing the interaural (ITD and ILD) cues. Ex. 1038, 41; Sauk (Ex. 1007) 7:1-8:33, 11:26-41, Fig. 5.

180. Specifically, as discussed, binaural sound includes ITD, ILD and HRTFs. Begault (Ex. 1038), 40-41, 49-50. This results in unique signals arriving at the left and right ears. In contrast, stereophonic sound relies on the inter-channel (not interaural) differences of intensity and time. These typically are a result of microphones being placed in various configurations around a sound source, or a room, and mixed down to two channels. Lipshitz (Ex. 1079) (“In stereo sound reproduction using two loudspeakers, each ear hears both loudspeakers, and is intended to do so.”). Stereo panning relies on inter-channel intensity differences to move a sound source left/right. This amplitude panning principle and stereophonic panning was originally proposed by Blumlein in 1931. Blumlein (Ex. 1077).

Monophonic sound is sound involving a single transmission path with no differences between the sound received by the two ears.

181. Because binaural and stereo sound differ through the inclusion of ITDs in binaural sound (and not in stereo sound), one way a POSA would switch sound from binaural to stereo is through removing or degrading ITDs in the binaural sound. And because binaural and mono sound differ through of the inclusion of ITDs and ILDs in binaural sound (and not in mono sound), and one way a POSA would switch sound from binaural to mono is through removing or degrading ITDs and ILDs, along with other localization cues in the binaural sound. Thus, a POSA would have understood that switching sound from binaural to stereophonic (or monophonic) sound, as described in Sauk, such that the user perceives less external localization, would be performed by removing or degrading audio cues (including ITDs, ILDs and HRFTs) in the binaural sound. Sauk (Ex. 1007), 11:26-41 (“This process also removes any localization cues for that audio.”)

182. For example, methods for removing audio cues were well known in the art. For example, U.S. Pat. App. Pub. No. 2011/0206209 (Ojala, Ex. 1010) describes that “spatial audio techniques have been used in connection with multi-channel audio reproduction” and that well-known Binaural Cue Coding (BCC) methods can be used to “encode the multi-channel audio signal by down mixing the various input audio signals into either a single (“sum”) channel or a smaller number of channels

conveying the “sum” signal.” Ojala (Ex. 1010), [0005]-[0006]. In particular, Ojala (Ex. 1010) describes methods for removing the “phase difference” between channels of multi-channel audio to create a single channel of audio. Ojala (Ex. 1010), [0023]-[0024], claim 1. A POSA would understand that the phase difference between channels represents the time delay between the channels. And in the context of spatial audio, a POSA would further understand that the specific time delay between channels is known to be the ITD, and producing single channel sound is producing monophonic sound. Further a POSA would understand that in generating monophonic sound from binaural sound, other localization cues would be removed or degraded as well. Sauk (Ex. 1007), 11:26-41 (“This process also removes any localization cues for that audio.”)

183. Korean Patent No. KR20110022252A (Ex. 1035, Moon) similarly describes removing phase differences in multichannel sound and that doing so will generate a “mono audio signal.” Moon (Ex. 1035), claim 1. Thus a POSA would have known how to remove ITDs by removing phase differences from the multichannel sound such as binaural sound, as described in Ojala (Ex. 1010) and Moon (Ex. 1035), as well as other localization cues when generating monophonic sound from binaural sound. And as described in the Motivation to Combine section, a POSA would have been motivated to do so to more effectively communicate warnings to the user in order to ensure user safety.

- c. **Claim 17. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.**

184. Clement and Osman with Sauk teaches “alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user” for the same reasons as Claims 15 and 16.

185. Switching sound from binaural to monophonic sound would be performed by the removal of time delays, such as ITDs. *See* Ojala (Ex. 1010), [0006], [0023]-[0024], claim 1; (discussing the removal of time delays from multichannel (spatialized) audio to produce a single channel (monophonic) audio); Moon (Ex. 1035), claim 1 (same); Begault (Ex. 1038), 40-41, 49-50. A POSA would have understood that switching sound from binaural to monophonic sound, as described in Sauk, includes removing ITDs. Begault (Ex. 1038), 41; Sauk (Ex. 1007) 7:1-8:33, 11:26-41, Fig. 5.

- d. **Claim 18. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by eliminating audio cues in the binaural sound being provided to the user.**

186. Clement and Osman with Sauk teaches “alerting the user that the user is leaving the environment by eliminating audio cues in the binaural sound being provided to the user.” for the same reasons as Claims 15 to 17.

187. As described for Claims 15 to 17, a POSA understood that switching from binaural to stereo sound eliminates audio cues in the binaural sound provided to the user. Sauk (Ex. 1007) at 11:35-40 (“This process also removes any localization cues for that audio.”).

- e. **Claim 19. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to stereo sound.**

188. Clement and Osman with Sauk teaches “alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to stereo sound” for the same reasons as Claim 15.

5. '179 Patent - Independent Claim 7

189. Independent Claim 7 is generally similar to Claim 14, except that it alerts the user in a different manner.

a. Elements 7[pre] to 7[d]

190. Counsel instructed me to assume this preamble was limiting. Elements 7[pre] to 7[d] are taught by Clement with Osman and Sauk for the same reasons as Elements 14[pre] to 14[d].

191. A POSA would understand that there is no relevant distinction between displaying “with” the WED as recited in Element 14[b] and displaying “by” the WED as recited Element 7[b]. And a POSA would understand that the localization

“to the empty space and to the virtual image” of Element 7[c] is taught for the same reasons as localization “in empty space from the virtual image” of Element 14[c].

- b. Element 7[e] alerting the user that the user is leaving the environment by altering processing of the binaural sound to degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound.**

192. Clement with Osman and Sauk teaches or renders obvious “alerting the user that the user is leaving the environment by altering processing of the binaural sound to degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound” for the same reasons as Claims 15 to 19.

193. As described for Claims 15 to 19, altering the processing of the binaural sound by degrading it by converting it to stereo or mono prevents the user from experiencing external localization of the binaural sound.

6. '179 Patent - Dependent Claims 8-10

- a. Claim 8. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by reducing a volume of the binaural sound being provided to the user.**

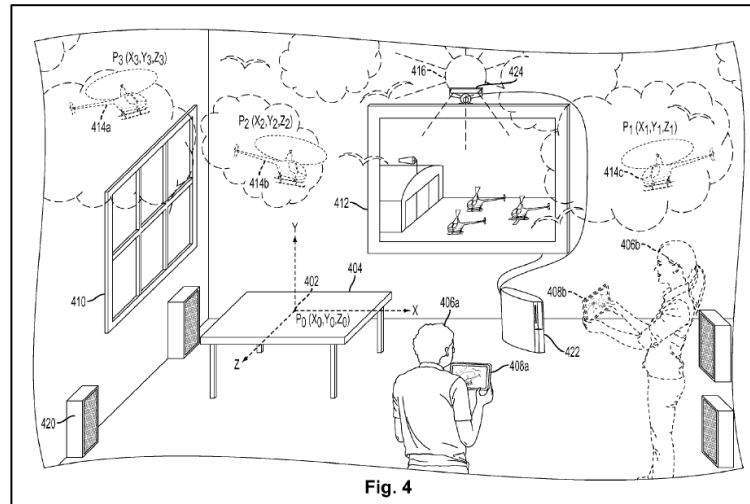
194. Clement with Osman and Sauk teaches “alerting the user that the user is leaving the environment by reducing a volume of the binaural sound being provided to the user” for the same reasons as Element 14[e] and Claim 15.

- b. Claim 9. The method of claim 7 further comprising: displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character**

from the binaural sound to stereo sound when the character vanishes from the environment during the game.

195. Clement with Osman and Sauk teaches or renders obvious “displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character from the binaural sound to stereo sound when the character vanishes from the environment during the game.”

196. Clement with Osman teaches, or renders obvious, “displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image.” Clement teaches that its system involves “multiple players or users ... in the same space 300, engaged in a virtual reality immersive experience.” Clement (Ex. 1005), 8:10-35. A POSA would have understood that referencing “multiple players” describes a virtual game and that virtual environments were commonly used for games. Likewise, Osman teaches a 3D virtual environment as part of a game the user is playing. *See* Osman (Ex. 1006), 6:6-7:3 (describing a “shared-space game” in “multi-player augmented-reality”), 7:43-49 (describing “augmented reality games” with “virtual game characters”), 8:65-9:11 (describing an “augmented-reality environment” or “virtual scene” for players playing a game),



Osman (Ex. 1006), Fig. 4. A POSA would have understood that in the combination of Clement and Osman that a user would have used the system to play games that included virtual game characters, as taught by Osman. A POSA further understood that virtual games, which include game characters, were one of the main uses of 3D VR systems at the time. *See, e.g.*, VR Book (Ex. 1067) at 12 (“VR is a great fit for the entertainment industry and will certainly be the driving force behind VR in the short term.”), 257 (describing “Characters, Avatars, and Social Networking” in VR); Pedrotti (Ex. 1080), 1:14-2:13 (describing a “virtual world”); 11:43-65 (describing “playing a game within a virtual world”); 9:26-40 (describing control of “a character on a television screen” in the game).

197. Clement with Osman teaches, or renders obvious “switching a voice of the character from the binaural sound to stereo sound when the character vanishes from the environment during the game.” A POSA would have understood that sound processing would have been changed when the user does not see a virtual character,

such as when the user is not facing the virtual character, and that sound emanating from a virtual character has lower priority or relevance when the user's face is directed away from the character as compared with sounds the user is facing. *See, e.g.,* Osman (Ex. 1006), 10:33-43 (“In one embodiment, the acoustic signals for the localized sounds are delivered with a higher volume when the object or person originating the sound is shown on the display of the portable device, or in the display 412 connected to game console 422. . . . When the sound origin is not on the display, the volume of the sound is lower.”); Jang (Ex. 1015), 8:25-9:3, Fig. 2 (describing the focus area as the area “within the visual field” and the non-focus area as areas outside the visual field), 6:39-7:2 (describing processing “the focus area differently than . . . the non-focus areas”), *see also* Jang (Ex. 1015), 2:32-59 (“Sound output of the focus area may be processed differently than sound output of the un-focus area.”), 3:39-51 (“The sensory processing unit may output sound sources corresponding to the non-focus area at a volume that is less than sound sources corresponding to the focus area.”), 4:20-24 (“The focus area may correspond to an area in the virtual environment in the immediate foreground of a visual perspective of a character corresponding to a user, and the non-focus area may correspond to an area that is not in the immediate foreground of the visual perspective of the character.”), 7:38-63.

198. As described for Claim 15, Sauk teaches that lower priority sounds may be switched from binaural to stereophonic sound. Sauk (Ex. 1007), 11:26-41. A POSA would have been motivated to combine switching from binaural to stereophonic sound as described in Sauk with changing sound processing when the face of the user is directed away from the virtual character because doing so would conserve processing resources, indicate a change in relevance, and direct the user's attention to sounds that the user is facing.

199. For example, a POSA would understand that in the case of a large number of sounds in the environment, there would be computational savings if only the specific sound sources within the view of the user (and not those outside the view the display or that the user is not facing) were processed to provide binaural sound. Nair (Ex. 1033), [0194]-[0224].

200. Further, a POSA would have been motivated to switch binaural sound to stereo to ensure that the sound from different sources remained intelligible to users as they leave the environment. Switching to monophonic would also have been obvious, but may be less preferable than switching to stereo in some situations. For example, switching to a monophonic presentation of multiple sound sources would result in a collapse of the audio to the center of the listener's head, resulting in a sound cacophony—the so-called “cocktail party effect” which POSAs often would have avoided when the user is or may be in a virtual environment with many different

sound sources (e.g., a cocktail party with many different conversations). *See* Cherry (Ex. 1043), 975-977; *see also* Growney (Ex. 1049) (describing the cocktail party effect).

c. Claim 10. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.

201. Clement with Osman and Sauk teaches “alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound” for the same reasons as Element 7[e].

7. '179 Patent - Independent Claim 1

202. Independent Claim 1 is generally similar to Claim 14, except that it alerts the user in a different manner.

a. Elements 1[pre] to 1[d]

203. Counsel instructed me to assume this preamble was limiting. Elements 1[pre] to 1[d] are taught by Clement with Osman and Sauk for the same reasons as Elements 14[pre] to 14[d].

204. A POSA would understand that the HMD described in Element 1[b] is same as the wearable electronic device (WED) of Element 14[b]. A POSA would understand that the HMD described for Element 14[d] includes sensors that perform the claimed sensing of Element 1[d] and memory that performs the storing of Element 1[b].

- b. Element 1[e] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter audio cues to prevent the user from experiencing external localization of the sounds and by lowering a volume of the sounds being provided to the user.**

205. Clement with Osman and Sauk teaches “warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter audio cues to prevent the user from experiencing external localization of the sounds and by lowering a volume of the sounds being provided to the user” for the same reasons as Elements 14[d]-[e] and Claims 15 to 18.

8. '179 Patent - Dependent Claims 3-4, and 6

- a. Claim 3. The method of claim 1, wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment.**

206. Clement and Osman teaches “wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment” for the same reasons as Element 1[e].

207. Further, a POSA understood that transitions between virtual rooms were well-known in virtual reality environments. For example, in the Background section describing the prior art, Jang (Ex. 1015) explains that “in a virtual environment, a user can shift between virtual spaces without having to portray the passage of distance or time to travel from one space to another.” Jang (Ex. 1015), 1:57-2:12. Jang (Ex. 1015) further provides “[a]n example of such a virtual reality

system includes a virtual reality game, “Second Life,” developed by Linden Lab of the United States.” Jang (Ex. 1015), 1:57-2:12. Jang (Ex. 1015) further explains that “[i]n such a virtual reality system, when a character of a user shifts between virtual spaces, the image viewed by the user and the sound heard by the user should change according to the shift between spaces or the change of the spaces.” Jang (Ex. 1015), 1:57-2:12. As illustrated by Jang (Ex. 1015), a POSA would have been familiar with virtual reality games with virtual rooms, such as Second Life, as well as the need to provide the user with transitions as they switch between virtual rooms. Jang (Ex. 1015), 1:57-2:12.

208. Jang (Ex. 1015) further describes that when shifting between these virtual rooms, “a volume of sound in a first space is gradually decreased while a volume of sound in a second space is gradually increased.” Jang (Ex. 1015), Abstract, 2:15-5:21. A POSA would understand that such a transition signals to the user that the user is leaving one virtual room and entering another virtual room, while the user remains in the 3D virtual environment.

209. A POSA would further understand that the transition between virtual rooms would also have been implemented in the combination of Clement and Osman. As described in Jang (Ex. 1015), such transitions were well-understood. Further, because Clement already teaches audio transitions as warnings when approaching the boundary, a POSA would have had a reasonable expectation of

success in implementing similar audio and visual transitions between virtual rooms. Such an implementation would be able to use the elements already present in the combination, including the existing position tracking systems and binaural audio processing systems. Thus, a POSA would have found it obvious to implement the audio transitions known in the prior art to also warn the user of switching between virtual rooms while the user remains in the virtual environment.

b. Claim 4. The method of claim 1 further comprising: tracking, with head tracking in the HMD, head movements of the user; and switching the sounds from the binaural sound to stereo sound when a face of the user is directed away from the virtual image.

210. Clement with Osman and Sauk, teaches or renders obvious “tracking, with head tracking in the HMD, head movements of the user; and switching the sounds from the binaural sound to stereo sound when a face of the user is directed away from the virtual image.”

211. First, Clement with Osman teaches, or renders obvious, “tracking, with head tracking in the HMD, head movements of the user.” In addition, Clement teaches a device that tracks “physical movement in real world space.” Clement (Ex. 1005), Abstract, 4:6-14, 7:62-8:11, 12:27-55, 16:27-35, Figs. 1A-1C, 2, 3A-3C. In particular, Clement teaches “determining the location of the players face, or ears,” which “is done utilizing image analysis of video images of play area 360, including the detection and tracking of user features such as eyes 366 and 368, face, nose 372,

mouth 370, torso 364, etc.” Clement (Ex. 1005), 7:62-8:11. Clement also teaches that “facial features are tracked to estimate the three dimensional location of the head and to derive the positions of the ears” and “[t]he more features are tracked, the more reliable head tracking is.” Clement (Ex. 1005), 7:62-8:11; *see also* Clement (Ex. 1005), 12:36-55 (“Head Tracking module 738 may determine the position of the user’s ears using one or more different technologies such as image recognition, RFID, ultrasound, infrared, triangulation, etc.”), 16:27-35 (“In another embodiment, the head of the user is tracked, which means that the location of the ears is known during the test. In one embodiment, the head tracking is performed by analyzing images taken of the user, but other methods may also be utilized, such as using headphones with a magnetometer, etc.”). A POSA would have understood that this tracking can be used to localize sound based on the position and location of the user. In addition, a POSA implementing would have incorporated Osman’s accelerometer, gyroscope, magnetometer, GPS, and compass into Clement’s HMD (to the extent it is not already included) to locate the position of the user’s HMD and to localize sound using HRTFs. Osman (Ex. 1006), 8:56-64 (“The position of the user is continuously tracked, and sound localization is based on the current location of the user. For example, if the user turns the head, sound localization changes so the sound appears to come from the correct location, even as the user is turning the head.”), 10:54-12:14 (describing GPS, RFID, and triangulation for head and position

tracking), 19:48-67 (“Information from different devices can be used by the Position Module to calculate the position of the portable device. These modules include a magnetometer, an accelerometer, a gyroscope, a GPS, and a compass. Additionally, the Position Module can analyze sound or image data captured with the cameras and the microphone to calculate the position. Further yet, the Position Module can perform tests to determine the position of the portable device or the position of other devices in the vicinity, such as WiFi ping test or ultrasound tests.”).

212. In addition, head tracking to implement 3D/binaural sound using HRTFs was known and well within the skill of a POSA. Begault (Ex. 1038), 8, 132-136 (“If 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 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.”), 167-172 (“In 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. For this reason, 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 spatial hearing.”).

213. Next, Clement with Osman teaches, or renders obvious, “switching the sounds from the binaural sound to stereo sound when a face of the user is directed away from the virtual image.” As discussed for Claim 9, a POSA would have understood that sound processing would have been changed when the user is not facing the virtual image, and that sound emanating from a virtual image has lower priority or relevance when the user’s face is directed away from the image as compared with sounds the user is facing. *See, e.g.*, Osman (Ex. 1006), 10:33-43; Jang (Ex. 1015), 8:25-9:3, Fig. 2, 6:39-7:2, *see also* Jang (Ex. 1015), 2:32-59, 3:39-51, 4:20-24, 7:38-63. And Sauk teaches switching from binaural to stereophonic sound to signify relevance changes. Sauk (Ex. 1007), 11:26-41.

214. A POSA would have been motivated to combine switching from binaural to stereophonic sound as described in Sauk with changing sound processing when the face of the user is directed away from the virtual object because doing so would conserve processing resources, indicate a change in relevance, and direct the user’s attention to sounds that the user is facing.

**c. Claim 6. The method of claim 1 further comprising:
switching the sounds from the binaural sound to stereo
sound when the virtual image vanishes in the 3D
virtual environment.**

215. Clement with Osman and Sauk, teaches or renders obvious “switching the sounds from the binaural sound to stereo sound when the virtual image vanishes in the 3D virtual environment” for the same reasons as Claims 4 and 9.

216. In particular, a POSA understood that when a virtual image vanishes from the 3D virtual environment, such as when the virtual object is no longer within the visual field of the user, the virtual object is of lower priority. *See, e.g.*, Osman (Ex. 1006), 10:33-43, Jang (Ex. 1015), 8:25-9:3, 2:32-59, 3:39-51, 4:20-24, 6:39-7:2, 7:38-63, Fig. 2. And POSAs would have understood that lower priority sounds may be switched to stereo and that POSAs would have been motivated to do so.

B. '179 Patent - Ground 2: All Claims Are Unpatentable Over Lyren and Reif

217. I explain below, all claims of the '179 Patent are either disclosed in or obvious over the combination of Lyren and Reif, as well as why a POSA would have considered it obvious to implement Lyren and Reif.

218. I note that, as I stated in the Priority Date section at the beginning of the section, counsel has asked me to assume that the '179 Patent's actual filing date (May 2, 2019) is the date of the alleged invention—and therefore the date for my obviousness analyses (e.g., the understanding of a POSA and what would have been obvious to that POSA).

219. I further note that for Ground 2 of the '184 Patent, which I address starting at Section X.B below, I will instead use the '184 Patent's actual filing date (June 16, 2019). This slight difference in date (May 2, 2019 compared to June 16, 2019) does not change my analysis.

1. The '179 Patent's claims lack written description support in its parent, the as-filed '140 Application

220. I have reviewed Application Ser. No. 16/106,140 which is in Exhibit 1028 at pages 92-191 (“'140 Application”), which is the immediate parent of the '179 Patent.

221. Counsel has asked me to evaluate whether the '140 Application provides sufficient written description support for the subject matter claimed in the '179 Patent.

222. As I explained in the Legal Understanding section near the beginning of this declaration, a claim has sufficient written description support when the disclosure (including the as-filed claims) would have reasonably conveyed to a POSA that the inventor was in possession of the claimed subject matter as of the filing date.

223. *The '140 Application*—including the claims that were filed with the application (*see* '140 Application (Ex. 1028) at 92-191, with the claims at pages 80-176, and Abstract and figures after that)—***does not provide sufficient support*** to show the requisite possession of the claimed subject matter.

224. In particular, all of the '179 Patent's claims directly (or indirectly via their independent claims) recite the HMD/WED sensing or determining when the user is leaving the 3D environment, and alerting or warning that user of that fact by

reducing the sound's volume and/or otherwise altering it. The patent does not mention these features at all.

225. The '140 Application does describe “a wide variety of different types of events” that may trigger “changing sound from binaural sound to stereo sound, mono sound, or altered binaural sound.” '140 Application (Ex. 1028) at 22-24, Figs. 4-16. However, the '140 application does not describe anywhere (including the claims) the triggering event of a user leaving the environment. In other words, the '140 application does not give an example of using the user's location to determine that (and to warn or alert) the user that the user is leaving the virtual environment, as recited in the claims.

226. For example, the '140 Application describes determining when the user moves into a “restricted area” where providing binaural sound is dangerous or prohibited. '140 Application (Ex. 1028) at 23 (“Examples of events and things that can trigger events include ... position or movement or orientation of a person or head of the person (such as direction the person walks or head orientation of the person), distance from the person to an object (such as distance from the person to a wall or an obstruction), available space (such as how much physical 3D space is available to receive and/or localize a sound or voice), safety (such as not localizing sound when the person is driving a vehicle), proximity to or being at a restricted area (such

as state, local, or United States Federal regulations prohibiting externally localizing sound while in a certain building or on an airplane)’’).

227. However, none of these is the claimed determining when and alerting or “warning the user that the user is leaving the 3D virtual environment.”

228. The ’140 Application also describes detecting or determining when people or things other than the user enter or leave the environment. *E.g.*, ’140 Application (Ex. 1028) at 107 (“a person walking into a room”), 141 (“visually rendered character vanishing from an augmented or virtual environment”), 144 (“any SLP that travels ‘off stage’ by leaving the room”). But there is no description of detecting when or altering the sound to warn or alert the user that *the user* is leaving the environment.

229. To be clear, my analysis includes the claims filed with the application. The claims recite, for example, sensing with an HPED that “an object interferes with the SLP” (’140 Application (Ex. 1028) at 171), Claim 4 recites determining with an HPED that two SLPs overlap (’140 Application (Ex. 1028) at 172), Claim 8 recites “sensing, by the computer system ... an object at the SLP” (’140 Application (Ex. 1028) at 173), etc. But none of the claims recite the determining/sensing and alerting/warning recited in the ’179 Patent’s independent claims.

2. Combination of Lyren and Reif

230. I have been asked to assume that the '140 Application, filed on August 21, 2018, is prior art for the '179 Patent.

231. Lyren teaches nearly each and every claim element of the '179 Patent, but, as described above for the '140 Application and in my Lyren overview, Lyren fails to teach or suggest the that its HMD/WED senses or determines when the user is leaving the 3D environment, and fails to teach or suggest alerting or warning that user that the user is leaving the environment by altering the binaural sound.

232. A POSA would have combined the teachings of Lyren and Reif to arrive at the claimed invention of the '179 Patent for the reasons discussed below at paragraphs 247-263. Specifically, while, as I mentioned above, Lyren does not teach alerting or warning users that the user is leaving the environment by altering binaural sound, it invites those of skill in the art to use its teachings in other situations or in response to other events. *See e.g.*, Lyren, [0065]. And Reif teaches detecting when a user is leaving the environment and warning/alerting the user that the user is leaving the environment and explains that these warnings/alerts are critical to ensure the safety of the user. Reif, Fig. 2B, 2A, Abstract, [0003]-[0005]. As I described above in the context of Pedrotti and Ground 1, these types of safe areas were well-known and used in the field—those of ordinary skill would have added these boundaries to designate a safe area in Lyren's VR environment and would have

further implemented alerts/warnings to inform the user when the user is leaving the safe area or environment. As I noted, I explain my reasoning further at paragraphs 247-263 below.

3. '179 Patent - Independent Claim 14

a. Element 14[pre] A method executed by one or more electronic devices, the method comprising:

233. Counsel instructed me to assume this preamble was limiting. It is my opinion that Lyren with Reif discloses the recited “a method executed by one or more electronic devices.”

234. Lyren discloses that “a user 150 wears an optical head mounted display (OHMD) 152 that simultaneously provides a plurality of sound localization points 154, 155, 156, and 157 during a conference call.” Lyren (Ex. 1008), [0053], *see also* Lyren (Ex. 1008), [0077] (“Consider an example in which Alice wears electronic glasses with a see-thru display, OHMD, or a head-mounted display.”), [0251]-[0253] (describing a computer system including a head mounted display), [0258]-[0260] (describing the sound location system of the computer system), [0281]-[0282] (describing a computer system including a wearable electronic device), Figs. 17, 18, 20.

235. Likewise, Reif discloses a “virtual reality (VR) system” that is implemented on computer hardware. Reif (Ex. 1009), Abstract, 12:65-13:18.

b. Element 14[a] storing head-related transfer functions (HRTFs) in memory;

236. Lyren teaches “storing head-related transfer functions (HRTFs) in memory.”

237. Specifically, Lyren describes that “a SLS [sound localization system] renders a source sound using an HRTF” and that “[m]emory 1726 and/or dedicated memory in the SLS 1728 stores ... HRTFs.” Lyren (Ex. 1008), [0046], [0256]; *see also* Lyren (Ex. 1008), [0283] (HRTFs stored in the electronic device), [0287] (wearable electronic device stores HRTFs).

238. As described for this limitation in Ground 1, a POSA would understand that the memory would have been in the WED because that is where the processing takes place to reduce latency, facilitate signal processing, and enhance the user experience.

c. Element 14[b] displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

239. Lyren teaches “displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment.”

240. Specifically, Lyren discloses that “A SLP in empty space can include images or video (e.g., images that are part of an augmented or virtual reality).” Lyren (Ex. 1008), [0077]. Lyren discloses “an example in which Alice wears electronic glasses with a see-thru display, OHMD, or a head-mounted display” such that

“[d]uring a call with Bob, the system localizes a voice of Bob to a SLP in empty space that is proximate to but away from Alice.” Lyren (Ex. 1008), [0077]. In this case, “[t]he electronic glasses or head-mounted display provides or displays an image of Bob that coincides with the SLP in empty space” where “[t]he image appears to exist in space at the location in empty space with the SLP that is proximate to but away from Alice.” Lyren (Ex. 1008), [0077]; *see also* Lyren (Ex. 1008), [0085] (“The SLP can exist in empty or unoccupied space, such as appearing in front of the person, next to the person, above the person, below the person, etc. This empty space can include virtual images or images per an augmented reality, such as 2D or 3D images that appear through electronic glasses. Alternatively, the SLP can exist in non-empty or occupied space, such as appearing to emanate from a physical object or tangible thing. For example, sound localizes to a moving remote control car or a teddy bear sitting on a chair.”), [0163]-[0164] (“Consider an example in which Alice and Bob are both in a 3D visually rendered space talking binaurally face-to-face in full-duplex with each person in a medial plane of the other. Because they can see each other's visual representation, they experience accurate sound localization of each other's voices.”), [0251]-[0252] (“The computer system 1700 includes ... optical head mounted display (OHMD) or smart glasses or wearable electronic glasses 1750 (including one or more sensors 1752, a processor 1753, a memory 1754, speakers 1755, sound localization system 1756, a display 1757, and

microphones 1758)”), [0257] (“Consider an example in which WEG 1750 localizes binaural sound to a location that is proximate to but away from a wearer of the WEG. ... The display 1757 displays an image on or over the SLP so the wearer sees the position of this SLP.”), [0258]-[0259] (“FIG. 18 is a portion of a computer system 1800 that includes a sound localization system 1810, sound hardware 1820, a codec selector 1830, codecs 1840, SLP sound sources 1850, input data 1860, a network and/or other electronic devices 1870, and a file system 1880.”), [0281]-[0283] (“FIG. 20 is a computer system 2000 that includes an electronic device 2002, a server 2004, a server 2006, a wearable electronic device 2008, storage 2010 with user profiles 2012, and an electronic device 2014 with one or more sensors 2016 in communication with each other over one or more networks 2018.”), [0287]-[0288] (“Wearable electronic device 2008 includes computer readable medium (CRM) or memory 2070, one or more displays 2072, a processor or processing unit 2074, one or more interfaces 2076 (such as an interface discussed herein), a camera 2078, one or more sensors 2080 (such as a sensor discussed herein), a motion or location tracker 2082, one or more speakers 2084, HRTFs 2086, a head tracking system or head tracker 2088, an imagery system 2090, a sound localization system 2092, and one or more microphones 2094.”), Figs. 17, 18, 20.

241. Further, Reif also teaches displaying a 3D virtual environment in an HMD that includes virtual images in the environment. *E.g.*, Reif (Ex. 1009),

Abstract, 1:9-40 (“Embodiments of the disclosure can include a virtual reality (VR) system that enables a user wearing a head mounted display (HMD) to safely move around while still experiencing a virtual world or environment displayed via the HMD to the user.”), 4:7-40 (“In some cases, the safe region can correspond to a portion inside the set of one or more boundaries within which the user wearing the HMD 150 is presented with a virtual scene of a virtual environment.”).

242. For example, a well-known VR system for displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment is Petitioners’ Oculus Quest system. An example of the Oculus HMD in use was provided during the Oculus Connect 5 presentation by Petitioners. *See* Oculus Connect 5 Keynote Day 1 (Ex. 1087) at 1:02:18-1:03:8 (video clip of Oculus Quest in use); 26:33-28:07 (video clip of Oculus Rift in use). Specifically, as shown in the video, the Oculus Quest and Oculus Rift HMDs display virtual images, such as tennis players and other game characters, to the user when they are playing games in a virtual environment. Oculus Connect 5 Keynote Day 1 (Ex. 1087), 1:02:18-1:03:8 (Oculus Quest video); 26:33-28:07 (Oculus Rift video).

- d. Element 14[c] processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes in empty space from the virtual image in the environment;**

243. Lyren teaches “processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes in empty space from the virtual image in the environment.”

244. Lyren teaches that its SLSs (sound localization systems) use HRTFs to generate binaural sound. Lyren (Ex. 1008), [0046] (“Another way of changing a sound from being perceived as binaurally captured audio or binaurally manufactured audio to being perceived as non-binaural audio is to render the original source with different spatial parameters or to re-render a sound source with different spatial parameters in order to adjust, degrade, or eliminate certain human audio cues. For example, a SLS renders a source sound using an HRTF to adjust SLPs, renders the sound source to specific alterations per an ITD and/or interaural level difference (ILD), or discontinues rendering or using the HRTF or ITD/ILD calculations while continuing to render other aspects of the audio without pause.”), [0077] (“A SLP in empty space can include images or video (e.g., images that are part of an augmented or virtual reality). Consider an example in which Alice wears electronic glasses with a see-thru display, OHMD, or a head-mounted display. During a call with Bob, the system localizes a voice of Bob to a SLP in empty space that is proximate to but away from Alice. The electronic glasses or head-mounted display provides or displays an image of Bob that coincides with the SLP in empty space. The image appears to exist in space at the location in empty space with the SLP that is proximate

to but away from Alice. Thus, the SLP of Bob's voice and the image of Bob exist in empty space at the same location that is proximate to but away from Alice. To Alice, the voice of Bob appears to emanate from the image of Bob.”), [0085] (“The SLP can exist in empty or unoccupied space, such as appearing in front of the person, next to the person, above the person, below the person, etc. This empty space can include virtual images or images per an augmented reality, such as 2D or 3D images that appear through electronic glasses. Alternatively, the SLP can exist in non-empty or occupied space, such as appearing to emanate from a physical object or tangible thing.”), [0163]-[0164] (“Consider an example in which Alice and Bob are both in a 3D visually rendered space talking binaurally face-to-face in full-duplex with each person in a medial plane of the other. Because they can see each other's visual representation, they experience accurate sound localization of each other's voices.”).

245. In addition, Lyren also teaches that its WEDs worn on the user's head include one or more DSPs to process the sound as part of the sound localization system (SLS). Lyren (Ex. 1008), [0251]-[0252] (“The sound localization system performs or executes one or more functions or methods discussed herein (such as one or more blocks discussed in FIG. 2-16). By way of example, the sound localization system executes or assists in executing one or more of ... switching among binaural and stereo and mono sounds, localizing sound (such as localizing sound to a SLP that is away from but proximate to a user), ... managing binaural

sound and binaural sound localization, rendering and altering binaural sound, ... and other functions relating to binaural sound.”), [0256] (“Sound localization system 1728 includes ... a digital-to-audio converter (DAC), decompressor/decoder, digital signal processor (DSP), and includes hardware and/or software to process and localize sound to a SLP.”), [0257] (“Consider an example in which WEG 1750 localizes binaural sound to a location that is proximate to but away from a wearer of the WEG. Sensors 1752 include a specific or customized sensor with a MEMS-based inertial measurement unit (IMU). ... The IMU communicates head-tracking data to the sound localization system 1756 to provide a static SLP that localizes near the wearer of the WEG. The display 1757 displays an image on or over the SLP so the wearer sees the position of this SLP.”). [0258]-[0259] (“FIG. 18 is a portion of a computer system 1800 that includes a sound localization system 1810, sound hardware 1820, a codec selector 1830, codecs 1840, SLP sound sources 1850, input data 1860, a network and/or other electronic devices 1870, and a file system 1880.”), [0281]-[0283] (“FIG. 20 is a computer system 2000 that includes an electronic device 2002, a server 2004, a server 2006, a wearable electronic device 2008, storage 2010 with user profiles 2012, and an electronic device 2014 with one or more sensors 2016 in communication with each other over one or more networks 2018.”), [0287]-[0288] (“Wearable electronic device 2008 includes computer readable medium (CRM) or memory 2070, one or more displays 2072, a processor or processing unit

2074, one or more interfaces 2076 (such as an interface discussed herein), a camera 2078, one or more sensors 2080 (such as a sensor discussed herein), a motion or location tracker 2082, one or more speakers 2084, HRTFs 2086, a head tracking system or head tracker 2088, an imagery system 2090, a sound localization system 2092, and one or more microphones 2094.”), Figs. 17, 18, 20.

246. In addition, a POSA would have found it obvious to process the HRTFs using a DSP. As discussed for Ground 1, Element 14[c], using DSPs for processing HRTFs well-known, common, and routine. And a POSA would have understood the advantages of using a DSP for processing of spatial sound, including efficient processing of digital signals to maximize power efficiency.

- e. **Elements 14[d]-14[e] determining, with the WED, when the user is leaving the environment; and alerting the user that the user is leaving the environment by reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user**

247. Lyren does not teach these limitations.

248. Lyren teaches that certain events trigger altering binaural sound. Lyren, [0064] (“Example embodiments can switch in response to or based on a wide variety of different types of events.”). For example Lyren teaches:

Example embodiments are not limited to a specific type of an event or a specific time or duration of an event. As noted, such events can be dynamic or static and selected by one or more of a user, a person, apparatus or machine, method, etc. Examples of events and things that can trigger events include, but are not limited to, one or more of a time

of day, a calendar day (such as a specific day of the week or day in a month), a location (such as a location of an electronic device or of a person listening to the sound), actions of a third person (such as a person walking into a room), a command or request from a person (such as a person interacting with a user interface to switch the sound), a command or request from a machine (such as a process, software program, intelligent user agent, or intelligent personal assistant commanding, requesting, initiating, or executing the switch), processing power (such as available processing power of an electronic device during a voice exchange or sound localization), bandwidth (such as available transmission and receiving wireless bandwidth of an electronic device during a voice exchange or sound localization), memory (such as available memory of an electronic device during a voice exchange or sound localization), position or movement or orientation of a person or head of the person (such as direction the person walks or head orientation of the person), distance from the person to an object (such as distance from the person to a wall or an obstruction), available space (such as how much physical 3D space is available to receive and/or localize a sound or voice), safety (such as not localizing sound when the person is driving a vehicle), proximity to or being at a restricted area (such as state, local, or United States Federal regulations prohibiting externally localizing sound while in a certain building or on an airplane), time (such as to switch the sound after a predetermined or given amount of time), a person's identity in a communication (such as to switch a call from binaural sound to stereo when a certain person calls using voice-over internet protocol, VoIP), and other examples provided herein.

Lyren (Ex. 1008), [0065]; *see also* Lyren (Ex. 1008), [0063]-[0096] (providing additional examples); [0097] (“Figures 4-16 provide examples of events for changing sound from binaural sound to stereo sound, mono sound, or altered binaural sound” and “other types of actions”). However, Lyren does not teach leaving the environment as a potential trigger event.

249. Lyren further teaches that altering binaural sound can involve switching binaural sound to stereo or mono sound, or lowering the volume. Lyren (Ex. 1008), [0025] (“Example embodiments include systems, apparatus, and methods that change binaural sound in response to an event. When the event occurs, binaural sound changes, such as switching to stereo sound, switching mono sound, switching to altered binaural sound, removing or changing a sound localization point (SLP), moving a SLP (such as moving the SLP from being externally localized to being internally localized), or taking another action in accordance with an example embodiment.”). For example, Lyren teaches the following:

The sound changes from being provided in binaural sound to being provided in stereo sound, mono sound, or altered binaural sound. Alternatively, the sound changes from being provided in stereo sound or mono sound to being provided in binaural sound. Sound can switch back and forth from being provided in binaural, stereo, and mono sounds (including switching between different variations of binaural sound, such as binaural sounds having different SLPs, different volumes at SLPs, etc.).

Lyren (Ex. 1008), [0067];

The event triggers or initiates a switch from binaural sound to stereo sound, from stereo sound to binaural sound, from binaural sound to mono sound, from mono sound to binaural sound, or from binaural sound to altered binaural sound. For example, a person receives binaural sound with a first codec, and a switch occurs such that the person receives binaural sound with a second codec. As another example, a person receives binaural sound rendered with a first set of HRTFs, and a switch occurs such that the person receives a second binaural sound rendered from a second set of HRTFs. As another example, a person receives binaural sound with a first set of SLPs, and a switch occurs such that the person receives binaural sound with a

second set of SLPs. As yet another example, a person receives binaural sound with a first set of background sound, and a switch occurs such that the person receives binaural sound with a second set of background sound. A person can receive binaural sound corresponding to one virtual or real or augmented space, and a switch occurs such that the person receives binaural sound from a second virtual or real or augmented space. As yet another example, after an event is detected, a change to the original binaural sound occurs while still providing the listener with altered or changed binaural sound (such as changing one or more SLPs, ITDs, ILDs, HRTFs, etc. in the original binaural sound while still maintaining binaural sound).

Lyren (Ex. 1008), [0089];

In FIGS. 4-16, example actions include, but are not limited to, one or more of switch the sound from binaural sound to stereo sound, switch the sound from binaural sound to mono sound, switch the sound from stereo sound to binaural sound, switch the sound from mono sound to binaural sound, maintain binaural sound but alter the binaural sound, stop binaural sound, discontinue playing sound, mute the sound, lower a volume of the sound, raise a volume of the sound, “cancel-out” or quiet a sound or part of a sound by processing it with Active Noise Control (ANC), provide a sound or audio alert, provide a visual alert, move one or more SLPs, adjust or alter a SLP, cancel a SLP, replace a SLP with a different SLP, replace a binaural environment with a different binaural environment, switch one or more codecs, cancel a command, execute a command or instruction, alter a HRTF of a person, change or alter an ITD or an ILD, end a computer program or process, start a computer program or process, provide a notification to a computer program or a person, and other actions discussed herein.

Lyren (Ex. 1008), [0103]; *see also* Lyren (Ex. 1008), [0063]-[0096] (providing additional examples), [0097] (“Figures 4-16 provide examples of events for changing sound from binaural sound to stereo sound, mono sound, or altered binaural sound” and “other types of actions”); [0046] (“Another way of changing a sound from being perceived as binaurally captured audio or binaurally manufactured

audio to being perceived as non-binaural audio is to render the original source with different spatial parameters or to re-render a sound source with different spatial parameters in order to adjust, degrade, or eliminate certain human audio cues. For example, a SLS renders a source sound using an HRTF to adjust SLPs, renders the sound source to specific alterations per an ITD and/or interaural level difference (ILD), or discontinues rendering or using the HRTF or ITD/ILD calculations while continuing to render other aspects of the audio without pause.”).

250. For example, Lyren teaches triggering an alteration of the sound provided to the user that localizes from the image (e.g., reducing volume and/or shifting from binaural to stereo or mono) based on events such as other people or characters entering or leaving a room or the environment. Lyren (Ex. 1008), [0065] (“Examples of events and things that can trigger events include, but are not limited to, one or more of ... actions of a third person (such as a person walking into a room) ...”), [0225] (“Events can be triggered by a change in a user’s ability to associate visual cues or images with the associated audio cues, such as a visually rendered character vanishing from an augmented or virtual environment, the presence or absence of a physical object ...”), [0232] (“As another example, if there are several people/SLPs in a physical or virtual space, any SLP that travels “off stage” by leaving the room can still be included in the conversation, but the sound can switch to mono or stereo.”). Lyren also teaches altering the sound based on determining “a

location (such as the location of an electronic device or of a person listening to the sound).” Lyren (Ex. 1008), [0065]; *see also* Lyren (Ex. 1008), [0123] (“Examples of restricted areas include, but are not limited to, an area, a location, or a point that prohibits SLPs or sound localization, an area in which it is dangerous to localize sound, a vehicle, or other location. Examples of such locations include at or near a construction zone or other inherently dangerous or hazardous area, inside an automobile, on a motorcycle or other motorized vehicle, in a library or a hospital or a sports arena or an elevator or a school or classroom, on a public transport (such as a bus, train, or airplane).”), [0240] (“The intelligent user agent surmises that externally localizing Alice's voice to Bob now might be dangerous for Bob since he is on a steep incline. The HPED receives the call, and the intelligent user agent adjusts the call so Alice's voice internally localizes to Bob through his earphones. In spite of the settings to externally localize Alice's voice, the intelligent user agent made a determination to trump or override the settings and have her voice internally localize to Bob. This decision was made as being in the best interest of Bob's safety.”).

251. Although, Lyren does not teach leaving the environment as a specific event that would warrant alerting or warning the user, Lyren discloses user safety is an important concern and that unsafe things should trigger such warnings. Lyren (Ex. 1008), [0065] (“Examples of events and things that can trigger events include,

but are not limited to, one or more of ... safety (such as not localizing sound when the person is driving a vehicle)"). Lyren also states that its techniques "are not limited to a specific type of an event or a specific time or duration of an event," and that "events can be dynamic or static and selected by one or more of a user, a person, apparatus or machine, method, etc." Lyren (Ex. 1008), [0065].

252. Reif, however, further teaches that detecting when the user was leaving the environment and warning the user of that event was crucial for user safety. A POSA implementing Lyren would have included Reif's additional safety features to keep users safe.

253. In particular, Reif teaches identifying a "safe region," where the user may move without bumping into real world objects and potentially suffering a real-world injury. Reif (Ex. 1009), Abstract ("A virtual reality (VR) system identifies a bounding box including a safe region. The bounding box encloses an area that allows a user wearing a head mounted display (HMD) to safely view content of a virtual world. The VR system identifies the user's location relative to the bounding box and tracks the user's location. When it is determined that the user is located outside the safe region of the bounding box, the VR system can modify what is presented to the user via the HMD, can warn the user, and/or can suggest that the user move back to the safe region."), 1:9-25 ("However, in some conventional VR systems, while a user wearing an HMD is moving around, the user may move "outside" the virtual

environment being experienced, which degrades the user experience associated with the virtual environment. In some cases, under conventional approaches, the user may undesirably bump, hit, crash, or otherwise make physical contact with a real-world object and hurt himself/herself, which also degrades the user experience associated with the virtual environment.”), Fig. 2B. *see also* Reif (Ex. 1009), Fig. 2A; [0002]-[0005] (“Embodiments of the disclosure can include a virtual reality (VR) system that enables a user wearing a head mounted display (HMD) to safely move around while still experiencing a virtual world or environment displayed via the HMD to the user. ... In some implementations, the VR system can define or identify a set of boundaries forming a bounding box within which a user wearing an HMD can safely move. ... The user is capable of safely moving inside the bounding box to view the virtual world displayed to the user. ... In one embodiment, the bounding box includes a safe region and a warning region.”), [0016]-[0023] (“The bounding box can include or encompass a set of one or more boundaries that define a safe region.”).

254. Reif teaches that “when it is determined that the user is located outside the safe region of the bounding box, the VR system can modify what is presented to the user via the HMD, can warn the user, and/or can suggest that the user move back to the safe region.” Reif (Ex. 1009), Abstract, [0002]-[0005] (“As the distance between the user and the center of the bounding box increases (e.g., when the user moves past the safe region and reaches the warning region), the view of the virtual

world can “transition” to a view of the real world, to provide a warning that the user is approaching the boundary of the region in which the user may safely move.”), [0016]-[0019]. A POSA would have implemented this feature to help protect the users from injury by bumping, hitting, tripping over, etc., over real-world objects when moving around while viewing the virtual environment in the HMD. Reif (Ex. 1009), [0002] (“In some cases, under conventional approaches, the user may undesirably bump, hit, crash, or otherwise make physical contact with a real-world object and hurt himself/herself, which also degrades the user experience associated with the virtual environment.”), [0004] (“Accordingly, this can reduce the likelihood that the user will bump, hit, crash, or otherwise undesirably make physical contact with a real-world object while wearing the HMD and moving.”), [0037] (“Accordingly, only the real-world scene is displayed to the user when the user is located outside the warning region. This can reduce the likelihood of the user undesirably bumping, hitting, or crashing, etc., into one or more objects in the real-world environment.”); U.S. Patent Application Publication No. 2019/0043259 (Wang, Ex. 1083) at [0001] (“However, such devices are often power inefficient due to continuously performing depth sensing or require extensive calibration to designate certain areas of a room to be safe for navigation without colliding into objects.”), [0032] (“With this information, the electronic device 100 can display notifications or other visual indications to the user 110 while navigating through a

VR environment that enables the user 110 to avoid collision with objects in the local environment 112, such as by staying within the designated virtual bounded floor plan.”), [0059] (“Accordingly, display of the VR environment is paused (i.e., no longer rendered for display in any form, faded or not, on the display of device 100) and a warning 802 (e.g., “Go Back to Play Area” message) is displayed to encourage the user 110 to return to the virtual bounded floor plan 800. This restricts navigation of VR environments to areas which have already been verified to be clear of collision risks.”).

255. Reif teaches that: “In some cases, the indication provided by the notification module 350 can include at least one of a visual indication, an audio indication, or a haptic indication. For example, the visual indication can correspond to a text message or a visual animation, the audio indication can include a warning beep or a voice message, and the haptic indication can include a vibration.” Reif (Ex. 1009), [0039]. Reif suggests a warning beep or voice message, but informs the POSA that “it should be appreciated that many variations are possible.” Reif (Ex. 1009), [0039].

256. In addition, Reif also teaches warning the user by gradually fading out the virtual embodiment as the user transitions from inside the safe region, through the warning zone, and to an area outside the bounding box, where the virtual environment is no longer visible. Reif (Ex. 1009), [0016]-[0018] (“Continuing with

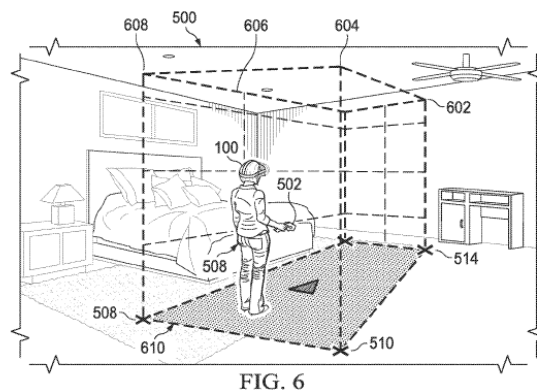
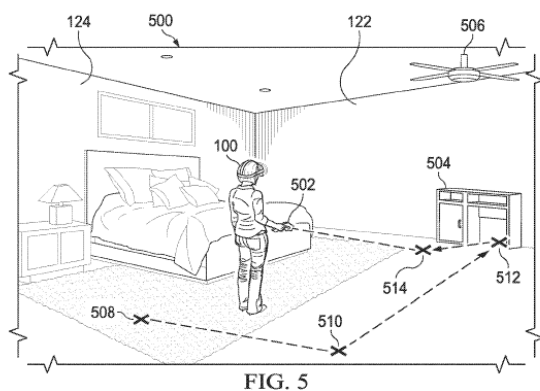
this example, when the user is located in the warning region, and as the distance between the user and the center of the bounding box increases, the opacity of the virtual scene displayed to the user can be decreased by the VR system 300 from 100% to 0% and the opacity of the real-world scene displayed to the user can be increased by the VR system 300 from 0% to 100%.”), [0036] (“As this distance between the user and the outer boundary(ies) of the warning region decreases, the real-world scene can be displayed to the user at a higher opacity level (i.e., at a lower transparency level) and the virtual-scene can be displayed at a lower opacity level (i.e., at a higher transparency level).”); *see also id.* [0002] (“However, in some conventional VR systems, while a user wearing an HMD is moving around, the user may move “outside” the virtual environment being experienced, which degrades the user experience associated with the virtual environment.”), [0004]-[0005] (“As the distance between the user and the center of the bounding box increases (e.g., when the user moves past the safe region and reaches the warning region), the view of the virtual world can “transition” to a view of the real world, to provide a warning that the user is approaching the boundary of the region in which the user may safely move. Thus, the opacity of the rendered virtual world can be reduced and the opacity of the real world, which is presentable via the HMD based on image data captured by a camera of the HMD, can be increased.”), [0019]-[0028] (further examples), Figs. 2A-2D (graphical illustration of fade out).

257. Reif further teaches sensors in the user's HMD. Specifically Reif teaches a "location module" as part of the "VR system" for determining the user's location in the real world and thus the user's location relative to the "bounding box." Reif (Ex. 1009), [0033]. Reif teaches that the location module takes input from input components (sensors) such as a camera, accelerometers, and gyroscopes to determine the user position. Reif (Ex. 1009), [0033]. Reif further teaches that the HMD may include both the VR system and the input component. Reif (Ex. 1009), [0015] ("In some embodiments, the HMD 150 can be included as a part of the VR system 300. ... Additionally, in some implementations, the HMD 150 can include the input component 120 (e.g., a camera facing the same direction as the user) configured to capture a view(s) or scene(s) of a real-world environment that the user would perceive if he/she was not wearing the HMD 150.").

258. As discussed for the Technology Background and for Ground 1, Element 14[d], such safe regions and establishing boundaries for safe regions were a well-known safety feature, along with warning the user that the user is leaving the safe region. Pedrotti (Ex. 1080), 4:49-5:24 ("The tracker 108 provides information about the sensed magnetic field to a processor (also not shown), which derives the position and orientation of the tracker 108, and thus HMD 106, relative to a source of the electromagnetic field from such information. Such trackers, and the processors which determine the position and orientation of such trackers, are well known to

those of skill in the art.”), 11:66-12:50 (“Thus, it is desirable to provide a means by which the user may be made aware that the user is approaching a boundary of a selected area in which the user desires to remain, to avoid physical objects or for any other reason, without distracting the user from the immersive virtual world. In one embodiment, a selected area having an outer boundary is determined, such as a “safe” area in which the user can move freely without coming into contact with the physical objects in the user's vicinity.”), 13:34-45 (“if the user's head goes beyond the boundary, the processor can cause the HMD to fade to black, and present, for example, only a warning message and/or a beacon or arrows directing the user to move in the direction of the boundary and the selected area.”), 15:25-65 (“Next, at step 1508, the processor determines from the position of the handheld controller that the handheld controller is within a preselected distance of the boundary. Finally, at step 1510 the processor generates and outputs instructions to the HMD to display on the HMD a visual warning to the user. ... By alerting a user in this fashion that the user is approaching the boundary of the selected area, the described system and method allows a user to take action to avoid the boundary and stay within the selected area defined by that boundary. This can assist the user in avoiding collision with physical objects in the area of the user, and any possible resulting damage to such objects or injury to the user.”), 16:10-32 (“For example, it is expected that the described apparatus may be implemented in numerous ways. ... One of skill in the

art will appreciate how to determine which and how many processors will be appropriate for a specific intended application. Similarly, as explained above, various tracking techniques may be used in addition to electromagnetic tracking, and in some embodiments some of the other components used for tracking, such as an electromagnetic field source or a camera, may be located in the base station, while in other embodiments they may be located elsewhere.”), Fig. 15;



Wang (Ex. 1083) at Figs. 5-6, [0060] (“Accordingly, display of the VR environment is paused (i.e., no longer rendered for display in any form, faded or not, on the display of device 100) and a warning 802 (e.g., “Go Back to Play Area” message) is displayed to encourage the user 110 to return to the virtual bounded floor plan 800. This restricts navigation of VR environments to areas which have already been verified to be clear of collision risks.”), Fig. 9; U.S. Patent Application Publication No. 2015/0094142 (Stafford, Ex. 1082) at [0046] (“For example, the triggering device 270 can initiate a display of a displayable image on the display 250 or the fading of brightness of images displayable by the head mounted display to the user.

In one example embodiment, the triggering device 270 stops displaying images by the head mounted device.”), [0057] (“At operation 540, the user is warned that the user approaches the edge of the scene based on the determination performed at operation 530. The warning can include an audio message, a visible message, turning on light emitting diodes, gradual fading of images displayed on the head mounted display, blending displayed images in the real world view, vibrating the head mounted display, disengaging the head mounted display or parts thereof, moving the head mounted display away from the eyes of the user, and the like.”).

259. Further, as discussed for Ground 1, Element 14[d], placing sensors in the HMD to determine the user’s position was also well-known in the art. *See, e.g.*, Pedrotti (Ex. 1080), 4:49-5:24 (“Attached to or incorporated in the HMD 106 is a tracker 108, which can be used to determine the position and orientation of the headset 104, and thus of the head of user 102, by, for example, sensing an electromagnetic field generated from a local base station, video game console or other apparatus (not shown).”); Wang (Ex. 1083), [0056] (“To this end, the electronic device 100 [HMD] may use location estimations of spatial features using non-image sensor data representative of global pose information, such as sensor data captured from a GPS receiver, a magnetometer, gyrocompass, and the like.”).

260. As discussed above in this subsection, Lyren discloses many variations—including altering the binaural sound provided in the virtual environment by reducing its volume and/or switching it to stereo or mono.

261. Lyren expressly teaches when the HMD detects that the “event” of the user leaving the environment triggers switching to stereo to mono or lowering the sound’s volume. *See, e.g.*, Lyren (Ex. 1008), [0103] (“In FIGS. 4-16, example actions include, but are not limited to, one or more of switch the sound from binaural sound to stereo sound, switch the sound from binaural sound to mono sound, [and] lower a volume of the sound”).

262. A POSA would have known that Lyren’s techniques of lowering the volume or otherwise altering/degrading the environment’s binaural would have been particularly helpful in virtual games/environments and would have implemented them. This was confirmed by Oculus engineers at the Oculus Connect 5 conference in September 2018: lowering the volume of binaural audio serves as a “soft notification” that was not “jarring” to user, but still warns the user that they are leaving the play space:

Also, there is a Guardian built in, but some experiences are going to have a smaller play space. Bogo had a small play space, so we needed some way to warn the player that you were leaving this play space, so there’s loud parts and soft parts in the experience. We were thinking, well, if we put sounds on it, if we put a loud sound on, let’s say during the fireworks part, that’s sound is going to be way too loud during the soft parts. If we put a soft sound that sits over here, you’re not going to hear it in the fireworks part. So we’re just like, hmm, you know what

we could do, we'll just try the absence of sound. We did a low-pass filter that ramps nice and smooth, so it's not jarring to the player, because you still want to keep everybody immersed, just let them know that they're, you know, about to step out of the play area.

Oculus Connect 5: Spatial Audio for Oculus Quest and Beyond (Ex. 1085), 13:00 to 14:40. Such an audio fade is demonstrated by the Oculus engineers during Oculus Connect 5, in which a video is shown where both the images and sounds fade when the user is walks out of the play area. Oculus Connect 5: Spatial Audio for Oculus Quest and Beyond (Ex. 1085), 13:00 to 14:40; *see also* Clement (Ex. 1005), 9:20-31, 12:66-13:10 (similarly fading out audio); Sauk (Ex. 1007) at 11:40-41 (similarly reducing the “decibel level” of the sound).

263. Thus, a POSA would have used Lyren's well-known technique of helping ensure user safety by warning the user that the user is leaving the environment (as taught by Reif) by determining or sensing that the user is leaving the environment and, in response to that determination, altering the environment's binaural sound by reducing its volume.

4. '179 Patent - Dependent Claims 15-20

- a. Claim 15. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by degrading the binaural sound being provided to the user as the user is leaving the environment.**

264. Lyren and Reif teaches “alerting the user that the user is leaving the environment by degrading the binaural sound being provided to the user as the user is leaving the environment.” *See* Elements 14[d]-[e] above.

265. In particular, Lyren teaches degrading the binaural sound by lowering its volume and/or switching to stereo or mono upon detecting a triggering event. *See* Elements 14[d]-[e] above; Lyren (Ex. 1008), [0103] (“In FIGS. 4-16, example actions include, but are not limited to, one or more of switch the sound from binaural sound to stereo sound, switch the sound from binaural sound to mono sound, [and] lower a volume of the sound”).

266. Lyren teaches: “An example embodiment uses a DSP or other processor to filter the binaural sound and degrade, alter, or eliminate sufficient audio cues to prevent a listener from experiencing external localization from a binaural audio source. For example, after DSP processing, the user perceives the sound with less, little, or no external localization. For example, a DSP process removes or re-normalizes interaural time differences (ITDs) in source impulses to cause imprecise or zero azimuth angle perception.” Lyren (Ex. 1008), [0044]-[0046] (“Another way

of changing a sound from being perceived as binaurally captured audio or binaurally manufactured audio to being perceived as non-binaural audio is to render the original source with different spatial parameters or to re-render a sound source with different spatial parameters in order to adjust, degrade, or eliminate certain human audio cues. For example, a SLS renders a source sound using an HRTF to adjust SLPs, renders the sound source to specific alterations per an ITD and/or interaural level difference (ILD), or discontinues rendering or using the HRTF or ITD/ILD calculations while continuing to render other aspects of the audio without pause.”).

267. As discussed in the Motivation to Combine for Ground 1, a POSA would have known that these other alerts may be more effective and needed in some situations to address the relevancy or immediacy of the transition, and others less so. Specifically, Clement describes providing warning of hazards such as obstacles and Sauk’s describes various methods of doing so, including changing binaural sound to stereo or monophonic sound.

268. And as discussed for Claim 15 in Ground 1, a POSA would have understood users leaving the boundary of a safe play area required an immediate warning to prevent injury. For instance, a POSA would have understood that leaving the safe play area (the boundary) would have risked the user colliding with physical objects etc. A POSA would have degraded the audio, as described in Lyren to communicate the urgency of the warning. *See* Sauk (Ex. 1007), 11:25-41 (“The

effect of changing an audio format for sounds from binaural to monophonic or stereophonic audio signifies a different level of relevancy or importance for such audio.”). A POSA would have implemented that technique here to provide various levels of transitions/warnings to ensure user safety—which Lyren expressly suggests is a valid concern. Lyren (Ex. 1008), [0065] (“Examples of events and things that can trigger events include, but are not limited to, one or more of ... safety (such as not localizing sound when the person is driving a vehicle)”).

- b. Claim 16. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by altering, by the DSP, processing of the binaural sound so the user perceives less external localization.**

269. Lyren and Reif teaches “alerting the user that the user is leaving the environment by altering, by the DSP, processing of the binaural sound so the user perceives less external localization” for the same reasons as Claim 15. *See also* Lyren (Ex. 1008), [0043]-[0046] (“For example, after DSP processing, the user perceives the sound with less, little, or no external localization.”).

- c. Claim 17. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.**

270. Lyren and Reif teaches “alerting the user that the user is leaving the environment by altering, by the DSP, processing of the binaural sound so the user

perceives less external localization” for the same reasons as Claims 15 and 16. *See also* Lyren (Ex. 1008), [0045]-[0046] (“For example, a DSP process removes or re-normalizes interaural time differences (ITDs) in source impulses to cause imprecise or zero azimuth angle perception.”), [103] (“In FIGS. 4-16, example actions include, but are not limited to, one or more of switch the sound from binaural sound to stereo sound, switch the sound from binaural sound to mono sound, ... alter a HRTF of a person, [and] change or alter an ITD or an ILD,”).

**d. Claim 18. The method of claim 14 further comprising:
alerting the user that the user is leaving the
environment by eliminating audio cues in the binaural
sound being provided to the user.**

271. Lyren and Reif teaches “alerting the user that the user is leaving the environment by eliminating audio cues in the binaural sound being provided to the user.” for the same reasons as Claims 15-17. *See also* Lyren (Ex. 1008), [0045] (“An example embodiment uses a DSP or other processor to filter the binaural sound and degrade, alter, or eliminate sufficient audio cues to prevent a listener from experiencing external localization from a binaural audio source. For example, after DSP processing, the user perceives the sound with less, little, or no external localization. For example, a DSP process removes or re-normalizes interaural time differences (ITDs) in source impulses to cause imprecise or zero azimuth angle perception.”)

- e. **Claim 19. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to stereo sound.**

272. Lyren and Reif teaches “alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to stereo sound” for the same reasons as Claim 15.

- f. **Claim 20. The method of claim 14 further comprising: detecting, with a microphone in the WED, a voice of a person in the environment of the user; and switching the binaural sound to stereo sound in response to detecting the voice.**

273. Lyren teaches or renders obvious “detecting, with a microphone in the WED, a voice of a person in the environment of the user; and switching the binaural sound to stereo sound in response to detecting the voice.”

274. Lyren teaches switching binaural sound in response to detecting a wide variety of events, such as “actions of a third person (such as a person walking into the room” or when the WED “detects a presence of a sound or an image.” Lyren (Ex. 1008), [0065] (“Example embodiments are not limited to a specific type of an event or a specific time or duration of an event. ... Examples of events and things that can trigger events include, but are not limited to, one or more of ... actions of a third person (such as a person walking into a room), a command or request from a person (such as a person interacting with a user interface to switch the sound)”), [0104]

(“For example, an event occurs when an electronic device detects a presence of a sound or an image.”).

275. As described in Lyren, a triggering event can be if a person in a user’s environment physically walks up to the user and a microphone on the HMD detects the person’s speech. *E.g.*, Lyren (Ex. 1008), [0107] (“Later, Bob (a friend of Alice) walks near her and says ‘Hello.’ The voice recognizer recognizes Bob’s voice, and the system automatically mutes the SLPs since Bob is on a list as one of Alice’s friends.”).

276. One of the actions that Lyren may take upon detecting such an event is switching the binaural sound to stereo. Lyren (Ex. 1008), Abstract, [0023] (“When an event occurs, the binaural sound switches or changes to stereo sound, to mono sound, or to altered binaural sound.”), [0025] (“When the event occurs, binaural sound changes, such as switching to stereo sound, switching mono sound, switching to altered binaural sound, removing or changing a sound localization point (SLP), moving a SLP (such as moving the SLP from being externally localized to being internally localized), or taking another action in accordance with an example embodiment.”), [0066] (“Block 220 states change the sound from binaural sound to stereo sound, mono sound, or altered binaural sound.”), [0105] (“For example, when Charlie comes within three feet of Alice, the voice of Bob that Alice hears switches from binaural sound to stereo or mono sound.”). Thus, Lyren therefore teaches

switching the binaural sound to stereo sound in response to detecting the person's voice, as recited in the claim.

277. A POSA would find this obvious. Lyren explains that incorrect localization is confusing or frustrating to a user, and accordingly a POSA would be motivated to avoid this: “[i]f the sound continued to localize at the SLP, then the voice appears to originate from another person in the elevator or from a wall in the elevator, and this confuses or frustrates the listening person.”. *E.g.*, Lyren (Ex. 1008), [0072], [0106] (“If the SLP of Max did not move, then the voice of Max appears to originate from the mirror or from the wall behind the mirror or from the visage of Alice in the mirror, and such localization confuses or disquiets Alice.”), [0217] (“When Alice awakes, she forgot that SLPs are being provided through her earphones and, as such, is confused or unable to distinguish between sounds that originate in her physical environment and other sounds provided by her earphones.”), [0225] (“a user is irritated by the positional blurring he perceives to SLPs that do not have a corresponding visual anchor, and the system switches accordingly.”). A POSA would therefore switch to stereo to avoid a confusing/irritating situation for users.

5. '179 Patent - Independent Claim 7

a. Elements 7[pre] to 7[c]

278. Counsel instructed me to assume this preamble was limiting. Lyren teaches these elements for the same reasons discussed above for corresponding Elements 14[pre] to 14[c].

279. A POSA would understand that there is no relevant distinction between displaying “with” the WED as recited in Element 14[b] and displaying “by” the WED as recited Element 7[b]. Similarly, Element 7[c]’s recitation of localizing “to the empty space and to the virtual image” (instead of “in empty space from the virtual image”) does not change the analysis.

b. Element 7[d] determining, with the WED, when the user is leaving the environment; and

280. Lyren and Reif teaches or renders obvious “determining, with the WED, when the user is leaving the environment” for the same reasons as Element 14[d].

c. Element 7[e] alerting the user that the user is leaving the environment by altering processing of the binaural sound to degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound.

281. Lyren and Reif teaches or renders obvious “alerting the user that the user is leaving the environment by altering processing of the binaural sound to degrade audio cues in order to prevent the user from experiencing external

localization of the binaural sound” for the same reasons discussed above for Element 14[e] and dependent Claims 15 and 16.

6. '179 Patent - Dependent claims 8-13

- a. Claim 8. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by reducing a volume of the binaural sound being provided to the user.**

282. Lyren and Reif teaches “alerting the user that the user is leaving the environment by reducing a volume of the binaural sound being provided to the user” for the same reasons as Element 14[e] and Claim 15.

- b. Claim 9. The method of claim 7 further comprising: displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character from the binaural sound to stereo sound when the character vanishes from the environment during the game.**

283. Lyren and Reif teaches “displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character from the binaural sound to stereo sound when the character vanishes from the environment during the game.”

284. Both Lyren and Reif teach playing a computer game. Reif (Ex. 1009), [0040]-[0044] (“In some implementations, the bounding box can have modifiers based on a game type selected via an API. For instance, a game that involves rapid hand motion may cause the bounding box to be modified to accommodate or

compensate for the rapid hand motion.”); Lyren (Ex. 1008), [0282] (“By way of example, electronic devices include, but are not limited to, a computer, [or] electronic and computer game consoles.”). Further, as discussed for Ground 1, a POSA understood that virtual games, which include game characters, were one of the main uses of 3D VR systems.

285. Further, Lyren teaches detecting the triggering event of “a visually rendered character vanishing from [the] virtual environment,” and in response, switches the SLP of the now-vanished character to stereo or mono to prevent confusing or irritating the user. Lyren (Ex. 1008), [0225] (“Events can be triggered by a change in a user's ability to associate visual cues or images with the associated audio cues, such as a visually rendered character vanishing from an augmented or virtual environment For example, when a listener externalizes a SLP in his cone of confusion, a system switches the SLP to stereo or mono to prevent irritation of the user”).

c. Claim 10. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.

286. Lyren and Reif teaches “alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound” for the same reasons as Element 7[e].

- d. Claim 11. The method of claim 7 further comprising: switching the binaural sound to stereo sound upon activation of a switch on the WED.**

287. Lyren and Reif teaches “switching the binaural sound to stereo sound upon activation of a switch on the WED.”

288. Lyren teaches that “[e]lectronic earphones or electronic headphones include a switch or button that when activated causes binaural sound to discontinue or continue (such as providing an on/off switch on the earphones or headphones).” Lyren (Ex. 1008), [0236]; *see also* Lyren (Ex. 1008), [0238] (“Consider an example in which electronic earphones include a switch that turns on and off binaural sound (such as an infrared sensor, push button switch, slide switch, or other physical or electrical switch). ... Movement of a hand to the switch or activation of the switch can switch between binaural and stereo, switch off binaural, switch on binaural, etc.”).

- e. Claim 12. The method of claim 7 further comprising: wherein the WED includes electronic glasses, the environment is a physical room where the user is located, the virtual image is an augmented reality (AR) character that speaks to the user while the user plays a game, the WED includes head tracking, and a digital signal processor (DSP) in the WED reduces a volume of the binaural sound to alert the user when a face of the user is not directed to the virtual image in the environment.**

289. Lyren and Reif teaches “wherein the WED includes electronic glasses, the environment is a physical room where the user is located, the virtual image is an

augmented reality (AR) character that speaks to the user while the user plays a game, the WED includes head tracking, and a digital signal processor (DSP) in the WED reduces a volume of the binaural sound to alert the user when a face of the user is not directed to the virtual image in the environment.”

290. As discussed for Claim 9, Lyren and Reif teach performing the claimed methods while playing a computer game. Further, Lyren teaches tracking the user’s head, and detecting which SLP/image the user is facing (or not facing). Lyren (Ex. 1008), [0242] (“Consider an example in which Alice wears an electronic device that performs head tracking or is in the presence of a device that performs head tracking (such as a head tracking system included in her notebook, in her desktop computer, or in her HPED). Multiple SLPs externally localize around her such that each SLP includes a corresponding image that Alice can see. When she looks at, gazes at, or focuses on a particular SLP and image, then the voice or sound from the other SLPs localizes internally, while the SLP in her focus is perceived at the location of its corresponding image. The system thus switches or changes between internally and externally localizing SLPs based on sensing a gaze of Alice and/or a position of her head with regard to the SLP. Alternatively, a situation exists for her to perceive each SLP around her, except for the SLP that she is looking at, which switches to stereo or mono sound during the time her focus is in its direction.”); *see also* Lyren (Ex.

1008), [0172]-[0176] (“Consider an example in which Alice wears an AR headset or electronic glasses that track her head movement and eye gaze.”).

291. Lyren teaches reducing the volume of an SLP/image (e.g., the voice of the character) when an event, such as not facing the SLP/image (character), is detected. Lyren (Ex. 1008), [0067] (“Sound can switch back and forth from being provided in binaural, stereo, and mono sounds (including switching between different variations of binaural sound, such as binaural sounds having different SLPs, different volumes at SLPs, etc.).”), [0103] (upon detecting event, “lower a volume of the sound” or “raise the volume of the sound”). A POSA would have thus lowered the volume of the speaking character’s voice when the user is not facing the character to help the user focus on what she is facing, and because users expect to better hear sounds in front of them.

- f. Claim 13. The method of claim 7 further comprising: sensing, with the WED, when a physical object interferes with a location of the virtual image in the environment; and switching the binaural sound to stereo sound in response to sensing the physical object interferes with the location of the virtual image in the environment.**

292. Lyren and Reif teaches “sensing, with the WED, when a physical object interferes with a location of the virtual image in the environment; and switching the binaural sound to stereo sound in response to sensing the physical object interferes with the location of the virtual image in the environment.”

293. Lyren teaches that “[a] switch, change, or other action with regard to the SLP or the binaural sound can occur when the object conflicts, interferes (i.e., collides with, comes near, overlaps, touches, or hinders), overlaps, approaches, exists in, or exists near the person or a SLP of the person.”). Lyren (Ex. 1008), [0110]; *see also* Lyren (Ex. 1008), [0052] (“As shown in the transition from box 140B to box 140C, when a third person 146 enters and sits at the chair next to user 134, the system removes the sound localization point 144 since the third person 146 physically occupies the space where the sound localization point 144 existed. The third person 146 collides, interferes, or overlaps with the SLP 144. ... As such, the system decides to switch the call to stereo.”), [0072] (“While listening to this recording, the person enters an elevator full of people. If the sound continued to localize at the SLP, then the voice appears to originate from another person in the elevator or from a wall in the elevator, and this confuses or frustrates the listening person. In response to this event of entering the elevator, the HPED automatically switches the sound of the recording so that the earphones present the sound of the voice recording in stereo sound when the listener enters the elevator.”), [0105] (“Charlie walks up to Alice and interferes with the SLP by entering within a predetermined area or zone of Alice. When Charlie enters this zone, an event occurs (i.e., Charlie's presence interferes with the SLP). For example, when Charlie comes within three feet of Alice, the voice of Bob that Alice hears switches from binaural

sound to stereo or mono sound. As another example, when Charlie moves within or proximate to a zone or area of the SLP (i.e., location of Bob's voice), the voice of Bob that Alice hears switches from binaural sound localized three feet from Alice to binaural sound localized one foot from Alice.”).

7. '179 Patent - Independent Claim 1

a. Elements 1[pre] to 1[c]

294. Counsel instructed me to assume this preamble was limiting. Lyren teaches these elements for the same reasons discussed above for corresponding Elements 14[pre] to 14[c].

295. A POSA would understand that the HMD described in Element 1[b] is same as the wearable electronic device (WED) of Element 14[b]. A POSA would understand that the HMD described for Element 14[b] includes sensors that perform the claimed sensing of Element 1[d] and includes memory that performs the storing of Element 1[b]. There is no relevant distinction between Element 1[c]’s localizing “in empty space to the virtual image” and “in empty space from the virtual image” in Element 14[c].

b. Element 1[d] sensing, with one or more sensors in the HMD, when the user leaves the 3D virtual environment that includes the virtual image; and

296. Lyren and Reif teaches or renders obvious “sensing, with one or more sensors in the HMD, when the user leaves the 3D virtual environment that includes the virtual image” for the same reasons as Element 14[d].

- c. **Element 1[e] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter audio cues to prevent the user from experiencing external localization of the sounds and by lowering a volume of the sounds being provided to the user.**

297. Lyren and Reif teaches or renders obvious “warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter audio cues to prevent the user from experiencing external localization of the sounds and by lowering a volume of the sounds being provided to the user” for the same reasons discussed above for Element 14[e] and dependent Claims 15 to 18.

8. '179 Patent - Dependent Claims 2-6

- a. **Claim 2. The method of claim 1 further comprising: warning the user that the user is leaving the 3D virtual environment by removing, by the DSP, background sounds that occur in the 3D virtual environment as part of a game the user is playing.**

298. Lyren and Reif teaches “warning the user that the user is leaving the 3D virtual environment by removing, by the DSP, background sounds that occur in the 3D virtual environment as part of a game the user is playing” for the same reasons as Element 14[e].

299. As discussed for Claim 9, Lyren and Reif teach performing the claimed methods while playing a computer game. Lyren expressly discloses triggers will change/switch the background sounds. Lyren (Ex. 1008), [0089] (“As yet another

example, a person receives binaural sound with a first set of background sound, and a switch occurs such that the person receives binaural sound with a second set of background sound. A person can receive binaural sound corresponding to one virtual or real or augmented space, and a switch occurs such that the person receives binaural sound from a second virtual or real or augmented space.”). Lyren further teaches that the environment’s sounds, including background sounds, may be reduced or muted. *See* Lyren (Ex. 1008), [0103] (“In FIGS. 4-16, example actions include, but are not limited to, one or more of ... stop binaural sound, discontinue playing sound, mute the sound, lower a volume of the sound”); [0260] (sound sources subject to such changes include background grounds).

b. Claim 3. The method of claim 1, wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment.

300. Lyren and Reif teaches or renders obvious “wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment” for the same reasons as Element 14[e].

301. Lyren teaches that when a third person “travels ‘off stage’ by leaving the room can still be included in the conversation, but the sound can switch to mono or stereo.” Lyren (Ex. 1008), [0232]; *see also* Lyren (Ex. 1008), [0053] (person walks out the door and out of the sight of the user), [0065] (“a person walking into a room” is a triggering event); [0067] (also lowering the volume), [0103] (same).

302. A POSA would have found it obvious to implement the converse. When the user is the one who leaves the (virtual) room, the system would alter the audio cues to prevent the user from experiencing external localization of the sounds from the other people remaining in the room and by lowering a volume of those sounds. The reasoning is the same whether three people leave the room and the user remains behind, or if the user leaves the room and three people remain. In either case, the user is now in another room apart from the others and would expect the volume to be reduced and to be unable to localize the persons' locations because they are in a different room.

303. In addition, Reif teaches first triggering the warning when the user crosses from the safe region (i.e., a virtual room) into the warning region (which is still in the environment), and gradually fading out the virtual environment as the user traverses the warning region. *E.g.*, Reif (Ex. 1009), [0017]-[0018] (fading out the virtual environment until “the user’s location is determined to be outside the bounding box” which causes the “HMD 150 to cease displaying the virtual environment to the user.”). A POSA implementing the safe and warning zones of Reif would have likewise reduced the volume of the sounds and the audio cues gradually and while still inside the environment (i.e., still able to perceive it) to correspond to the visual fade out as the user travels through the warning region. A POSA would understand that such a smooth transition would provide an improved,

more immersive experience, as well as being more pleasing, coherent, and seamless. Oculus Connect 5: Spatial Audio for Oculus Quest and Beyond (Ex. 1085), 13:35-13:45 (“ramps nice and smooth ... because you want to keep everybody immersed”).

- c. **Claim 4. The method of claim 1 further comprising: tracking, with head tracking in the HMD, head movements of the user; and switching the sounds from the binaural sound to stereo sound when a face of the user is directed away from the virtual image.**

304. Lyren teaches or renders obvious “tracking, with head tracking in the HMD, head movements of the user; and switching the sounds from the binaural sound to stereo sound when a face of the user is directed away from the virtual image.”

305. As discussed for Claim 12, Lyren teaches tracking the user’s head, and detecting which SLP/image the user is facing (or not facing). Lyren (Ex. 1008), [0242] (“Consider an example in which Alice wears an electronic device that performs head tracking or is in the presence of a device that performs head tracking (such as a head tracking system included in her notebook, in her desktop computer, or in her HPED).”), [0172]-[0176]. Lyren further teaches switching the binaural sound of an SLP/image to stereo when an event, such as not facing the SLP/image, is detected because that indicates the user is not focused on that object, while providing binaural sound from the object that is the focus of the user’s attention. Lyren (Ex. 1008), [0242] (“The system thus switches or changes between internally

and externally localizing SLPs based on sensing a gaze of Alice and/or a position of her head with regard to the SLP. The system thus switches or changes between internally and externally localizing SLPs based on sensing a gaze of Alice and/or a position of her head with regard to the SLP.”), [0067] (“The sound changes from being provided in binaural sound to being provided in stereo sound.”), [0103] (upon detecting event, “switch the sound from binaural sound to stereo sound”); [0233] (“Focus of the listener, for example, can be judged by a gaze”). Thus, a POSA would have switched the less-relevant audio (i.e., the audio from images the user is not facing) to stereo, as taught by Lyren. *See also* Sauk (Ex. 1007), 11:26-41 (switching less-relevant audio to stereo).

- d. **Claim 5. The method of claim 1 further comprising: switching the sounds from the binaural sound to stereo sound when the user activates a physical switch located in the HMD.**

306. Lyren teaches “switching the sounds from the binaural sound to stereo sound when the user activates a physical switch located in the HMD” for the same reasons as Claim 11. Lyren (Ex. 1008), [0236], [0238].

- e. **Claim 6. The method of claim 1 further comprising: switching the sounds from the binaural sound to stereo sound when the virtual image vanishes in the 3D virtual environment.**

307. Lyren teaches “switching the sounds from the binaural sound to stereo sound when the virtual image vanishes in the 3D virtual environment.”

308. Lyren teaches “[e]vents can be triggered by a change in a user's ability to associate visual cues or images with the associated audio cues, such as a visually rendered character vanishing from an augmented or virtual environment.” Lyren (Ex. 1008), [0225]; *see also* Claim 9.

X. The '184 Patent

309. The '184 Patent's claims are similar (and some limitations are identical) to the '179 Patent's claims, and would have been obvious for the same reasons, with exceptions that I will discuss as they arise.

310. Accordingly, my analysis of the '184 Patent will frequently refer back to the corresponding claim or claim element in the '179 Patent's analysis. This is true for both Grounds (i.e., the Clement-based Ground 1, and the Lyren-based Ground 2).

A. '184 Patent - Ground 1: Claims 1, 3-4, 6-10, and 14-19 Are Unpatentable Over Clement, Osman, and Sauk

311. I explain below, on an element-by-element basis, how every element of Claims 1, 3-4, 6-10, and 14-19 of the '184 Patent is either disclosed in or obvious over the combination of Clement with Osman and Sauk, as well as why a POSA would have considered it obvious to implement Clement with Osman and Sauk.

1. Reasons to Implement Clement with Osman

312. A POSA would have considered it obvious to, and have good reason to, implement the system described in Clement using 3D or binaural sound, as taught in

Osman for the same reasons as discussed for the Motivation to Combine this combination for the '179 Patent.

2. Reasons to Implement Clement and Osman with Sauk

313. A POSA would have considered it obvious to, and have good reason to, implement the system described in Clement and Osman with the switch from binaural sound to stereo or mono sound described in Sauk for the same reasons as discussed for the Motivation to Combine this combination for the '179 Patent.

3. '184 Patent - Independent Claim 14

a. Elements 14[pre] to 14[c]

314. Counsel instructed me to assume this preamble was limiting. Elements 14[pre] to 14[c] are taught or rendered obvious by Clement and Osman for the same reasons as Elements 14[pre] to 14[c] of Ground 1 for the '179 Patent.

315. Elements 14[pre] to 14[c] in the '184 Patent are substantively identical to Elements 14[pre] to 14[c] in the '179 Patent. The only difference is that Element 14[c] of the '179 Patent recites that the binaural sound “localizes in empty space from the virtual image,” where the '184 Patent recites that sound “localizes to the user at the virtual image.” In other words, the '179 Patent is narrower than the '184 Patent because it requires that virtual image to be at a point of empty space. These elements are therefore taught by Clement and Osman for the same reasons as Elements 14[pre] to 14[c] of Ground 1 for the '179 Patent.

- b. Element 14[d]: alerting the user that the user is leaving the virtual image in the environment by altering processing of the binaural sound by the DSP and reducing a volume of the binaural sound that localizes to the user at the virtual image.**

316. Unlike its counterpart in the '179 Patent, Claim 14 in the '184 Patent does not recite determining when the user is leaving the environment. It instead jumps forward to an alert, without regard how the triggering event is detected.

317. Element 14[d] in the '184 Patent is the same as Element 14[e] in the '179 Patent, except that it recites “altering processing of the binaural sound by the DSP *and* reducing a volume of the binaural sound that localizes to the user at the virtual image” rather than merely “reducing the volume of the binaural sound that localizes from the virtual image and that is being provided to the user” as in the '179 Patent. Accordingly, Element 14[d] in this patent is taught or rendered obvious by the Clement combination for the same reasons discussed for Elements 14[d]-[e] (reducing the volume) and Claim 16 (altering processing of the binaural sound by the DSP) in Ground 1 of the '179 Patent.

318. Further, a POSA would understand that there is no meaningful difference between “binaural sound that localizes from the virtual image and that is being provided to the user” and “binaural sound that localizes to the user at the virtual image.”

4. '184 Patent - Dependent Claims 15-19

- a. **Claim 15. The method of claim 14, wherein the DSP alters processing of the binaural sound to alert the user by changing a spatial cue that includes at least one of an interaural time difference (ITD), an interaural level difference (ILD), and the HRTFs.**

319. Clement and Osman with Sauk teaches “wherein the DSP alters processing of the binaural sound to alert the user by changing a spatial cue that includes at least one of an interaural time difference (ITD), an interaural level difference (ILD), and the HRTFs.”

320. As discussed for Claim 17 of Ground 1 for the '179 Patent, switching from binaural sound to stereo or mono sound can be performed by the removal of time delays, such as ITDs. *See* Ojala (Ex. 1010), [0006], [0023]-[0024], claim 1; (discussing the removal of time delays from multichannel (spatialized) audio to produce a single channel (monophonic) audio); Moon (Ex. 1035), claim 1 (same); Begault (Ex. 1038), 40-41, 49-50. A POSA would have understood that switching sound from binaural to monophonic sound, as described in Sauk, includes removing, and thus changing, ITDs. Begault (Ex. 1038), 41; Sauk (Ex. 1007), 7:1-8:33, 11:26-41, Fig. 5.

321. Further, binaural sound is derived from audio cues including ITDs, as well as ILDs and HRTFs to provide localized sound. '184 Patent, 3:43-58. A POSA understood that switching from binaural to stereo sound, eliminates these audio cues, as well as spectral coloration caused by the shape of the human head, torso and

pinnae or other cues such as filtering based on the orientation of the head, in the binaural sound provided to the user. Sauk (Ex. 1007), 11:35-40 (“This process also removes any localization cues for that audio.”).

b. Claim 16. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by stopping, by the DSP, the processing of the sound with the HRTFs.

322. Clement and Osman with Sauk teaches “alerting the user that the user is leaving the environment by stopping, by the DSP, the processing of the sound with the HRTFs.”

323. First, as discussed above with respect to Element 14[d] in this Ground for this patent, Clement teaches degrading the sound by fading it out (reducing its volume) such that “[i]f, at block 1057, it is determined that the transition profile includes audio parameters, then, at block 1058, the virtual audio output associated with the virtual world and output to the user via the audio output device 130 of the HMD 100 may be replaced with transition audio output in accordance with the profile” and “this replacement of the virtual audio output with transition audio output may include, for example, a serial fade out of the virtual audio followed by a fade in of the transition audio output.” Clement (Ex. 1005), 12:47-65. A POSA would understand that after the binaural sound has completely faded out, it is no longer processed.

324. In addition and in the alternative, as discussed for Element 14[d], Clement, Osman, and Sauk teaches switching from binaural to stereo or mono sound. One way that the DSP does this is by stopping the DSP's processing of the binaural sound with HRTFs. As discussed for Element 14[c] of Ground 1 for the '179 Patent, the DSP uses HRTFs to process stereo or mono sound, adding audio cues to generate the binaural sound that is provided to the user. If the DSP stops processing the sound with the HRTFs, the original unprocessed stereo or mono audio will be provided instead. For example, it was known that one form for this unprocessed sound is "a valid mono signal," which can be directly played. Faller (Ex. 1044), 2 ("Since the transmitted single audio channel is a valid mono signal, it is suitable for playback by the legacy receivers.").

- c. **Claim 17. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.**

325. The additional limitation recited in Claim 17 of the '184 Patent is identical to the additional limitation of Claim 17 of the '179 Patent. This additional limitation is therefore taught by Clement, Osman, and Sauk for the same reasons as described for Claim 17 of Ground 1 for the '179 Patent.

- d. **Claim 18. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by eliminating one or more spatial cues in the binaural sound being provided to the user.**

326. Clement and Osman with Sauk teaches “alerting the user that the user is leaving the environment by eliminating one or more spatial cues in the binaural sound being provided to the user.”

327. According to the ’184 Patent, spatial cues are ITDs and ILDs or head-related-transfer-functions (HRTFs). *See, e.g.*, ’184 Patent, 3:10-17. As described for Claim 17 in this Ground for this patent, and for Claim 17 of Ground 1 for the ’179 Patent, Clement and Osman with Sauk teaches removing ITDs in the binaural sound being provided to the user.

328. Further, as described for Claim 15 in this Ground for this patent, a POSA understood that switching from binaural to stereo sound eliminates audio cues, including ITDs, ILDs, and HRTFs, as well as spectral coloration caused by the shape of the human head, torso and pinnae or other cues such as filtering based on the orientation of the head, in the binaural sound provided to the user. Sauk (Ex. 1007), 11:35-40 (“This process also removes any localization cues for that audio.”).

**e. Claim 19. The method of claim 14 further comprising:
alerting the user that the user is leaving the
environment by switching the binaural sound playing
to the user to one of stereo sound and mono sound.**

329. Clement and Osman with Sauk teaches “alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to one of stereo sound and mono sound” for the same reasons as Element 14[d] of this Ground of this patent.

5. '184 Patent - Independent Claim 7

330. Independent Claim 7 is generally similar to Claim 14, except that it alerts the user by “altering processing of the sound by the DSP to degrade a spatial cue and to lower a volume of the sound playing to the user.”

a. Elements 7[pre] to 7[c]

331. Counsel instructed me to assume this preamble was limiting. Elements 7[pre] to 7[c] are taught by Clement and Osman for the same reasons as Elements 7[pre] to 7[c] of Ground 1 for the '179 Patent.

332. Elements 7[pre] to 7[c] in the '184 Patent are substantively identical to Elements 7[pre] to 7[c] in the '179 Patent. The only differences are that Element 7[b] of the '179 Patent recites “storing head-related transfer functions (HRTFs) in memory” where the '184 Patent recites merely “storing head-related transfer functions (HRTFs),” and that Element 7[c] of the '179 Patent recites that the DSP “generate binaural sound in empty space that is provided to the user and that externally localizes to the empty space and to the virtual image in the environment,” where the '184 Patent recites that the DSP “generate binaural sound that plays to the user and that externally localizes at the virtual image in the environment.” In other words, the '179 Patent is narrower than the '184 Patent because it requires that the storage of HRTFs be in memory and that the virtual image be at a point of empty

space. These elements are therefore taught by Clement and Osman for the same reasons as Elements 7[pre] to 7[c] of Ground 1 for the '179 Patent.

333. Likewise, the differences between Elements 7[pre] to 7[c] and Elements 14[pre] to 14[c] of the '184 Patent do not change my analysis. Specifically, there is no relevant distinction between displaying “with” the WED as recited in Element 14[b] and displaying “by” the WED as recited Element 7[b] or between “binaural sound that plays to the user and that externally localizes at the virtual image in the environment” in Element 7[c] compared to “binaural sound that externally localizes to the user at the virtual image in the environment” in Element 14[c].

- b. Element 7[d] alerting the user that the user is leaving the environment by altering processing of the sound by the DSP to degrade a spatial cue and to lower a volume of the sound playing to the user.**

334. Unlike its counterpart in the '179 Patent, Claim 7 in the '184 Patent does not recite determining when the user is leaving the environment. It instead jumps forward to an alert, without regard how the triggering event is detected.

335. Element 7[d] of the '184 Patent is nearly the same as Element 7[e] of the '179 Patent. Rather than “degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound,” however, the '184 Patent recites “degrade a spatial cue and to lower a volume of the sound playing to the user.”

336. Accordingly, Element 7[d] in this patent is taught or rendered obvious by the Clement combination. Clement teaches degrading the sound by fading it out (reducing its volume) as discussed for Element 14[d] in this Ground of this patent. In addition and in the alternative, converting the audio to stereo or mono alters or removes the inserted ITDs and ILDs in the binaural sound as discussed for Claims 15 and 18 in this Ground of this patent. A POSA understood that modifying or removing spatial cues are ways to degrade them.

6. '184 Patent - Dependent Claims 8-10

a. Claim 8. The method of claim 7, wherein the spatial cue includes an interaural time difference (ITD).

337. Clement with Osman and Sauk teaches “wherein the spatial cue includes an interaural time difference (ITD)” for the same reasons as Element 7[d] of this Ground for this Patent.

b. Claim 9. The method of claim 7 further comprising: displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character from the binaural sound to stereo sound when the WED no longer displays the virtual image during the game.

338. The additional limitation recited in Claim 9 of the '184 Patent is nearly identical to the additional limitation of Claim 9 of the '179 Patent. Rather than triggering “when the character vanishes from the environment during the game,” however, the '184 Patent recites triggering “when the WED no longer displays the

virtual image during the game.” A POSA would understand that there is no relevant distinction between the character vanishing and the character no longer being displayed by the WED. Accordingly, this additional limitation is therefore taught or rendered obvious by Clement, Osman, and Sauk for the same reasons as described for Claim 9 of Ground 1 for the ’179 Patent.

c. Claim 10. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.

339. The additional limitation recited in Claim 10 of the ’184 Patent is identical to the additional limitation of Claim 10 of the ’179 Patent. This additional limitation is therefore taught by Clement, Osman, and Sauk for the same reasons as described for Claim 10 of Ground 1 for the ’179 Patent.

7. ’184 Patent - Independent Claim 1

340. Independent Claim 1 is generally similar to Claim 14, except that it alerts the user in a different manner.

a. Elements 1[pre] to 1[c]

341. Counsel instructed me to assume this preamble was limiting. Elements 1[pre] to 1[c] are taught by Clement with Osman and Sauk for the same reasons as Elements 1[pre] to 1[c] of Ground 1 for the ’179 Patent.

342. Elements 1[pre] to 1[c] in the ’184 Patent are substantively identical to Elements 1[pre] to 1[c] in the ’179 Patent. The only differences are that Element

1[a] of the '179 Patent recites “storing head-related transfer functions (HRTFs) in memory of a head mounted display (HMD) worn on a head of a user,” where the '184 Patent merely recites “storing head-related transfer functions (HRTFs),” that Element 1[c] of the '179 Patent recites “a digital signal processor (DSP) *in the HMD*,” where the '184 Patent merely recites “a digital signal processor (DSP).” In other words, the '179 Patent is narrower than the '184 Patent because it requires storage of the HRTFs in memory at the HMD and locating the DSP at the HMD. These elements are therefore taught by Clement and Osman for the same reasons as Elements 1[pre] to 1[c] of Ground 1 for the '179 Patent.

343. Further a POSA would understand that there is no meaningful difference between “binaural sound that *is provided* to the user and that externally localizes in empty space *to* the virtual image in the 3D virtual environment” and “binaural sound that *plays* to the user and that externally localizes in empty space *at* the virtual image in the 3D virtual environment.”

344. Likewise, the differences between Elements 1[pre] to 1[c] and Elements 14[pre] to 14[c] of the '184 Patent do not change my analysis. Specifically, a POSA would understand that the HMD described in Element 1[b] is same as the wearable electronic device (WED) of Element 14[b] and that there is no relevant distinction between “binaural sound that plays to the user and that externally localizes in empty space at the virtual image “ in Element 1[c] and “binaural sound

that externally localizes to the user at the virtual image in the environment” in Element 14[c].

- b. Element 1[d] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter a spectral cue and lower a volume of the sound playing to the user.**

345. Unlike its counterpart in the '179 Patent, Claim 1 in the '184 Patent does not recite sensing, with one or more sensors in the HMD, when the user leaves the 3D virtual environment that includes the virtual image. It instead jumps forward to an alert, without regard how the triggering event is detected.

346. Element 1[d] of the '184 Patent is nearly the same as Element 1[e] of the '179 Patent. Rather than “alter audio cues to prevent the user from experiencing external localization of the sounds,” however, the '184 Patent recites “alter a spectral cue.”

347. Spectral cues include ILDs which are also audio cues for providing localization of sounds. *See* '184 Patent, 3:10-17. Accordingly, Element 1[d] in this patent is taught or rendered obvious by the Clement combination for the same reasons discussed for Elements 1[d]-[e] in Ground 1 of the '179 Patent.

8. '184 Patent - Dependent Claims 3-4, and 6

- a. Claim 3. The method of claim 1, wherein the warning occurs when the user is leaving a virtual area where the virtual image is located while remaining in the 3D virtual environment.**

348. The additional limitation recited in Claim 3 of the '184 Patent is nearly identical to the additional limitation of Claim 3 of the '179 Patent. Rather than a “virtual room,” however, the '184 Patent recites a “virtual area where the virtual image is located.” A POSA would understand that there is no relevant distinction between a virtual room and a virtual area. This additional limitation is therefore taught by Clement and Osman for the same reasons as described for Claim 3 of Ground 1 for the '179 Patent.

- b. Claim 4. The method of claim 1 further comprising: tracking, with head tracking in the HMD, head movements of the user; and switching the sound from the binaural sound to stereo sound when a face of the user is directed away from the virtual image such that the virtual image is no longer visible to the user.**

349. The additional limitations recited in Claim 4 of the '184 Patent are nearly identical to the additional limitations of Claim 4 of the '179 Patent. Rather than a “when a face of the user is directed away from the virtual image,” however, the '184 Patent recites a “when a face of the user is directed away from the virtual image such that the virtual image is no longer visible to the user.” This difference does not change the analysis as in both instances the user has directed their gaze away from the virtual image. These additional limitations are therefore taught or rendered obvious by Clement with Osman and Sauk for the same reasons as described for Claim 4 of Ground 1 for the '179 Patent.

- c. **Claim 6. The method of claim 1 further comprising: switching the sound from the binaural sound to stereo sound when the virtual image vanishes from the 3D virtual environment.**

350. The additional limitation recited in Claim 6 of the '184 Patent is nearly identical to the additional limitation of Claim 6 of the '179 Patent. Rather than “sounds,” however, the '184 Patent recites a “sound.” A POSA would understand that for the same reasons Clement with Osman and Sauk renders obvious switching “sounds,” the combination also renders obvious switching the “sound.” This additional limitation is therefore taught by Clement with Osman and Sauk for the same reasons as described for Claim 6 of Ground 1 for the '179 Patent.

B. '184 Patent - Ground 2: All Claims Are Unpatentable Over Lyren and Reif

351. I explain below, all claims of the '184 Patent are either disclosed in or obvious over the combination of Lyren and Reif, as well as why a POSA would have considered it obvious to implement Lyren and Reif.

1. The '184 Patent's claims are not supported by its parent, the '140 Application

352. Counsel has asked me to evaluate whether the '140 Application provides sufficient written description support for the subject matter claimed in the '184 Patent. My conclusion is the same as it was for the '179 Patent: the '140 Application lacks sufficient written description support for the claims of the '184 Patent.

353. My analysis is nearly the same as the one I performed above for the '179 Patent. The '184 Patent's independent claims are similar to those of the '179 Patent, but they lack the sensing/determining elements. That is, Elements 1[d], 7[d], and 14[d] of the '179 Patent are not found in corresponding Claims 1, 7, and 14 in the '184 Patent. However, the altering/warning Elements 1[d], 7[d], and 14[d] of the '184 Patent are very similar to the corresponding alerting/warning elements 1[e], 7[e], and 14[e] of the '179 Patent.

354. As I explained with respect to the '179 Patent above, I have reviewed the '140 Application (i.e., Exhibit 1028 at pages 92-191) and it lacks sufficient support for the alerting/sensing Elements in Claims 1, 7, and 14 of the '179 Patent. The same remains true for the corresponding elements of Claims 1, 7, 14 of the '184 Patent.

355. Further, and as I discussed with respect to the '179 Patent's claims, the '140 Application lacks any disclosure of sensing or detecting the user leaving the environment. Lack of disclosure of a means for sensing or detecting the user "leaving the environment" further confirms my opinion that "leaving the environment" is not a trigger taught by the '140 Application.

2. Combination of Lyren and Reif

356. My analysis of this combination for the '184 Patent is, for the most part, identical to my analysis above for the '179 Patent. I will point out differences as

they occur. But in brief, Lyren teaches nearly each and every claim element of the '184 Patent, except the recited alerting/warning that user that the user is leaving the environment.

3. '184 Patent - Independent Claim 14

a. 14[pre] to 14[c]

357. Counsel instructed me to assume this preamble was limiting. Elements 14[pre] to 14[c] of the '184 Patent are substantively identical to the corresponding elements in the '179 Patent. The only difference is that Element 14[c] of the '179 Patent recites that the binaural sound “localizes in empty space from the virtual image,” where the '184 Patent recites that sound “localizes to the user at the virtual image.” In other words, the '179 Patent is narrower than the '184 Patent because it requires that virtual image to be at a point of empty space.

358. These Elements are therefore taught by Lyren for the same reasons as Elements 14[pre] to 14[c] of Ground 2 for the '179 Patent.

b. Element 14[d]: alerting the user that the user is leaving the virtual image in the environment by altering processing of the binaural sound by the DSP and reducing a volume of the binaural sound that localizes to the user at the virtual image.

359. Unlike its counterpart in the '179 Patent, Claim 14 in the '184 Patent does not recite determining when the user is leaving the environment. It instead jumps forward to an alert, without regard how the triggering event is detected.

360. Element 14[d] of the '184 Patent is nearly the same as Element 14[e] of the '179 Patent. Rather than merely reducing the sound to alert the user, however, the '184 Patent also recites “altering the processing of the binaural sound by the DSP.”

361. As I explained for Elements 14[d]-[e] in Ground 2 for the '179 Patent, the combination teaches or renders obvious this claim element. Regarding the “altering the processing” feature, I note that I specifically discussed, e.g., switching the binaural sound to stereo or mono—which is altering the processing—in addition to reducing the volume, citing, *e.g.*, Lyren (Ex. 1008), [0103].

362. Further, a POSA would understand that there is no meaningful difference between “binaural sound that localizes from the virtual image and that is being provided to the user” and “binaural sound that localizes to the user at the virtual image.”

4. '184 Patent - Dependent Claims 15-20

- a. **Claim 15. The method of claim 14, wherein the DSP alters processing of the binaural sound to alert the user by changing a spatial cue that includes at least one of an interaural time difference (ITD), an interaural level difference (ILD), and the HRTFs.**

363. Lyren and Reif teaches “wherein the DSP alters processing of the binaural sound to alert the user by changing a spatial cue that includes at least one

of an interaural time difference (ITD), an interaural level difference (ILD), and the HRTFs.”

364. As discussed for Claim 15 of Ground 2 of the '179 Patent, Lyren teaches switching binaural sound to stereo or mono upon detecting a triggering event. Specifically, Lyren teaches that this may be done by using a DSP to “degrade, alter, or eliminate sufficient audio cues to prevent a listener from experiencing external localization from a binaural audio source.” Lyren (Ex. 1008), [0044]-[0046]. Lyren teaches that “[f]or example, after DSP processing, the user perceives the sound with less, little, or no external localization” and “[f]or example, a DSP process removes or re-normalizes interaural time differences (ITDs) in source impulses to cause imprecise or zero azimuth angle perception.” Lyren (Ex. 1008), [0045]. Lyren further teaches that “[a]nother way of changing a sound from being perceived as binaurally captured audio or binaurally manufactured audio to being perceived as non-binaural audio is to render the original source with different spatial parameters or to re-render a sound source with different spatial parameters in order to adjust, degrade, or eliminate certain human audio cues [such as rendering] the sound source to specific alterations per an ITD and/or interaural level difference (ILD), or discontinue[ing] rendering or using the HRTF or ITD/ILD calculations while continuing to render other aspects of the audio without pause.” Lyren (Ex. 1008), [0046]; *see also* Lyren, [0103] (“In FIGS. 4-16, example actions include, but

are not limited to, one or more of ... alter a HRTF of a person [and] change or alter an ITD or an ILD”).

b. Claim 16. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by stopping, by the DSP, the processing of the sound with the HRTFs.

365. Lyren and Reif teaches “alerting the user that the user is leaving the environment by stopping, by the DSP, the processing of the sound with the HRTFs.”

366. As discussed for Claim 15 of Ground 1 for the ’179 Patent, the DSP alters its processing to provide stereo or mono audio. One way that the DSP may provide stereo or mono audio is to stop the DSP’s processing of the binaural sound with HRTFs. As discussed for Element 14[c] of Ground 1 for the ’179 Patent, the DSP uses HRTFs to generate binaural sound that is provided to the user. If the DSP stops processing the sound with the HRTFs, unprocessed stereo or mono audio will be provided instead. Lyren (Ex. 1008), [0046] (“Another way of changing a sound from being perceived as binaurally captured audio or binaurally manufactured audio to being perceived as non-binaural audio is to render the original source with different spatial parameters or to re-render a sound source with different spatial parameters in order to adjust, degrade, or eliminate certain human audio cues. For example, a SLS renders a source sound using an HRTF to adjust SLPs, renders the sound source to specific alterations per an ITD and/or interaural level difference (ILD), or discontinues rendering or using the HRTF or ITD/ILD calculations while

continuing to render other aspects of the audio without pause.”). Further, Lyren expressly discloses stopping HRTF processing in response to an event. *See* Lyren (Ex. 1008), [0103] (“stop binaural sound, discontinue playing sound, mute the sound ... cancel a SLP, ”), [0124] (“The earphones immediately stop localizing the voice of Bob”), [0246] (“the binaural music stops playing”), [0292] (“stopping transmission or emission of binaural sound, etc.”).

- c. **Claim 17. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.**

367. The additional limitation recited in Claim 17 of the '184 Patent is identical to the additional limitation of Claim 17 of the '179 Patent. This additional limitation is therefore taught by Lyren and Reif for the same reasons as described for Claim 17 of Ground 2 for the '179 Patent.

- d. **Claim 18. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by eliminating one or more spatial cues in the binaural sound being provided to the user.**

368. Lyren and Reif teaches “alerting the user that the user is leaving the environment by eliminating one or more spatial cues in the binaural sound being provided to the user.”

369. According to the '184 Patent, spatial cues are ITDs and ILDs or head-related-transfer-functions (HRTFs). *See, e.g.,* '184 Patent, 3:10-17. Lyren teaches

eliminating ITDs, ILDs, or HRTFs for the same reasons as Claim 15 of this Ground for this patent. *See also* Lyren (Ex. 1008), [0044]-[0045] (using “DSP or other processor to filter the binaural sound and degrade, alter, or eliminate sufficient audio cues to prevent a listener from experiencing external localization from a binaural audio source” and for example “remov[ing] or re-normaliz[ing] interaural time differences (ITDs) in source impulses to cause imprecise or zero azimuth angle perception.”), [0046] (“Another way of changing a sound from being perceived as binaurally captured audio or binaurally manufactured audio to being perceived as non-binaural audio is to render the original source with different spatial parameters or to re-render a sound source with different spatial parameters in order to adjust, degrade, or eliminate certain human audio cues [such as rendering] the sound source to specific alterations per an ITD and/or interaural level difference (ILD), or discontinue[ing] rendering or using the HRTF or ITD/ILD calculations while continuing to render other aspects of the audio without pause.”), [0103] (“In FIGS. 4-16, example actions include, but are not limited to, one or more of ... alter a HRTF of a person [and] change or alter an ITD or an ILD ...”).

- e. **Claim 19. The method of claim 14 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to one of stereo sound and mono sound.**

370. Lyren and Reif teaches “alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to one of stereo

sound and mono sound” for the same reasons as element 14[d] of this Ground of this patent.

- f. Claim 20. The method of claim 14 further comprising: detecting, with a microphone in the WED, a voice of a person in the environment of the user; and switching the binaural sound to stereo sound in response to detecting the voice.**

371. The additional limitations recited in Claim 20 of the '184 Patent are identical to the additional limitations of Claim 20 of the '179 Patent. These additional limitations are therefore taught by Lyren for the same reasons as described for Claim 20 of Ground 2 for the '179 Patent.

5. '184 Patent - Independent Claim 7

a. Elements 7[pre] to 7[c]

372. Counsel instructed me to assume this preamble was limiting. Elements 7[pre] to 7[c] are taught by Lyren for the same reasons as Elements 7[pre] to 7[c] of Ground 1 for the '179 Patent.

373. Elements 7[pre] to 7[c] in the '184 Patent are substantively identical to Elements 7[pre] to 7[c] in the '179 Patent. The only difference are that Element 7[b] of the '179 Patent recites “storing head-related transfer functions (HRTFs) in memory” where the '184 Patent recites merely “storing head-related transfer functions (HRTFs),” and that Element 7[c] of the '179 Patent recites that the DSP “generate binaural sound in empty space that is provided to the user and that externally localizes to the empty space and to the virtual image in the environment,”

where the '184 Patent recites that the DSP “generate binaural sound that plays to the user and that externally localizes at the virtual image in the environment.” In other words, the '179 Patent is narrower than the '184 Patent because it requires that the storage of HRTFs be in memory and that the virtual image be at a point of empty space. These elements are therefore taught by Lyren for the same reasons as Elements 7[pre] to 7[c] of Ground 2 for the '179 Patent.

374. Likewise, the differences between Elements 7[pre] to 7[c] and Elements 14[pre] to 14[c] of the '184 Patent do not change my analysis. Specifically, there is no relevant distinction between displaying “with” the WED as recited in Element 14[b] and displaying “by” the WED as recited Element 7[b] or between “binaural sound that plays to the user and that externally localizes at the virtual image in the environment” in Element 7[c] compared to “binaural sound that externally localizes to the user at the virtual image in the environment” in Element 14[c].

- b. Element 7[d] alerting the user that the user is leaving the environment by altering processing of the sound by the DSP to degrade a spatial cue and to lower a volume of the sound playing to the user.**

375. Unlike its counterpart in the '179 Patent, Claim 7 in the '184 Patent does not recite determining when the user is leaving the environment. It instead jumps forward to an alert, without regard how the triggering event is detected.

376. Element 7[d] of the '184 Patent is nearly the same as Element 7[e] of the '179 Patent. Rather than “degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound,” however, the '184 Patent recites “degrade a spatial cue and to lower a volume of the sound playing to the user.”

377. As discussed for Element 14[d] of this Ground for this patent, Lyren teaches lowering a volume. According to the '184 Patent, spatial cues are ITDs and ILDs or head-related-transfer-functions (HRTFs). *See, e.g.*, '184 Patent, 3:10-17. Lyren further teaches degrading ITDs, ILDs, or HRTFs. Lyren (Ex. 1008), [0044]-[0045] (using “DSP or other processor to filter the binaural sound and degrade, alter, or eliminate sufficient audio cues to prevent a listener from experiencing external localization from a binaural audio source” and for example “remov[ing] or re-normaliz[ing] interaural time differences (ITDs) in source impulses to cause imprecise or zero azimuth angle perception.”), [0046] (“Another way of changing a sound from being perceived as binaurally captured audio or binaurally manufactured audio to being perceived as non-binaural audio is to render the original source with different spatial parameters or to re-render a sound source with different spatial parameters in order to adjust, degrade, or eliminate certain human audio cues [such as rendering] the sound source to specific alterations per an ITD and/or interaural level difference (ILD), or discontinue[ing] rendering or using the HRTF or ITD/ILD

calculations while continuing to render other aspects of the audio without pause.”), [0103] (“In FIGS. 4-16, example actions include, but are not limited to, one or more of ... alter a HRTF of a person [and] change or alter an ITD or an ILD”).

6. '184 Patent - Dependent claims 8-13

a. Claim 8. The method of claim 7, wherein the spatial cue includes an interaural time difference (ITD).

378. Lyren teaches “wherein the spatial cue includes an interaural time difference (ITD)” for the same reasons as Element 7[d] of this Ground for this patent.

b. Claim 9. The method of claim 7 further comprising: displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and switching a voice of the character from the binaural sound to stereo sound when the WED no longer displays the virtual image during the game.

379. The additional limitation recited in Claim 9 of the '184 Patent is nearly identical to the additional limitation of Claim 9 of the '179 Patent. Rather than triggering “when the character vanishes from the environment during the game,” however, the '184 Patent recites triggering “when the WED no longer displays the virtual image during the game.” A POSA would understand that there is no relevant distinction between the character vanishing and the character no longer being displayed by the WED. Accordingly, this additional limitation is therefore taught by Lyren and Reif for the same reasons as described for Claim 9 of Ground 2 for the '179 Patent.

- c. **Claim 10. The method of claim 7 further comprising: alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.**

380. The additional limitation recited in Claim 10 of the '184 Patent is identical to the additional limitation of Claim 10 of the '179 Patent. This additional limitation is therefore taught by Lyren and Reif for the same reasons as described for Claim 10 of Ground 2 for the '179 Patent.

- d. **Claim 11. The method of claim 7 further comprising: switching the binaural sound to stereo sound upon activation of a button on the WED.**

381. The additional limitation recited in Claim 11 of the '184 Patent is nearly identical to the additional limitation of Claim 11 of the '179 Patent. Rather than a “switch,” however, the '184 Patent recites a “button.” A POSA would understand that there is no relevant distinction between a switch and a button. Accordingly, this additional limitation is therefore taught by Lyren for the same reasons as described for Claim 11 of Ground 2 for the '179 Patent.

- e. **Claim 12. The method of claim 7 further comprising: wherein the WED includes electronic glasses, the environment is a physical room where the user is located, the virtual image is an augmented reality (AR) character that speaks to the user while the user plays a game, the WED includes head tracking, and a digital signal processor (DSP) in the WED reduces a volume of the binaural sound to alert the user when a face of the user is not directed to the virtual image in the environment.**

382. The additional limitations recited in Claim 12 of the '184 Patent are identical to the additional limitations of Claim 12 of the '179 Patent. These additional limitations are therefore taught by Lyren and Reif for the same reasons as described for Claim 12 of Ground 2 for the '179 Patent.

- f. **Claim 13. The method of claim 7 further comprising: sensing, with the WED, when a physical object interferes with a location of the virtual image in the environment; and switching the binaural sound to stereo sound in response to sensing the physical object interferes with the location of the virtual image in the environment.**

383. The additional limitations recited in Claim 13 of the '184 Patent are identical to the additional limitations of Claim 13 of the '179 Patent. These additional limitations are therefore taught by Lyren and Reif for the same reasons as described for Claim 13 of Ground 2 for the '179 Patent.

7. '184 Patent - Independent Claim 1

a. Elements 1[pre] to 1[c]

384. Counsel instructed me to assume this preamble was limiting. Elements 1[pre] to 1[c] are taught by Lyren for the same reasons as Elements 1[pre] to 1[c] of Ground 1 for the '179 Patent.

385. Elements 1[pre] to 1[c] in the '184 Patent are substantively identical to Elements 1[pre] to 1[c] in the '179 Patent. The only differences are that Element 1[a] of the '179 Patent recites “storing head-related transfer functions (HRTFs) in memory of a head mounted display (HMD) worn on a head of a user,” where the '184 Patent merely recites “storing head-related transfer functions (HRTFs),” and that Element 1[c] of the '179 Patent recites “a digital signal processor (DSP) *in the HMD*,” where the '184 Patent merely recites “a digital signal processor (DSP).” In other words, the '179 Patent is narrower than the '184 Patent because it requires storage of the HRTFs in memory at the HMD and locating the DSP at the HMD. These elements are therefore taught by Lyren for the same reasons as Elements 1[pre] to 1[c] of Ground 2 for the '179 Patent.

386. Further a POSA would understand that there is no meaningful difference between “binaural sound that *is provided* to the user and that externally localizes in empty space *to* the virtual image in the 3D virtual environment” and

“binaural sound that *plays* to the user and that externally localizes in empty space *at* the virtual image in the 3D virtual environment.”

387. Likewise, the differences between Elements 1[pre] to 1[c] and Elements 14[pre] to 14[c] of the '184 Patent do not change my analysis. Specifically, a POSA would understand that the HMD described in Element 1[b] is same as the wearable electronic device (WED) of Element 14[b] and that there is no relevant distinction between “binaural sound that plays to the user and that externally localizes in empty space at the virtual image “ in Element 1[c] and “binaural sound that externally localizes to the user at the virtual image in the environment” in Element 14[c].

- b. Element 1[d] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter a spectral cue and lower a volume of the sound playing to the user.**

388. Unlike its counterpart in the '179 Patent, Claim 1 in the '184 Patent does not recite sensing, with one or more sensors in the HMD, when the user leaves the 3D virtual environment that includes the virtual image. It instead jumps forward to an alert, without regard how the triggering event is detected.

389. Element 1[d] of the '184 Patent is nearly the same as Element 1[e] of the '179 Patent. Rather than “alter audio cues to prevent the user from experiencing external localization of the sounds,” however, the '184 Patent recites “alter a spectral cue.”

390. Spectral cues include ILDs which are also audio cues for providing localization of sounds. *See* '184 Patent, 3:10-17. Accordingly, Element 1[d] in this patent is taught or rendered obvious by Lyren and Reif for the same reasons discussed for Elements 1[d]-[e] in Ground 2 of the '179 Patent.

8. '184 Patent - Dependent Claims 2-6

- a. Claim 2. The method of claim 1 further comprising: warning the user that the user is leaving the 3D virtual environment by removing, by the DSP, background sound that occurs in the 3D virtual environment as part of a game the user is playing.**

391. The additional limitation recited in Claim 2 of the '184 Patent is nearly identical to the additional limitation of Claim 2 of the '179 Patent. Rather than “background sounds,” however, the '184 Patent recites a “background sound.” A POSA would understand that there is no relevant distinction between background sounds and a background sound. This additional limitation is therefore taught by Lyren and Reif for the same reasons as described for Claim 2 of Ground 2 for the '179 Patent.

- b. Claim 3. The method of claim 1, wherein the warning occurs when the user is leaving a virtual area where the virtual image is located while remaining in the 3D virtual environment.**

392. The additional limitation recited in Claim 3 of the '184 Patent is nearly identical to the additional limitation of Claim 3 of the '179 Patent. Rather than a “virtual room,” however, the '184 Patent recites a “virtual area where the virtual

image is located.” A POSA would understand that there is no relevant distinction between a virtual room and a virtual area. This additional limitation is therefore taught by Lyren and Reif for the same reasons as described for Claim 3 of Ground 2 for the ’179 Patent.

- c. **Claim 4. The method of claim 1 further comprising: tracking, with head tracking in the HMD, head movements of the user; and switching the sound from the binaural sound to stereo sound when a face of the user is directed away from the virtual image such that the virtual image is no longer visible to the user.**

393. The additional limitations recited in Claim 4 of the ’184 Patent are nearly identical to the additional limitations of Claim 4 of the ’179 Patent. Rather than a “when a face of the user is directed away from the virtual image,” however, the ’184 Patent recites a “when a face of the user is directed away from the virtual image such that the virtual image is no longer visible to the user.” This difference does not change the analysis as in both instances the user has directed their gaze away from the virtual image. These additional limitations are therefore taught by Lyren and Reif for the same reasons as described for Claim 4 of Ground 2 for the ’179 Patent.

- d. Claim 5. The method of claim 1 further comprising: switching the sound from the binaural sound to stereo sound when the user activates a button located on the HMD.**

394. The additional limitation recited in Claim 5 of the '184 Patent is nearly identical to the additional limitation of Claim 5 of the '179 Patent. Rather than a “physical switch,” however, the '184 Patent recites a “button.” A POSA would understand that there is no relevant distinction between a physical switch and a button. Instead, a POSA would have understood that a button could and would be replaced with a switch, and, indeed, Lyren teaches either a “switch or button.” Lyren, [0238]. This limitation is therefore taught by Lyren for the same reasons as described for Claim 5 of Ground 2 for the '179 Patent.

- e. Claim 6. The method of claim 1 further comprising: switching the sound from the binaural sound to stereo sound when the virtual image vanishes from the 3D virtual environment.**

395. The additional limitation recited in Claim 6 of the '184 Patent is nearly identical to the additional limitation of Claim 6 of the '179 Patent. Rather than “sounds,” however, the '184 Patent recites a “sound.” A POSA would understand that for the same reasons Clement with Osman and Sauk renders obvious switching “sounds,” the combination also renders obvious switching the “sound.” This additional limitation is therefore taught by Lyren and Reif for the same reasons as described for Claim 6 of Ground 2 for the '179 Patent.

XI. Conclusion

396. I have provided these opinions based on the information currently available to me. I reserve the right to expand or modify my opinions as this proceeding continues and as my investigation and study continues. I may also supplement my opinions in response to any additional information that becomes available to me, including opinions provided by Patent Owner or its expert(s).

397. 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.

Executed on June 6th, 2023 in New York, NY.

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

Dr. Agnieszka Roginska, Ph.D.

Declaration for *Inter Partes Review* of USP 10,368,179 and 10,448,184

APPENDIX

The following listing of the challenged claims adds letters to the claim limitations for clarity:

'179 Patent Claims

1[pre] A method comprising:

[a] storing head-related transfer functions (HRTFs) in memory of a head mounted display (HMD) worn on a head of a user;

[b] displaying, with a display of the HMD worn on the head of the user, a three-dimensional (3D) virtual environment that includes a virtual image;

[c] processing, by a digital signal processor (DSP) in the HMD, sounds with the HRTFs to generate binaural sound that is provided to the user and that externally localizes in empty space to the virtual image in the 3D virtual environment;

[d] sensing, with one or more sensors in the HMD, when the user leaves the 3D virtual environment that includes the virtual image; and

[e] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter audio cues to prevent the user from experiencing external localization of the sounds and by lowering a volume of the sounds being provided to the user.

2. The method of claim 1 further comprising:

warning the user that the user is leaving the 3D virtual environment by removing, by the DSP, background sounds that occur in the 3D virtual environment as part of a game the user is playing.

3. The method of claim 1, wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment.

4. The method of claim 1 further comprising:

tracking, with head tracking in the HMD, head movements of the user;
and

switching the sounds from the binaural sound to stereo sound when a face of the user is directed away from the virtual image.

5. The method of claim 1 further comprising:

switching the sounds from the binaural sound to stereo sound when the user activates a physical switch located in the HMD.

6. The method of claim 1 further comprising:

switching the sounds from the binaural sound to stereo sound when the virtual image vanishes in the 3D virtual environment.

7[pre] A method executed by one or more electronic devices, the method comprising:

[a] storing head-related transfer functions (HRTFs) in memory;

[b] displaying, by a display of a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

[c] processing, by a digital signal processor (DSP), sounds with the HRTFs to generate binaural sound in empty space that is provided to the user and that externally localizes to the empty space and to the virtual image in the environment; and

[d] determining, with the WED, when the user is leaving the environment; and

[e] alerting the user that the user is leaving the environment by altering processing of the binaural sound to degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound.

8. The method of claim 7 further comprising:

alerting the user that the user is leaving the environment by reducing a volume of the binaural sound being provided to the user.

9. The method of claim 7 further comprising:

displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and

switching a voice of the character from the binaural sound to stereo sound when the character vanishes from the environment during the game.

10. The method of claim 7 further comprising:

alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.

11. The method of claim 7 further comprising:

switching the binaural sound to stereo sound upon activation of a switch on the WED.

12. The method of claim 7 further comprising:

wherein the WED includes electronic glasses, the environment is a physical room where the user is located, the virtual image is an augmented reality (AR) character that speaks to the user while the user plays a game, the WED includes head tracking, and a digital signal processor (DSP) in the WED reduces a volume of the binaural sound to alert the user when a face of the user is not directed to the virtual image in the environment.

13. The method of claim 7 further comprising:

sensing, with the WED, when a physical object interferes with a location of the virtual image in the environment; and

switching the binaural sound to stereo sound in response to sensing the physical object interferes with the location of the virtual image in the environment.

14[pre] A method executed by one or more electronic devices, the method comprising:

[a] storing head-related transfer functions (HRTFs) in memory;

[b] displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

[c] processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes in empty space from the virtual image in the environment;

[d] determining, with the WED, when the user is leaving the environment; and

[e] alerting the user that the user is leaving the environment by reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user.

15. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by degrading the binaural sound being provided to the user as the user is leaving the environment.

16. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by altering, by the DSP, processing of the binaural sound so the user perceives less external localization.

17. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.

18. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by eliminating audio cues in the binaural sound being provided to the user.

19. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to stereo sound.

20. The method of claim 14 further comprising:

detecting, with a microphone in the WED, a voice of a person in the environment of the user; and

switching the binaural sound to stereo sound in response to detecting the voice.

'184 Patent Claims

1[pre] A method comprising:

[a] storing head-related transfer functions (HRTFs);

[b] displaying, with a display of a head mounted display (HMD) worn on a head of a user, a three-dimensional (3D) virtual environment that includes a virtual image;

[c] processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that plays to the user and that externally localizes in empty space at the virtual image in the 3D virtual environment; and

[d] warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter a spectral cue and lower a volume of the sound playing to the user.

2. The method of claim 1 further comprising:

warning the user that the user is leaving the 3D virtual environment by removing, by the DSP, background sound that occurs in the 3D virtual environment as part of a game the user is playing.

3. The method of claim 1, wherein the warning occurs when the user is leaving a virtual area where the virtual image is located while remaining in the 3D virtual environment.

4. The method of claim 1 further comprising:

tracking, with head tracking in the HMD, head movements of the user; and

switching the sound from the binaural sound to stereo sound when a face of the user is directed away from the virtual image such that the virtual image is no longer visible to the user.

5. The method of claim 1 further comprising:

switching the sound from the binaural sound to stereo sound when the user activates a button located on the HMD.

6. The method of claim 1 further comprising:

switching the sound from the binaural sound to stereo sound when the virtual image vanishes from the 3D virtual environment.

7[pre] A method executed by one or more electronic devices, the method comprising:

[a] storing head-related transfer functions (HRTFs);

[b] displaying, by a display of a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

[c] processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that plays to the user and that externally localizes at the virtual image in the environment; and

[d] alerting the user that the user is leaving the environment by altering processing of the sound by the DSP to degrade a spatial cue and to lower a volume of the sound playing to the user.

8. The method of claim 7, wherein the spatial cue includes an interaural time difference (ITD).

9. The method of claim 7 further comprising:

displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and

switching a voice of the character from the binaural sound to stereo sound when the WED no longer displays the virtual image during the game.

10. The method of claim 7 further comprising:

alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.

11. The method of claim 7 further comprising:

switching the binaural sound to stereo sound upon activation of a button on the WED.

12. The method of claim 7 further comprising:

wherein the WED includes electronic glasses, the environment is a physical room where the user is located, the virtual image is an augmented reality (AR) character that speaks to the user while the user plays a game, the WED includes head tracking, and a digital signal processor (DSP) in the WED reduces a volume of the binaural sound to alert the user when a face of the user is not directed to the virtual image in the environment.

13. The method of claim 7 further comprising:

sensing, with the WED, when a physical object interferes with a location of the virtual image in the environment; and

switching the binaural sound to stereo sound in response to sensing the physical object interferes with the location of the virtual image in the environment.

14. A method executed by one or more electronic devices, the method comprising:

[a] storing head-related transfer functions (HRTFs);

[b] displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

[c] processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes to the user at the virtual image in the environment; and

[d] alerting the user that the user is leaving the virtual image in the environment by altering processing of the binaural sound by the DSP and reducing a volume of the binaural sound that localizes to the user at the virtual image.

15. The method of claim 14, wherein the DSP alters processing of the binaural sound to alert the user by changing a spatial cue that includes at least one

of an interaural time difference (ITD), an interaural level difference (ILD), and the HRTFs.

16. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by stopping, by the DSP, the processing of the sound with the HRTFs.

17. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.

18. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by eliminating one or more spatial cues in the binaural sound being provided to the user.

19. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to one of stereo sound and mono sound.

20. The method of claim 14 further comprising:

detecting, with a microphone in the WED, a voice of a person in the environment of the user; and

switching the binaural sound to stereo sound in response to detecting the voice.