

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re *Inter Partes* Review of:)
U.S. Patent No. 9,226,090)
Issued: Dec. 29, 2018)
Application No.: 14/311,532)
Filing Date: June 23, 2014)

For: **Sound Localization for an Electronic Call**

FILED VIA P-TACTS

**DECLARATION OF DR. CHRIS KYRIAKAKIS IN SUPPORT OF
PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 9,226,090**

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

TABLE OF CONTENTS

I.	INTRODUCTION.....	13
II.	EXPERIENCE, QUALIFICATIONS, AND STATUS AS AN INDEPENDENT EXPERT WITNESS	14
III.	UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES	19
A.	Claim Construction.....	20
B.	Obviousness	20
IV.	SUMMARY OF MATERIALS REVIEWED AND CONSIDERED.....	22
V.	LEVEL OF SKILL IN THE ART AND PERSPECTIVE APPLIED IN THIS DECLARATION	25
VI.	BACKGROUND.....	27
A.	Overview of the Technology	27
1.	Wearable virtual reality	27
2.	Binaural sound.....	30
3.	Sound source localization.....	35
4.	Creating binaural sound.....	39
5.	Compensating for head movement.....	42
6.	Virtual conferencing and hybrid collaboration environments.....	47
B.	The '090 Patent.....	50
C.	Prosecution History	53
D.	Challenged Claims of the '090 Patent	54
E.	Effective Priority Date	55
VII.	CLAIM CONSTRUCTION	55
VIII.	THE PRIOR ART	56
A.	Bedingfield.....	56
B.	Li	58
C.	Andrea.....	59
D.	Ballas	60
E.	Sandgren	60

IX.	GROUND 1: BEDINGFIELD IN VIEW OF ANDREA RENDERS OBVIOUS CLAIMS 1-3, 6, 15-17	61
A.	Motivation to Combine.....	61
B.	Reasonable Expectation of Success.....	65
C.	Independent Claims 1	66
1.	1[pre]: “ <i>A method, comprising:</i> ”	66
2.	1[a]: “ <i>capturing, with an electronic earphone located at a head of a talking person, binaural sound that will be provided to a listening person during a telephone call;</i> ”	68
3.	1[b]: “ <i>designating, with a computer system, a sound localization point in empty space that is away from and proximate to the listening person such that the sound localization point simulates an origin of the binaural sound at the empty space that the listening person hears during the telephone call;</i> ”	75
4.	1[c]: “ <i>adjusting, with the computer system, the binaural sound captured at the earphone of the talking person so the binaural sound originates during the telephone call from the sound localization point in empty space that is away from and proximate to the listening person; and</i> ”	81
5.	1[d]: “ <i>providing, with an electronic earphone located at a head of the listening person, the binaural sound to the listening person during the telephone call such that the origin of the binaural sound for the listening person occurs at the sound localization point in empty space that is away from and proximate to the listening person.</i> ”	82
D.	Dependent Claims 2, 3, 6.....	83
1.	Claim 2[a]: “ <i>The method of claim 1 further comprising: tracking a head orientation of the listening person with respect to the sound localization point;</i> ”	83
2.	Claim 2[b]: “ <i>adjusting an interaural time difference of the sound that the listening person hears in response to the listening person changing the head orientation with respect to the sound localization point.</i> ”	85

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

3.	Claim 3: “ <i>The method of claim 1 further comprising: receiving, from the listening person, a location that specifies a distance from the listening person where to position the sound localization point in empty space that is proximate to and away from the listening person.</i> ”	86
4.	Claim 6[a]: “ <i>The method of claim 1 further comprising: capturing binaural sound from a second talking person;</i> ”	86
5.	Claim 6[b]: “ <i>designating a second sound localization point in empty space that is away from and proximate to the listening person such that the second sound localization point simulates an origin of the binaural sound from the second talking person that the listener hears;</i> ”	87
6.	Claim 6[c]: “ <i>positioning the second sound localization point adjacent to the sound localization point such that the binaural sound captured from the talking person appears to originate from the sound localization point and the binaural sound captured from the second talking person appears to originate from the second sound localization point.</i> ”	87
E.	Independent Claim 15	88
1.	15[pre]: “ <i>A computer system, comprising:</i> ”	88
2.	15[a]: “ <i>a first wearable electronic device that a first person wears;</i> ”	89
3.	15[b]: “ <i>a second wearable electronic device that a second person wears and that captures a binaural voice of the second person during an electronic phone call with the first wearable electronic device that the first person wears;</i> ”	89
4.	15[c]: “ <i>a memory that stores instructions; and</i> ”	91
5.	15[d]: “ <i>a processor that executes the instructions to...</i> ” ...	92
6.	15[d1]: “ <i>adjust the binaural voice of the second person during the electronic phone call in order to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person,</i> ”	93

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

7.	15[d2]: “ <i>wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.</i> ”	93
F.	Dependent Claims 16-17	94
1.	Claim 16: “ <i>The computer system of claim 15, wherein the processor further executes the instructions to determine a location of an empty chair that is proximate to the first person and position the sound localization point at the empty chair.</i> ”	94
2.	Claim 17: “ <i>The computer system of claim 15, wherein the processor further executes the instructions to change an interaural time difference (ITD), an interaural level difference (ILD), and a head-related transfer function (HRTF) in order to position the sound localization point of the binaural voice to the location that is away from and proximate to the first person.</i> ”	96
X.	GROUND 2: BEDINGFIELD IN VIEW OF ANDREA AND BALLAS RENDERS OBVIOUS CLAIM 18	98
A.	Motivation to Combine Bedingfield and Andrea with Ballas	98
B.	Reasonable Expectation of Success.....	100
C.	Dependent Claim 18	101
1.	Claim 18: “ <i>The computer system of claim 15, wherein the processor further executes the instructions to change an interaural time difference, an interaural level difference, and a head-related transfer function of the binaural sound in order to maintain the sound localization point fixed at the location during the electronic phone call while the first person physically moves away from the location and changes a head orientation with respect to the location.</i> ”	101
XI.	GROUND 3: BEDINGFIELD IN VIEW OF ANDREA AND SANDGREN RENDERS OBVIOUS CLAIM 19	105
A.	Motivation to Combine.....	105
B.	Reasonable Expectation of Success.....	108
C.	Dependent Claim 19	109

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

1. Claim 19[a]: “*The computer system of claim 15, wherein the processor further executes the instructions to change attenuation and reverberation of the binaural sound in order to add artificial audial cues of virtual objects near the second person,*” 109
2. Claim 19[b]: “*wherein the virtual objects do not actually exist near the second person but are perceived to exist by the first person from the artificial audial cues.*” 111

XII. GROUND 4: BEDINGFIELD IN VIEW OF BALLAS RENDERS
OBVIOUS CLAIMS 8-10, 12, 14 112

- A. Motivation to Combine..... 112
- B. Reasonable Expectation of Success..... 112
- C. Independent Claim 8..... 113
 1. 8[Pre]: “*A non-transitory computer readable storage medium storing instructions that cause a computer system to execute a method, comprising:*” 113
 2. 8[a]: “*commence an electronic call between a first individual wearing an electronic earphone and a second individual that are remote from each other;*” 113
 3. 8[b]: “*designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;*” 113
 4. 8[c]: “*execute the electronic call such that the origin of the voice of the second individual appears to the first individual to originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and*” 114
 5. 8[d]: “*maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first*

	<i>individual hears in response to the first individual moving the head orientation with respect to the sound localization point.”</i>	115
D.	Dependent Claims 9, 10, 12, 14.....	116
1.	Claim 9: “ <i>The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: maintain the origin of the voice of the second individual at the sound localization point but alter an interaural level difference of the voice that the first individual hears in response to the first individual moving closer to and away from the sound localization point.</i> ”	116
2.	Claim 10[a]: “ <i>The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: monitor movements of the head orientation of the first individual in response to the first individual localizing the voice of the second individual during the electronic call;</i> ”	117
3.	Claim 10[b]: “ <i>adjust head-related transfer functions (HRTFs) of the first individual to more accurately simulate real HRTFs of the first individual in response to the movements of the head orientation of the first individual localizing the voice of the second individual during the electronic call.</i> ”	117
4.	Claim 12[a]: “ <i>The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: add a third individual to the electronic call;</i> ”	118
5.	Claim 12[b]: “ <i>designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the third individual occurs for the first individual through the electronic earphone such that the sound localization point of the third individual is adjacent to the sound localization point of the second individual.</i> ”	119

6. Claim 14: “*The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: receive, from the first individual, commands that move the sound localization point from a first point in empty space that is away from and proximate to the first individual to a second point in empty space that is away from and proximate to the first individual.*” 119

XIII. GROUND 5: LI IN VIEW OF ANDREA RENDERS OBVIOUS
CLAIMS 1, 3, 6, 15-17 121

- A. Motivation to Combine 121
- B. Reasonable Expectation of Success 122
- C. Independent Claim 1 124
1. 1[pre]: “*A method, comprising:*” 124
2. 1[a]: “*capturing, with an electronic earphone located at a head of a talking person, binaural sound that will be provided to a listening person during a telephone call*” 124
3. 1[b]: “*designating, with a computer system, a sound localization point in empty space that is away from and proximate to the listening person such that the sound localization point simulates an origin of the binaural sound at the empty space that the listening person hears during the telephone call*” 127
4. 1[c]: “*adjusting, with the computer system, the binaural sound captured at the earphone of the talking person so the binaural sound originates during the telephone call from the sound localization point in empty space that is away from and proximate to the listening person; and*” 132
5. 1[d]: “*providing, with an electronic earphone located at a head of the listening person, the binaural sound to the listening person during the telephone call such that the origin of the binaural sound for the listening person occurs at the sound localization point in empty space*” 132

	<i>that is away from and proximate to the listening person.”</i>	134
D.	Dependent Claims 3, 6.....	134
1.	Claim 3: “ <i>The method of claim 1 further comprising: receiving, from the listening person, a location that specifies a distance from the listening person where to position the sound localization point in empty space that is proximate to and away from the listening person.</i> ”	134
2.	Claim 6[a]: “ <i>The method of claim 1 further comprising: capturing binaural sound from a second talking person;</i> ”	135
3.	Claim 6[b]: “ <i>designating a second sound localization point in empty space that is away from and proximate to the listening person such that the second sound localization point simulates an origin of the binaural sound from the second talking person that the listener hears;</i> ”	136
4.	Claim 6[c]: “ <i>positioning the second sound localization point adjacent to the sound localization point such that the binaural sound captured from the talking person appears to originate from the sound localization point and the binaural sound captured from the second talking person appears to originate from the second sound localization point.</i> ”	136
E.	Independent Claim 15:.....	137
1.	15[Pre]: “ <i>A computer system, comprising:</i> ”.....	137
2.	15[a]: “ <i>a first wearable electronic device that a first person wears;</i> ”	137
3.	15[b]: “ <i>a second wearable electronic device that a second person wears and that captures a binaural voice of the second person during an electronic phone call with the first wearable electronic device that the first person wears;</i> ”	137
4.	15[c]: “ <i>a memory that stores instructions; and</i> ”	138
5.	15[d]: “ <i>a processor that executes the instructions to...</i> ” .	138
6.	15[d1]: “ <i>adjust the binaural voice of the second person during the electronic phone call in order to position a sound localization point of the binaural voice to a</i>	

	<i>location that is away from and proximate to the first person,”</i>	139
7.	15[d2]: “ <i>wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.</i> ”	139
F.	Dependent Claims 16-17	140
1.	Claim 16: “ <i>The computer system of claim 15, wherein the processor further executes the instructions to determine a location of an empty chair that is proximate to the first person and position the sound localization point at the empty chair.</i> ”	140
2.	Claim 17: “ <i>The computer system of claim 15, wherein the processor further executes the instructions to change an interaural time difference (ITD), an interaural level difference (ILD), and a head-related transfer function (HRTF) in order to position the sound localization point of the binaural voice to the location that is away from and proximate to the first person.</i> ”	141
XIV.	GROUND 6: LI IN VIEW OF ANDREA AND BALLAS RENDERS OBVIOUS CLAIMS 2, 18	143
A.	Motivation to Combine Li and Andrea with Ballas	143
B.	Reasonable Expectation of Success.....	145
C.	Dependent Claim 2	147
1.	Claim 2[a]: “ <i>The method of claim 1 further comprising: tracking a head orientation of the listening person with respect to the sound localization point;</i> ”	147
2.	Claim 2[b]: “ <i>adjusting an interaural time difference of the sound that the listening person hears in response to the listening person changing the head orientation with respect to the sound localization point.</i> ”	147
D.	Dependent Claim 18	148
1.	Claim 18: “ <i>The computer system of claim 15, wherein the processor further executes the instructions to change an interaural time difference, an interaural level difference, and a head-related transfer function of the binaural sound in order to maintain the sound localization point fixed at the location during the</i>	

	<i>electronic phone call while the first person physically moves away from the location and changes a head orientation with respect to the location.”</i>	148
XV.	GROUND 7: LI IN VIEW OF ANDREA AND SANDGREN RENDERS OBVIOUS CLAIM 19.....	149
A.	Motivation to Combine.....	149
B.	Reasonable Expectation of Success.....	152
C.	Dependent Claim 19	153
1.	Claim 19[a]: “ <i>The computer system of claim 15, wherein the processor further executes the instructions to change attenuation and reverberation of the binaural sound in order to add artificial audial cues of virtual objects near the second person,</i> ”	153
2.	Claim 19[b]: “ <i>wherein the virtual objects do not actually exist near the second person but are perceived to exist by the first person from the artificial audial cues.</i> ”	153
XVI.	GROUND 8: LI IN VIEW OF BALLAS RENDERS OBVIOUS CLAIMS 8-10, 12, 14	153
A.	Motivation to Combine.....	153
B.	Reasonable Expectation of Success.....	154
C.	Independent Claim 8.....	154
1.	8[pre]: “ <i>A non-transitory computer readable storage medium storing instructions that cause a computer system to execute a method, comprising:</i> ”	154
2.	8[a]: “ <i>commence an electronic call between a first individual wearing an electronic earphone and a second individual that are remote from each other;</i> ”	154
3.	8[b]: “ <i>designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;</i> ”	155

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

4.	8[c]: “execute the electronic call such that the origin of the voice of the second individual appears to the first individual to originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and”	155
5.	8[d]: “maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first individual hears in response to the first individual moving the head orientation with respect to the sound localization point.”	156
D.	Dependent Claims 9, 10, 12, 14.....	158
1.	Claim 9: “The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: maintain the origin of the voice of the second individual at the sound localization point but alter an interaural level difference of the voice that the first individual hears in response to the first individual moving closer to and away from the sound localization point.”	158
2.	Claim 10	159
a.	10[a]: “The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: monitor movements of the head orientation of the first individual in response to the first individual localizing the voice of the second individual during the electronic call;”	159
b.	10[b]: “adjust head-related transfer functions (HRTFs) of the first individual to more accurately simulate real HRTFs of the first individual in response to the movements of the head orientation of the first individual localizing the voice of the second individual during the electronic call.”	159
3.	Claim 12	161

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

a.	12[a]: “ <i>The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: add a third individual to the electronic call;</i> ”	161
b.	12[b]: “ <i>designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the third individual occurs for the first individual through the electronic earphone such that the sound localization point of the third individual is adjacent to the sound localization point of the second individual.</i> ”	161
4.	Claim 14: “ <i>The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: receive, from the first individual, commands that move the sound localization point from a first point in empty space that is away from and proximate to the first individual to a second point in empty space that is away from and proximate to the first individual.</i> ”	162
XVII. SECONDARY CONSIDERATIONS		164
XVIII. CONCLUSION.....		164

I. INTRODUCTION

1. I have been retained by Meta Platforms, Inc. (“Petitioner”) to provide my opinions concerning the validity of claims 1-3, 6, 8-10, 12, 14-19 of U.S. Patent No. 9,226,090.

2. I am simultaneously engaged by Petitioner to provide declarations for IPR petitions challenging U.S. Patent Nos. 9,282,196 and U.S. Patent No. 9,674,628.

3. This Declaration summarizes the opinions I have formed to date. I reserve the right to modify my opinions, if necessary, based on further review and analysis of information that I receive subsequent to the filing of this report, including in response to positions taken by Eight KHZ, LLC (“8KHZ”) or its experts that I have not yet seen, including any secondary consideration evidence that 8KHZ or its expert may consider and present.

4. It is my opinion that claims 1-3, 6, 8-10, 12, 14-15, and 17-19 of the ’090 patent are invalid based on the following grounds:

Ground 1	Claims 1-3, 6, 15-17 are obvious under 35 U.S.C. § 103 over Bedingfield in view of Andrea.
Ground 2	Claim 18 is obvious under 35 U.S.C. § 103 over Bedingfield in view of Andrea and Ballas.
Ground 3	Claim 19 is obvious under 35 U.S.C. § 103 over Bedingfield in view of Andrea and Sandgren.
Ground 4	Claims 8-10, 12, 14 are obvious under 35 U.S.C. § 103 over Bedingfield in view of Ballas.
Ground 5	Claims 1, 3, 6, 15-17 are obvious under 35 U.S.C. § 103 over Li in view of Andrea.

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

Ground 6	Claims 2, 18 are obvious under 35 U.S.C. § 103 over Li in view of Andrea and Ballas.
Ground 7	Claim 19 is obvious under 35 U.S.C. § 103 over Li in view of Andrea and Sandgren.
Ground 8	Claims 8-10, 12, 14 are obvious under 35 U.S.C. § 103 over Li in view of Ballas.

II. EXPERIENCE, QUALIFICATIONS, AND STATUS AS AN INDEPENDENT EXPERT WITNESS

5. My curriculum vitae, which a more detailed summary of my background, experience, publications that I have authored in the past ten years, and qualifications, is attached to the accompanying Petition as Ex. 1004.

6. I earned my Bachelor of Science degree in Engineering and Applied Science from the California Institute of Technology (Caltech) in 1985. I received my Master of Science degree in Electrical Engineering in 1987 and my Ph.D. in Electrical Engineering in 1993, both from the University of Southern California (“USC”).

7. I am a tenured Associate Professor in the Department of Electrical and Computer Engineering at USC.

8. My research interests lie at the intersection of acoustics, psychoacoustics (the science that studies human perception of sound), and audio signal processing. I have worked in the area of audio signal processing for more than 20 years with particular emphasis on spatial and 3D audio. I have published

nearly 100 technical papers, including several peer reviewed papers. I have published a book on spatial audio entitled *Immersive Audio Signal Processing* and hold several patents in audio processing for headphone optimization, acoustic measurement of loudspeakers in rooms, loudspeaker crossover optimization, and loudspeaker response correction using signal processing. My publications examine various aspects of sound measurement, echo cancellation, microphone arrays, noise removal, filter banks, how sound interacts with the acoustical elements of the environment, novel methods for surround sound recording and reproduction, and the perception of sound by human listeners.

9. A major focus of my early research was on head-related transfer functions (HRTFs) and how they can be measured and modeled using various signal processing approaches. I developed several implementations of HRTF-based 3D audio for rendering spatial sounds over headphones. I have also researched microphone arrays for applications in enhanced far field voice recognition. This includes methods for Direction of Arrival finding and beam forming from linear and circular microphone arrays. I have also investigated methods for noise removal and reverberation reduction from microphone signals. In addition, over the past few years year I have worked on methods for echo cancellation for applications in wearable devices. Previous other work includes loudspeakers and how they interact with their acoustical environment. In particular, I have researched the role of sound

reflections, absorption, and diffusion in the performance of loudspeakers. I have also published several technical papers on acoustical measurement methods in rooms and device enclosures and developed novel signal processing algorithms for optimizing sound system performance. Other topics I have researched include multichannel audio acquisition and rendering, virtual microphones and virtual speakers, hybrid headphone-loudspeaker rendering methods, and advanced signal processing techniques for optimizing sound quality from small portable devices.

10. I am the founding Director of the USC Immersive Audio Laboratory with facilities for experimental work in room acoustics, multichannel audio, and psychoacoustics. This laboratory also serves as a unique teaching facility for my undergraduate course in Introduction to Digital Audio and my graduate course in Immersive Audio Signal Processing. Both courses have a major acoustics component that examines the interaction of sound with the acoustical environment and the basic principles of microphones and loudspeaker design. The graduate course was developed through a two-year grant I received from the National Science Foundation entitled “Collaborative Learning in Engineering Using Immersive Environments,” and was the first of its kind to assess the impact of audio immersion in student learning. In addition to the courses I teach, I have also supervised and served on Ph.D. dissertation committees for more than 30 students.

11. From 2003-2018, I was the co-founder and Chief Technology Officer of Audyssey Laboratories, a USC spin-off company that develops and licenses audio technology to leading consumer electronics and automotive companies around the world. As part of my work at Audyssey, I developed a measurement methodology for headphones and earphones and created a database with 500+ measurements of products from nearly 100 manufacturers, created audio algorithms related to sound optimization that are used in millions of home theater systems, cars, and IMAX theaters world-wide.

12. I also developed audio algorithms and designed speakers. These speakers were novel acoustical designs that used a combination of unique enclosures, passive radiators (i.e., speaker diaphragms that are driven by changes in surrounding air pressure), and audio signal processing to optimize their performance and overcome limitations that arise from small drivers and enclosures. For example, we used signal processing technologies combined with novel acoustical design to extend the bass response of small woofers and passive radiators beyond what was previously possible in small speaker enclosures. The innovations in these designs have received awards, including Popular Science's "Best of What's New."

13. In 2019, I joined Syng as the Chief Audio Scientist. Syng creates innovative products for next generation spatial audio and my role there includes research and development of audio algorithms for sound capture using microphone

arrays to measure the acoustics of the room and loudspeaker beamforming for rendering sound in various directions.

14. In April 2018 I was elected as Senior Member of the Institute of Electrical and Electronic Engineers (IEEE) in recognition of my contributions to the field of engineering. I am also a member of the Audio Engineering Society (AES), an association for professionals in the audio industry.

15. In 2006, I received a World Technology Network Award. This organization presents awards to innovators in several areas in which technology can foster a paradigm change. My award was for innovations in immersive audio that enable new capabilities in media and journalism. Other award recipients at that event included Vice President Al Gore, Google, and Space-X. In addition, my work has been featured in multiple news articles, including pieces by [The Atlantic](#), the [BBC World Service](#), the [New York Times](#), and [National Public Radio](#).

16. In the late 1990s and early 2000s, I was a faculty researcher and later Deputy Director of the National Science Foundation's engineering research center established at USC. I studied the fundamental and technological limitations of immersive audio and the role of acoustics on the performance of microphones and loudspeakers. In 2003, together with one of my graduate students, I received the award for Best Paper at the Institute of Electrical and Electronics Engineers ("IEEE") Conference on Signals, Systems and Computers.

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

17. I am being compensated at the rate of \$525 per hour for my work on this declaration. My fee is not contingent on the outcome of any matter or on any of the technical positions I explain in this declaration.

18. I have no financial interest in Petitioner. I have been informed that 8KHZ claims ownership in the '628 patent. I have no financial interest in 8KHZ.

III. UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES

19. I understand that statutory and judicially created standards must be considered to determine the validity of a patent claim. I am not an attorney and, consequently, will offer no opinion on the law itself. My understanding of the pertinent law is described in this section and is the result of explanations provided by counsel. I have applied this understanding in my analysis.

20. I understand that a patent claim is unpatentable if it is obvious in view of the prior art. I further understand that the frame of reference for determining whether a claim is obvious is from the perspective of a person of ordinary skill in the relevant art ("POSA") at the time of the invention. I have been asked to assume that the effective priority date is June 23, 2014. I have not been asked to determine whether this date is appropriate.

A. Claim Construction

21. I understand that claim construction in an *Inter Partes* Review is controlled by the standard in *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc).

22. For example, claim construction begins with the language of the claims. *Id.*, 1312-14. The words are given their “ordinary and customary meaning,” which is “the meaning that the term would have to a person of ordinary skill in the art in question at the time of the invention, i.e., as of the effective filing date of the patent application.” *Id.* The specification and prosecution history are sources of intrinsic evidence that informs the meaning of the claim language. Extrinsic evidence, such as expert testimony and dictionaries, may be useful in educating the court regarding the field of the invention or helping determine what a person of ordinary skill in the art (“POSA”) would understand claim terms to mean. *Id.* 1318-19. However, extrinsic evidence in general is viewed as less reliable than intrinsic evidence. *Id.*

23. I understand that in this proceeding, any claim term that is not construed should be given its plain and ordinary meaning.

B. Obviousness

24. I understand that a prior art reference can render a patent claim obvious to one of ordinary skill in the art if the differences between the subject matter set forth in the patent claim and the prior art are such that the subject matter of the claim

would have been obvious at the time the claimed invention was made. In analyzing obviousness, I understand that it is important to consider the scope of the claims, the level of skill in the relevant art, the scope and content of the prior art, the differences between the prior art and the claims, and any secondary considerations.

25. I understand that when the claimed subject matter involves combining pre-existing elements to yield no more than one would expect from such an arrangement, the combination is obvious. I also understand that in assessing whether a claim is obvious one must consider whether the claimed improvement is more than the predictable use of prior art elements according to their established functions. I understand that there need not be a precise teaching in the prior art directed to the specific subject matter of a claim because one can take account of the inferences and creative steps that a person of skill in the art would employ. I further understand that a person of ordinary skill is a person of ordinary creativity, not an automation.

26. 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 control the inquiry. Any need or problem known in the field at the time of the invention and addressed by the patent can provide a reason for combining elements. For example, I understand that it is important to consider whether there existed at the time of the invention a known problem for which there was an obvious

solution encompassed by the patent's claims. I understand that known techniques can have obvious uses beyond their primary purposes, and that in many cases a person of ordinary skill can fit the teachings of multiple pieces of prior art together like pieces of a puzzle.

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

28. I understand that secondary considerations can include, for example, evidence of commercial success of the invention, evidence of a long-felt need that was solved by an invention, evidence that others copied an invention, or evidence that an invention achieved a surprising result. I further understand that such evidence must have a nexus, or causal relationship to the elements of a claim, in an order to be relevant. I am unaware of any such secondary considerations for the '090 patent.

IV. SUMMARY OF MATERIALS REVIEWED AND CONSIDERED

29. All of the opinions contained in this Declaration are based on the documents I reviewed and my knowledge and professional judgment. In forming

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

the opinions expressed in this Declaration, while drawing on my experience in the field of audio signal processing and sound localization (including my teaching, research, and consulting experiences), I reviewed the '090 patent, its file history, the prior art, and all exhibits cited in this declaration, including the documents in the table below.

Exhibit	Description
1001	U.S. Patent No. 9,226,090 (“’090 patent”)
1002	Prosecution History of U.S. Patent No. 9,226,090 (“’090 FH”)
1004	<i>Curriculum Vitae</i> of Dr. Kyriakakis
1005	U.S. Patent No. 8,406,439 to Bedingfield et al. (“Bedingfield”)
1006	U.S. Patent No. 9,131,305 to Li et al. (“Li”)
1007	U.S. Patent No. 8,818,000 to Andrea (“Andrea”)
1008	U.S. Patent Pub. No. 2003/0059070 to Ballas (“Ballas”)
1009	U.S. Patent No. 7,606,375 to Bailey, et al. (“Bailey”)
1013	U.S. Patent App. Pub. 2014/0064526 to Otto, et al. ("Otto")
1015	Rumsey, Francis "Whose head is it anyway? Optimizing binaural audio." J. Audio Eng. Soc., Vol. 59, No. 9, 2011 ("Rumsey")
1017	U.S. Patent App. Pub. 2008/0273708 to Sandgren, et al. ("Sandgren")
1020	Scheinerman, Matt “Exploring Augmented Reality.” Haverford College Computer Science (2009) (“Scheinerman”)
1021	Comeau, Charles et al. “Headsight TV System.” Electronics, 1961, pp. 86-90 (“Comeau”)

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

Exhibit	Description
1022	Sutherland, Ivan “The Ultimate Display.” Proceedings of IFIP Congress, pp. 506-508 (1965) (“Sutherland1965”)
1023	Sutherland, Ivan “A head-mounted three dimensional display.” Fall Joint Computer Conference, pp. 757-764 (1968) (“Sutherland1968”)
1024	Moller, Henrik, et al. “Head-Related Transfer Functions of Human Subjects.” J. Audio Eng. Soc., Vol. 43, No. 5, 1995, pp. 300-321 (“Moller”)
1025	Akeroyd, Michael. “The psychoacoustics of binaural hearing.” International Journal of Audiology, 45:sup1, 25-33 (2006) (“Akeroyd”)
1026	Middlebrooks, John, et al. “Sound Localization by Human Listeners.” Annual Review of Psychology, 1991. 42:135-39 (“Middlebrooks”)
1031	Cheng, Corey, et al. “Moving Sound Source Synthesis for Binaural Electroacoustic Music Using Interpolated Head-Related Transfer Functions (HRTFs).” (“Cheng”)
1032	May 22, 2023 Deposition Transcript of Durand Begault, Ph.D. from Eight KHZ, LLC v. Meta Platforms, Inc. et al., 6:22-cv-575-ADA-DTG (W.D. Tex.)
1033	U.S. Patent No. 257,453 to Ader (“Ader”)
1034	Olson, Harry. “Sound Reproducing Systems - Monaural, Binaural, Monophonic, and Stereophonic.” Audio, 1958 (“Olson”)
1035	Kyriakakis, Chris. “Fundamental and Technological Limitations of Immersive Audio Systems.” Proceedings of the IEEE, Vol. 86, No. 5, 1998 (“Kyriakakis 1998”)
1036	Begault, Durand. “3-D Sound for Virtual Reality and Multimedia.” 2000 (“Begault”)
1037	Vertesi, Janet. “Seamful Space: Heterogeneous Infrastructures in Interaction.” Science, Technology, & Human Values 2014, Vol. 39:2, 264-284 (“Vertesi”)
1038	Torrejon, Alfonso, et al. “Improving Communication and Presence in Online Telepresence Systems.” Workshop Proceedings of the 9th International Conference on Intelligent Environments, 2013 (“Torrejon”)
1044	U.S. Patent No. 2,955,156 to Heilig (“Heilig”)

30. I may rely upon these materials and/or additional materials to rebut arguments raised by 8KHZ. Further, I may also consider additional documents and information in forming any necessary opinions including documents that may not yet have been provided to me.

31. My opinions are additionally guided by my appreciation of how a person of ordinary skill in the art would have understood the claims of the '090 patent at the time of the alleged invention, which I have been asked to assume is June 23, 2014.

V. LEVEL OF SKILL IN THE ART AND PERSPECTIVE APPLIED IN THIS DECLARATION

32. I understand that certain issues relating to validity must be judged from the perspective of a person of ordinary skill in the relevant art, as I discuss below.

33. The invention of the '090 patent generally relates to well-known systems and methods of providing sound “comparable in quality to sound that listeners would hear if they were communicating face-to-face with each other.” Ex. 1001, 1:13-15.

34. In determining the level of ordinary skill, I have been asked to consider, for example, the types of problems encountered in the field, prior solutions to those problems, the rapidity with which innovations are made, the sophistication of the technology, and the educational level of active workers in the field. Taking those

factors into consideration, a person of ordinary skill in the art at the time of the claimed priority date of the '090 patent would have had at least a bachelor's degree in computer science or engineering with a focus on audio signal processing, with at least two years of relevant experience. It would have taken approximately at least two years of additional experience in the field for a person to become familiar with the problems encountered in the field and the prior and current solutions to those problems, although additional experience could substitute for less education, and additional education could likewise substitute for less experience.

35. My opinions do not turn on the precise definition, and the challenged claim would be unpatentable from the perspective of any reasonable POSA.

36. I believe that I would have qualified as a person of at least ordinary skill in the art as of the earliest claimed June 23, 2014 priority date on the face of the '090 patent. I believe that I have a sufficient level of knowledge, experience, and education to provide an expert opinion in the field of the '090 patent.

37. My opinions in this Declaration are based on the perspective of a person of ordinary skill in the art as of June 23, 2014 or somewhat before. This is true even if the testimony is stated in the present tense. Each of the statements below reflects my opinion based on my review of the disclosures of the '090 patent, the file history of the '090 patent, and the challenged claims.

VI. BACKGROUND

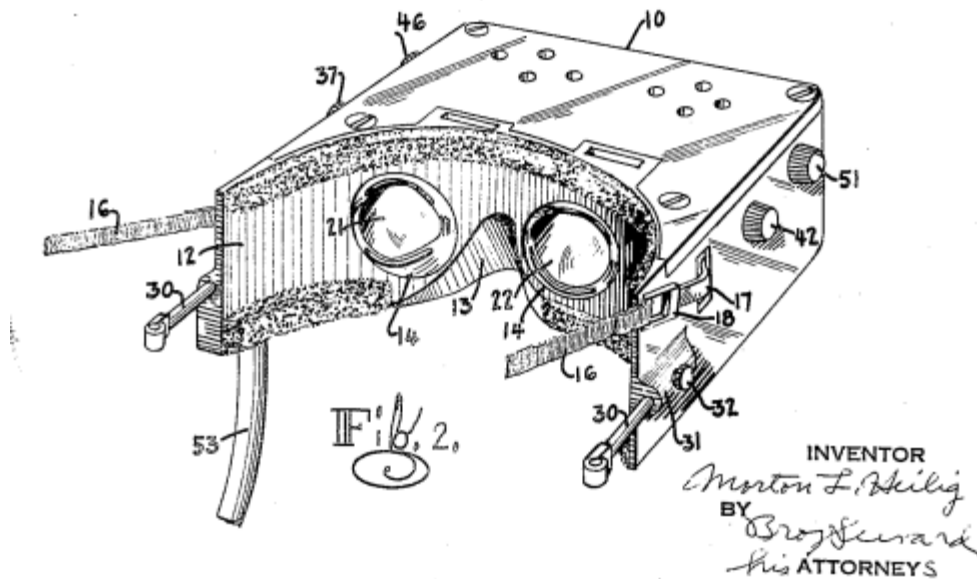
A. Overview of the Technology

1. Wearable virtual reality

38. The development of wearable virtual reality technology is decades old, dating back to the early 1960s. As I discuss below, the earliest wearable devices included, head mounted displays (“HMDs”) that were capable of reproducing sound. Later, the more sophisticated HMDs had head tracking systems that would adjust the displays and sound in response to the wearer’s head movements. By synchronizing the orientation of display to the direction of sound, and tracking and responding to the wearer’s head movements, these early devices provided immersive and realistic simulated experiences.

39. Morton Heilig was a cinematographer who in 1960 laid the foundation for these wearable devices by inventing an HMD, which he called the “Telesphere Mask.” Ex. 1020, 4. Heilig described his invention as a “Stereoscopic-television apparatus for individual use,” with earphones and a screen for each eye, in U.S. Patent No. 2,955,156 (Ex. 1044, Fig. 2 reproduced below, *see also* Ex. 1020). As Heilig envisioned, when wearing his device, “the spectator is given a complete sensation of reality, i.e., moving three dimensional images, which may be in color, with 100% peripheral vision, binaural sound, scents and air breezes.” Ex. 1044, 3:14-17 (emphasis added). Not only were the right and left eye images

“independent,” but Heilig also described his device as “providing independent ear phones, which are fed separately from sounds recorded by independent microphone systems, excellent stereophonic or binaural sound is given.” *Id.*, 3:22-25 (emphasis added). By providing binaural sound, Heilig’s device would have been capable of delivering spatialized sound to the wearer, which would have been consistent with Heilig’s vision of providing the “complete sensation of reality” in three dimensions.



40. Around the same time that Heilig developed his Telesphere Mask, engineers Charles Comeau and James Bryan developed the “Headsight” device in 1961. The Headsight device was the first HMD to have a head tracking system. Ex. 1021, 3-7. Images and sounds provided over the HMD were synchronized with and responded to the wearer’s head movements. *Id.* Specifically, Comeau and Bryan envisioned using the Headsight HMD as a surveillance device for a remote area. *Id.*,

3. The “camera at [a] remote location follows [the] operator’s head motions,” such that “[w]henver the operator points his head, right, left, up or down, the camera follows, viewing each portion of the scene in its correct relative position to the observer. If the observer tilts his head to one side the camera also tilts maintaining the viewed horizon at a constant level position.” *Id.*

41. Shortly after the Telesphere Mask and Headsight device were developed, Ivan Sutherland described an “Ultimate Display” in 1965 and another HMD in 1968. As Sutherland recognized, “[i]f the task of [a] display is to serve as a looking-glass into the mathematical wonderland constructed in computer memory, it should serve as many senses as possible,” including “meaningful sounds.” Ex. 1022, 1. Sutherland envisioned the Ultimate Display as a virtual environment in which “[a] chair displayed in such a room would be good enough to sit in.” *Id.*, 2.

42. Sutherland’s 1968 HMD was a three-dimensional display that “presented the user with a perspective image which changes as he moves.” Ex. 1023, 1. Sutherland’s HMD measured “the position and orientation of the optical system fastened to the user’s head” (pictured below), using a head position sensor. *Id.*, 1, 3-5. The display presented a three-dimensional image that “change[d] in exactly the way that the image of a real object would change for similar motions of the user’s head.” *Id.*, 1.

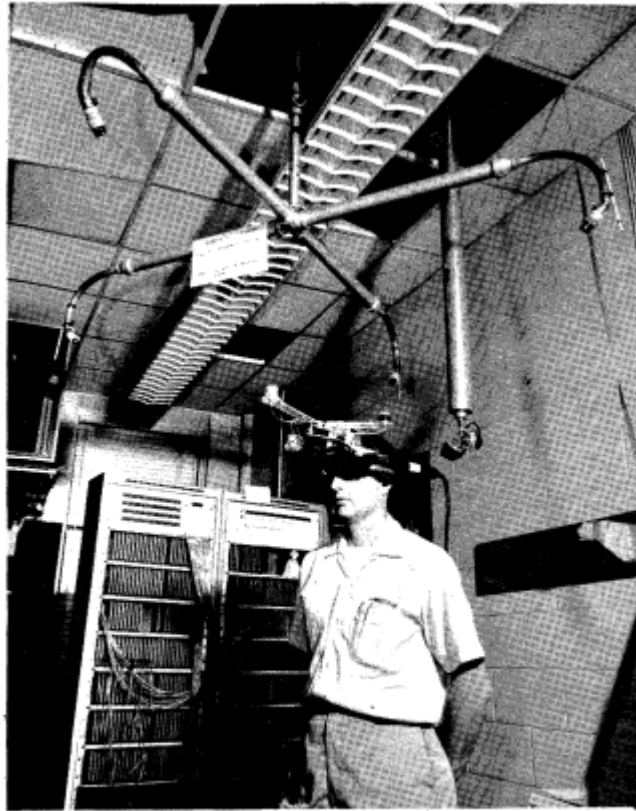


FIGURE 3—The mechanical head position sensor in use

2. Binaural sound

43. The concept of binaural sound has been well-known since the 19th century. The first public binaural sound exhibition occurred in 1881 during the International Exposition of Electricity in Paris. Using a device called the Theatrophone, its inventor Clement Ader (U.S. Patent No. 257,453, issued May 9, 1882) placed microphones with transmitters on the stage of the Paris Opera and transmitted the signals to the exhibition pavilion where listeners could listen to a

binaural live performance by placing a headphone in each ear. *See* Ex. 1033, *generally*.

44. Ader's patent teaches:

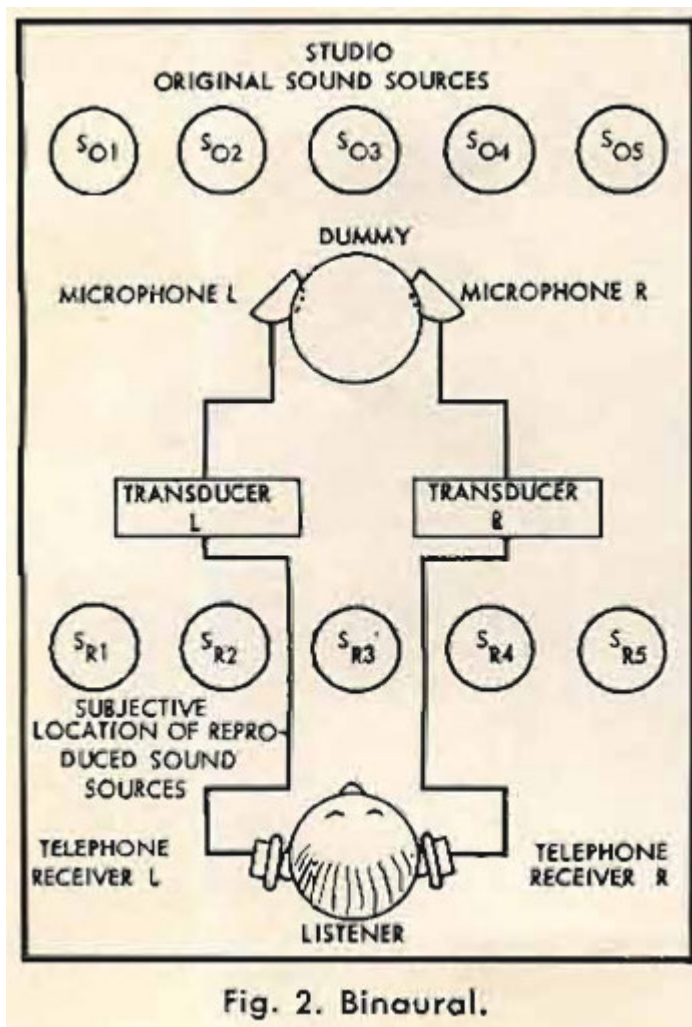
"I place a series of transmitters on the stage... each of these transmitters being arranged, as usual in a local circuit, with a battery and induction-coil and a line-circuit extending from each such coil to the houses of the respective subscribers. Two line-circuits extend to the house of each subscriber and there terminate in two receiving-telephones, one of which the auditor is to place at each ear. The receiver which he should place at his right ear is connected with a transmitter at the right of the stage, and the one which he should place at his left ear is connected with a transmitter at the left of the stage. **It is well known that the sounds transmitted telephonically vary in strength according to the distance of their source from the transmitter. Hence, when an actor or singer is at the right of the stage the sound heard by the auditor's right ear is the louder, and as he moves toward the left of the stage the sound heard by the auditor's left ear becomes the louder, while that heard by his right ear decreases.** An effect is thus produced upon the ears akin to that which the stereoscope produces on the eyes, so that the auditor is enabled to follow the actor's movements about the stage, and thus receives a more intelligible impression of the performance than by the arrangement heretofore employed."

Ex. 1033, 1:12-42 (emphasis added).

45. Later, Harry F. Olson formalized a definition for a system for reproducing binaural sound as having a listener wear two independent telephone

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

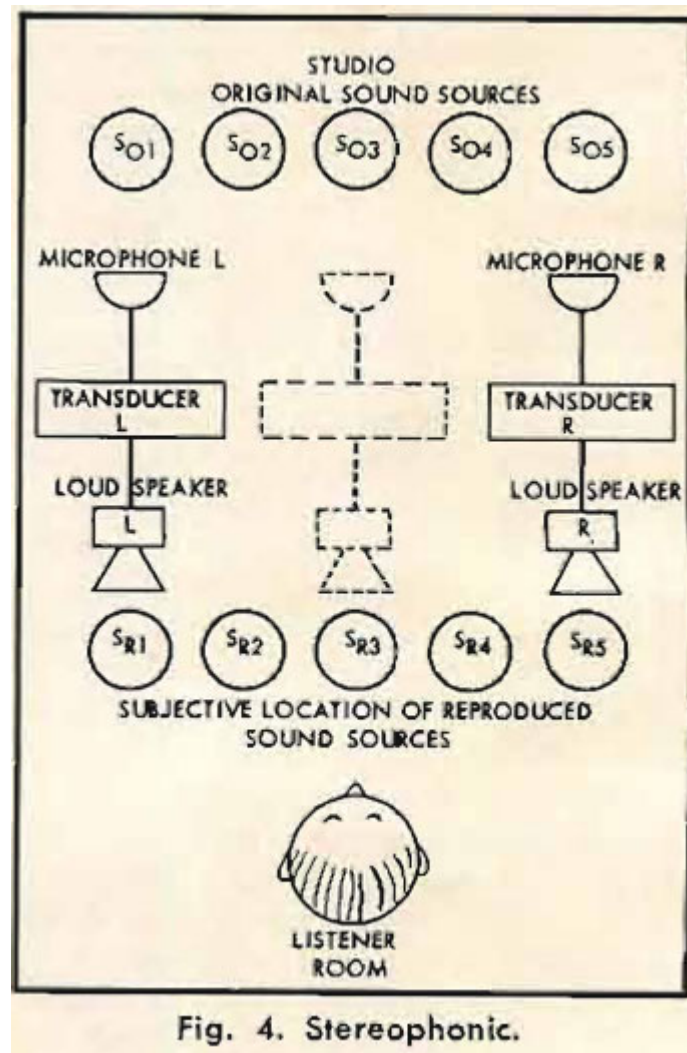
channels. Olson was a prominent acoustical engineering pioneer whose work in microphone and loudspeaker development would have been well-known to a POSA. In September 1958, Olson wrote in an article published in “Audio,” which was the most popular audio magazine when it was in print from 1947 to 2000. He wrote: “A binaural sound reproducing system is a closed circuit type of sound reproducing system in which two microphones, used to pick up the original sound, are each connected to two independent corresponding transducing channels which in turn are coupled to two independent corresponding telephone receivers worn by the listener, as in Fig. 2.” Ex. 1034, 1, Fig. 2 (reproduced below).



46. In the 1980s, Frederic Wightman and Doris Kistler at the University of Wisconsin published a set of papers that described the cues for locating external sound sources and methods to accurately simulate the same set of cues over headphones. Specifically, the researchers demonstrated how they were able to accurately simulate external sound sources at an SLP in an empty space (described as a free field) using headphones. Ex. 1026, 154. It was well-known that “binaural sound” or binaural audio was commonly used to refer to “systems or techniques that

capture or simulate ‘head-related’ signals containing natural acoustic differences between the two ears.” Ex. 1015, 1.

47. A related, but different type of sound is called stereophonic. Unlike binaural sound which is provided over wearable telephone receivers, or headphones, stereophonic sound is intended for playback over loudspeakers. Olson defines stereophonic sound as: “A stereophonic sound reproducing system is a field type sound reproducing system in which two or more microphones, used to pick up the original sound, are each coupled to a corresponding number of independent transducing channels which in turn are each coupled to a corresponding number of loudspeakers arranged, in substantial geometrical correspondence to that of the microphones, as in Fig. 4.” Ex. 1034, 1, Fig. 4 (reproduced below); *see also* Ex. 1013, ¶¶4, 16.



3. Sound source localization

48. It has been long recognized that binaural sound can simulate a spatialized sound field. In my 1998 review paper entitled “Fundamental and Technological Limitations of Immersive Audio Systems”, I wrote: “The human hearing process is based on the analysis of input signals to the two ears for differences in intensity, time of arrival, and directional filtering by the outer ear.”

Ex. 1035, p. 944.

49. This analysis of processing and comparing sound cues delivered to each ear is used to determine the location of a sound (i.e., the sound localization point, or SLP). As I summarized in that paper, sound localization theories date back to the 19th century, as “Several theories were proposed as early as 1882 [15] that identified two basic mechanisms as being responsible for source localization: 1) interaural time differences (ITDs) and 2) interaural level differences (ILDs). A later theory by Lord Rayleigh [16] was based on a combination of ITD and ILD cues that operated in different wavelength regimes.” *Id.*

50. **ITD**: The ITD is the “perceived difference[] in the time of arrival of sound at the two ears.” *Id.* For example, and as the patent recognizes, when a sound source is positioned directly in front of the human listener, the sound generally arrives at each ear of the human listener at the same time, as depicted in Figs. 13AB of the '628 patent.

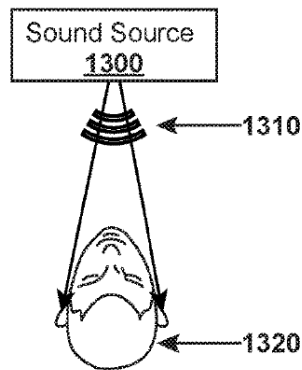


Figure 13A

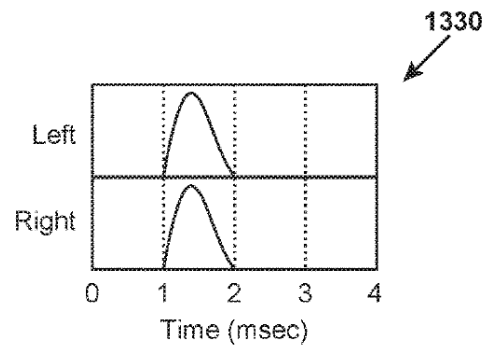


Figure 13B

Ex. 1001, Figs. 13A-13B

But sound from a source positioned to the left of the listener will arrive at the left ear slightly before it arrives at the right ear, as depicted in Figs. 13CD of the '628 patent.

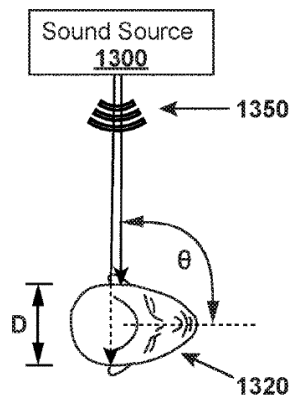


Figure 13C

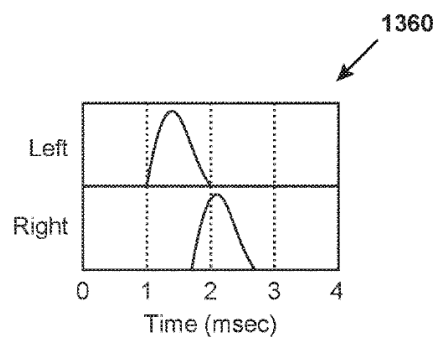


Figure 13D

Ex. 1001, Figs. 13C-13D

The difference in arrival time between the two ears is the ITD, which was well-known to be a basic cue for sound source localization. Ex. 1035, 944; Ex. 1013, ¶5.

51. **ILD**: Another well-known cue for sound source localization is the ILD. For a sound source positioned closer to one ear than the other, the sound level between the two ears also differs, especially in the higher frequencies in the range 4-20 kHz. This difference in sound level arises between the two ears because “the listener’s head casts an acoustical shadow giving rise to a lower sound level at the ear farthest from the sound source.” Ex. 1035, 944; *see also* Ex. 1013, ¶5.

52. **HRTF**: Sound sources directly in front of the listener or directly behind the listener arrive at the two ears with no difference in arrival time or level. Yet, humans are able to localize sounds positioned in the front versus the back. This is due to other factors that also impact the manner in which an ear receives sound from a point in space. Such factors include reverberation, reflection, refraction, attenuation, and even the shape of the listener’s ears, head, shoulders, and torso, which as the POSA would have readily recognized, all affect the sound that reaches a listener’s ears, depending on the environment. *See, e.g.*, Ex. 1009 (describing a virtual environment in which reverberation, reverberation delay, reflections, decay time, reflection delay, and other reverberation parameters change depending on the distance and hardness of features in the environment).

53. These factors are captured in the concept of head-related transfer functions (HRTFs). As I wrote in 1998:

“While time or intensity differences provide source direction information in the horizontal (azimuthal) plane, in the median plane, time differences are constant and localization is based on spectral filtering. The reflection and diffraction of sound waves from the head, torso, shoulders, and pinnae, combined with resonances caused by the ear canal, form the physical basis for the HRTF. [...] Spatial information is thus encoded onto the signals received by the eardrums through a combination of direction-dependent and direction-independent filters [18], [19]. The magnitude and phase of these head-related transfer functions vary significantly for each sound direction but also from person to person.”

Ex. 1035, 944.

4. Creating binaural sound

54. Binaural sound can be created in two ways. The **first** is to **record** binaural sound using two microphones that are spaced at the same distance as the width of a human head (*i.e.*, a binaural recording). This is often achieved by placing microphones on the sides of a mannequin (or dummy) head. When the left microphone signal is sent to the left ear and the right microphone signal is sent to the right ear (using headphones), the listener perceives the recorded scene with precise localization of each source in 3-D space.

55. The **second** is to **synthesize** binaural sound by processing non-binaural

content with the left and right HRTFs of the desired placement angle and position. For example, a monaural sound source, such as a person's voice, is processed (via convolution) with a left ear HRTF and a right ear HRTF, where each HRTF is designated at the desired placement angle and position, such as 30° to the front left of the listener and at a specified distance from the listener. Ex. 1025, 5; *see also* Ex. 1013, ¶¶11-12, 17, 20, 71, 75-77. This processing produces a “virtual auditory space,” with a two-channel synthetic binaural signal with the left channel sent to the listener's left ear and the right channel to the listener's right ear (using headphones). Ex. 1025, 5. At the relevant time period, HRTFs had been known for decades and were commonly used to characterize the manner in which an ear receives a sound from a point in space. Ex. 1024, 1; Ex. 1025, 5.

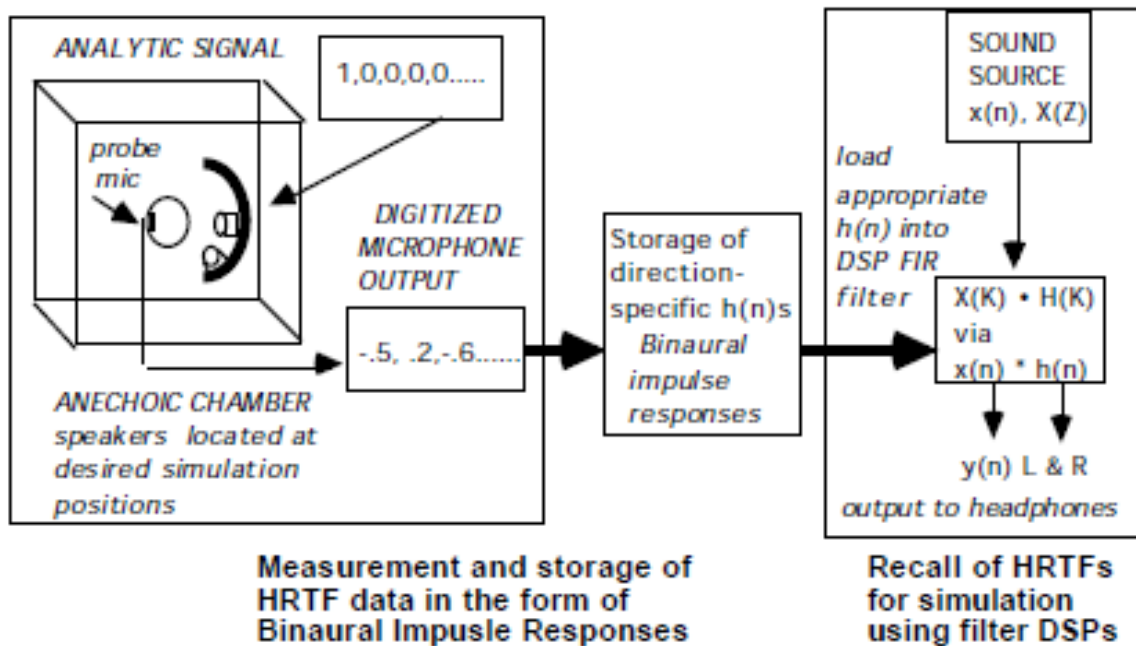
56. HRTFs are typically measured in advance (e.g., in a laboratory) and then made available to anyone wishing to produce synthetic binaural sound. A commonly used set of HRTFs for several hundred angles around the head was made publicly available by Bill Gardner and Keith Martin of the MIT Media Lab in 1994, using a mannequin with artificial ears referred to as KEMAR.

57. It was commonly known to measure HRTFs, by the “[m]easurement of [b]inaural [i]mpulse [r]esponses,” Durand Begault described in his April 2000 book “3-D Sound for Virtual Reality and Multimedia”:

“One can obtain the HRTF of an individual or of a dummy head by playing an analytic signal at a desired position at least a meter distant, and measuring the impulse response with probe microphones placed in the vicinity of the ear canals. This measurement then can be formatted for use directly within spatial filtering DSPs. Figure 4.17 shows the general procedures involving the measurement, storage, and recall for simulation of HRTF data. A sound source (e.g., a loudspeaker) generates an acoustical signal whose time and frequency characteristics are known in advance.”

Ex. 1036, 109.

58. Begault depicts in Fig. 4.17 (reproduced below) an “[o]verall plan of the HRTF measurement-storage-simulation technique. HRTFs are obtained using acoustic signals from loudspeakers that are recorded with probe microphones in the vicinity of the ear canals. The data are stored for input to DSP filters during the simulation stage.” *Id.*, 110.



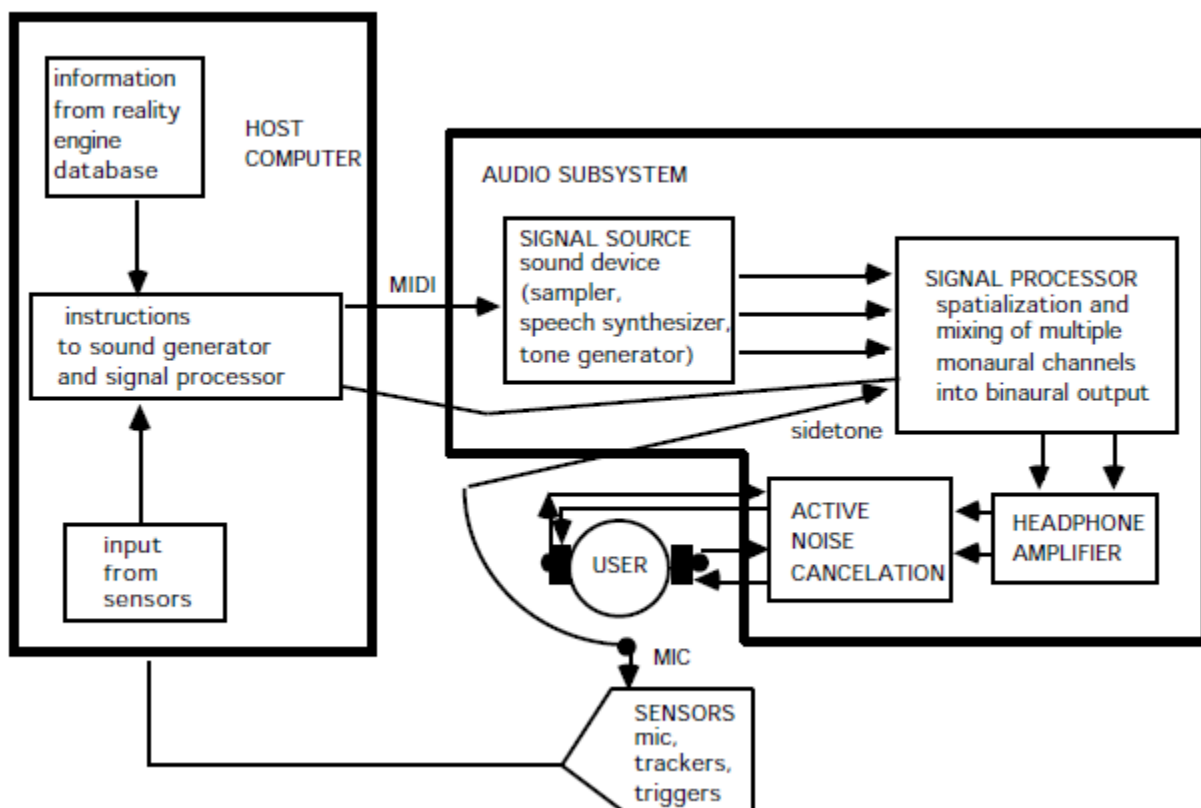
59. Such 3-D sound was useful in various virtual reality environments, such as videoconferencing. *Id.*, 174 (describing arranging spatialized audio signals in virtual space for multiconversation teleconferencing with broadcast video images of the participants).

5. Compensating for head movement

60. In the relevant time frame, it was commonly known for virtual reality systems to include head trackers. In the 1980s, Wightman and Kistler built on the technology from the 1960s head tracking devices described above (Comeau and Bryan's "Headsight" device and Sutherland's HMD). Wightman and Kistler's head tracking system determined the position of the user's head, and then used that head

position to select “which transfer function should be used in the headphone presentation to maintain an apparently stable source location.” Ex. 1026, 155.

61. In his book, Begault depicts the components of a virtual reality system as including audio and visual components in addition to sensors such as “trackers.” Ex. 1036, 9, 13, Figs. 1.7, 1.8 (reproduced below).

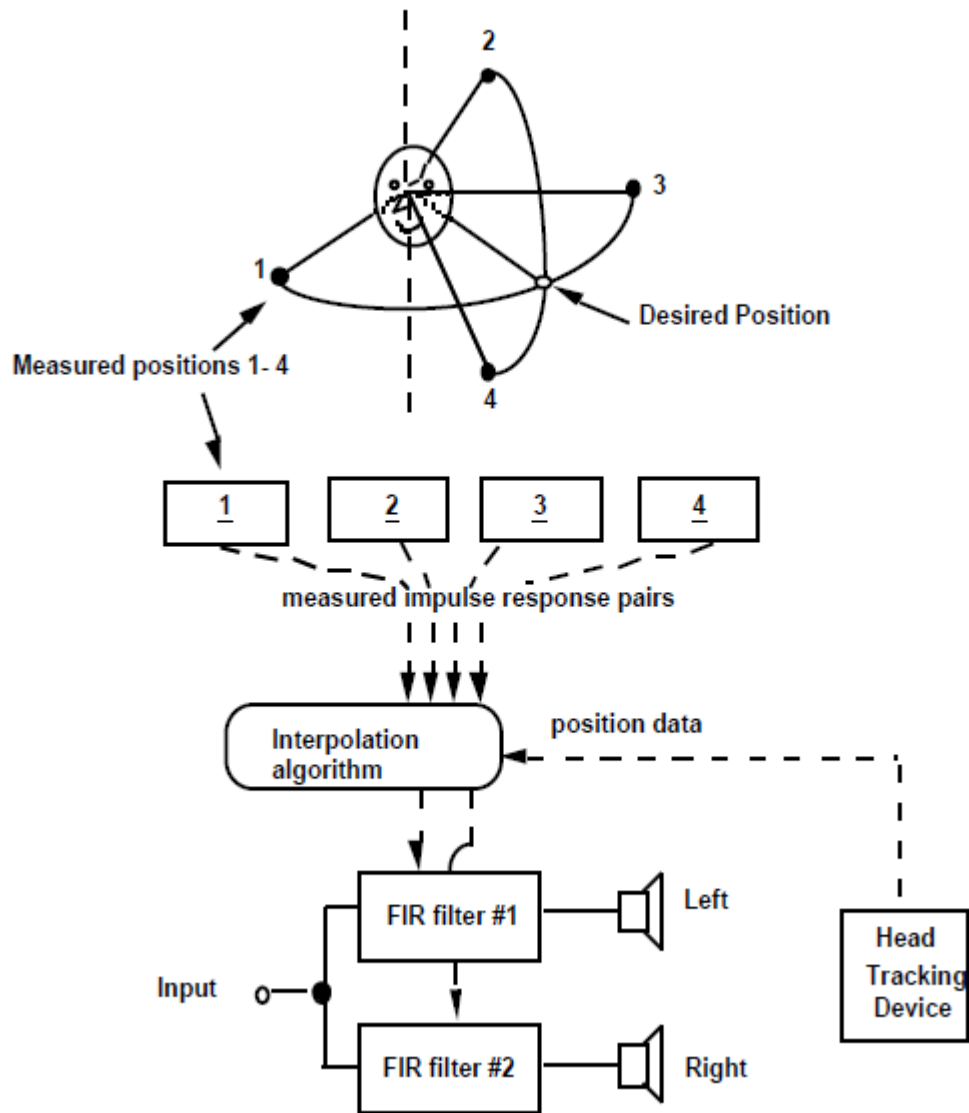


62. These “trackers” were for tracking the head movement of the listener, to adjust the binaural sound delivered to the listener. This was done so that the perceived location of the sound source remains fixed in virtual space, regardless of the listener’s head movement:

“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.”

Ex. 1036, 133; *see also id.*, 167.

63. As Begault depicts in Fig. 4.38 (reproduced below), the head tracking device measures the impulse response pairs for different positions of the user’s head, and uses an “[i]nterpolation procedure [] to compensate for head movement of the listener.” *Id.*, 134.



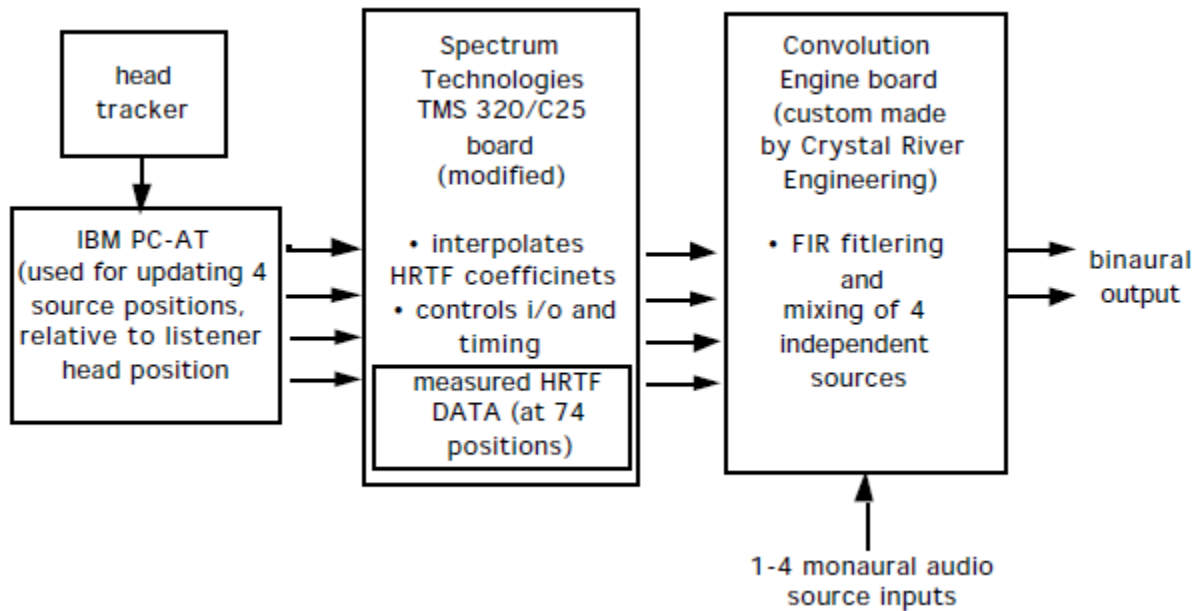
64. It was commonly known that the benefits of such systems was that the perceived location of the sound source can remain fixed in virtual space, regardless of the listener's head movement. Doing so would mimic spatial hearing in natural environments and enhance the immersivity of the environment:

“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.”

Ex. 1036, 167 (emphasis added).

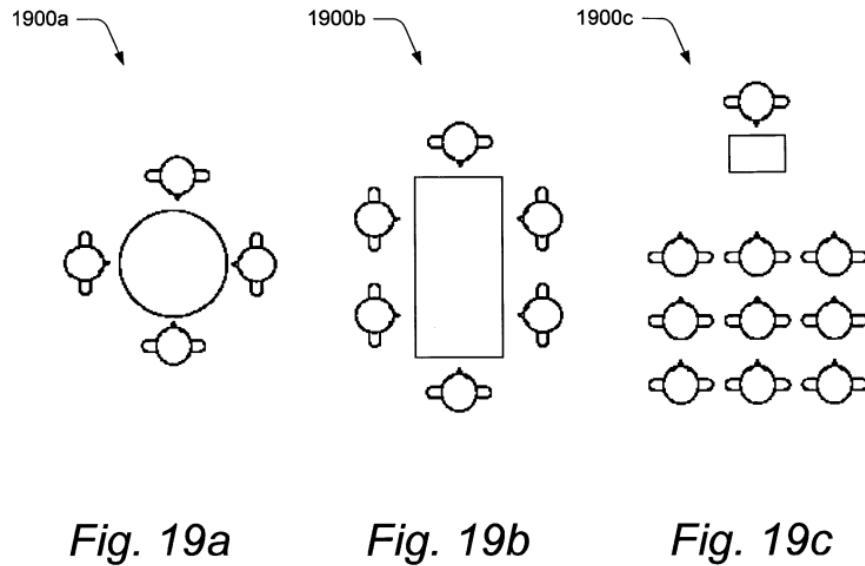
65. The Convolvotron was another head-tracked virtual audio system developed in the 1980s. *Id.* The system tracked the relative position of the head, and based on that position, selected HRTF filters from a library to dynamically move virtual sources. *Id.*, 167-169, Fig. 5.8 (reproduced below); Ex. 1032, 48:13:49:8, 50:12-52:23, 250:11-253:3 (8KHZ’s expert in the parallel district court litigation admitting that the Convolvotron maintained sound sources in the same virtual position when a user’s head was moving by “analyzing head position relative to the intended virtual source position, and then selecting or interpolating between the appropriate HRTFs.”).



6. Virtual conferencing and hybrid collaboration environments

66. In 2014, virtual and hybrid environments with spatialized audio was commonly known. For example, Bedingfield described different sound field configurations, and contemplated providing spatialized audio to remote participants as if they were seated in chairs “around a circular conference table.” Ex. 1005, 9:20-23, 16:36-48, Fig. 19a-c (reproduced below).

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090



In Bedingfield’s system, all participants are remote, and the environment is virtual. Hybrid environments were also commonplace. *See, e.g.,* Ex. 1038, Fig. 5, reproduced below, depicting a classroom setting with empty chairs indicated in red and designated for remote students, with in-person students indicated in blue. Specifically, each “immersive student” in a red circle is a remote student who “remotely choose[s] to sit on an empty chair” in the classroom. *Id.*, 5, 6 (describing remote students “tak[ing] ownership of a seating place”).

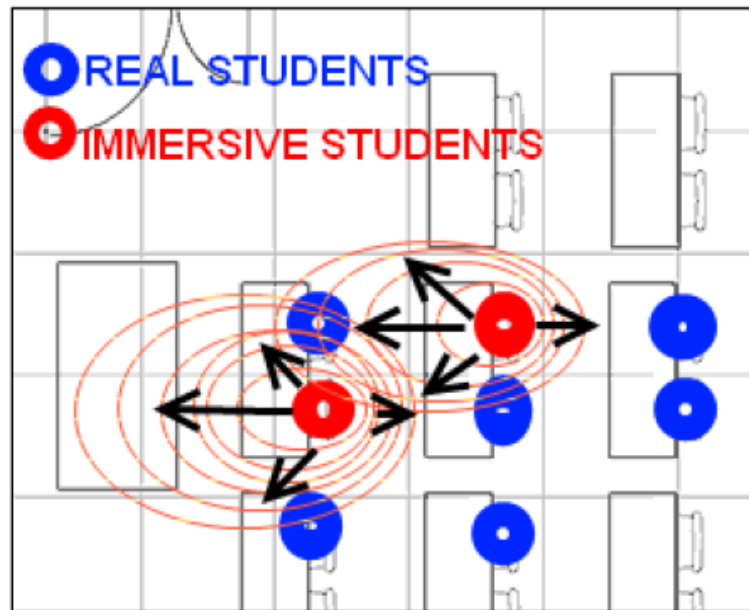


Figure 5. Communication between real and immersed elements

67. In such environments, it was known to provide spatialized sound to the remote or in-person participants at SLPs. For example, the red locations (above) were known to be SLPs because the audio is “directional” and “take[s] into account the location in space of the participants,” and “[b]inaural audio elements” are added to “enhance the ‘feeling of been there.’” Ex. 1038, 7-8 (describing how local students use headphones to “be able to hear and talk to [their peers seated among them], depending o[n] their seating distance”).

68. Spatialized audio was commonplace and desirable because it “closely simulate[s] an in-person conference setting” (Ex. 1005, 3:25-27) and provided “virtual bonds between people that are geographically dispersed,” which “gives the opportunity for more intimacy, leading to a better understanding of each other, [and]

providing the opportunity to balance the groups [sic] weaknesses in order to achieve their goals” (Ex. 1038, 2).

69. **Accordingly, in my opinion it was known well before the relevant time frame to (1) capture binaural sound, (2) designate an SLP, and (3) adjust the binaural sound to originate from the SLP.**

B. The’090 Patent

70. The ’090 patent is generally directed to well-known systems and methods of providing sound “comparable in quality to sound that listeners would hear if they were communicating face-to-face with each other.” Ex. 1001, 1:15-16. In the embodiments claimed by the ’090 patent, the computer system positions a voice of one person at a SLP that is “away from and proximate” to the other person. Ex. 1001, Abstract, Figure 1. Specifically, the system captures the binaural voice of the speaker, designates a SLP, adjusts the binaural voice so that it simulates a binaural voice originating from the sound localization point, and provides that adjusted binaural voice to the listener.

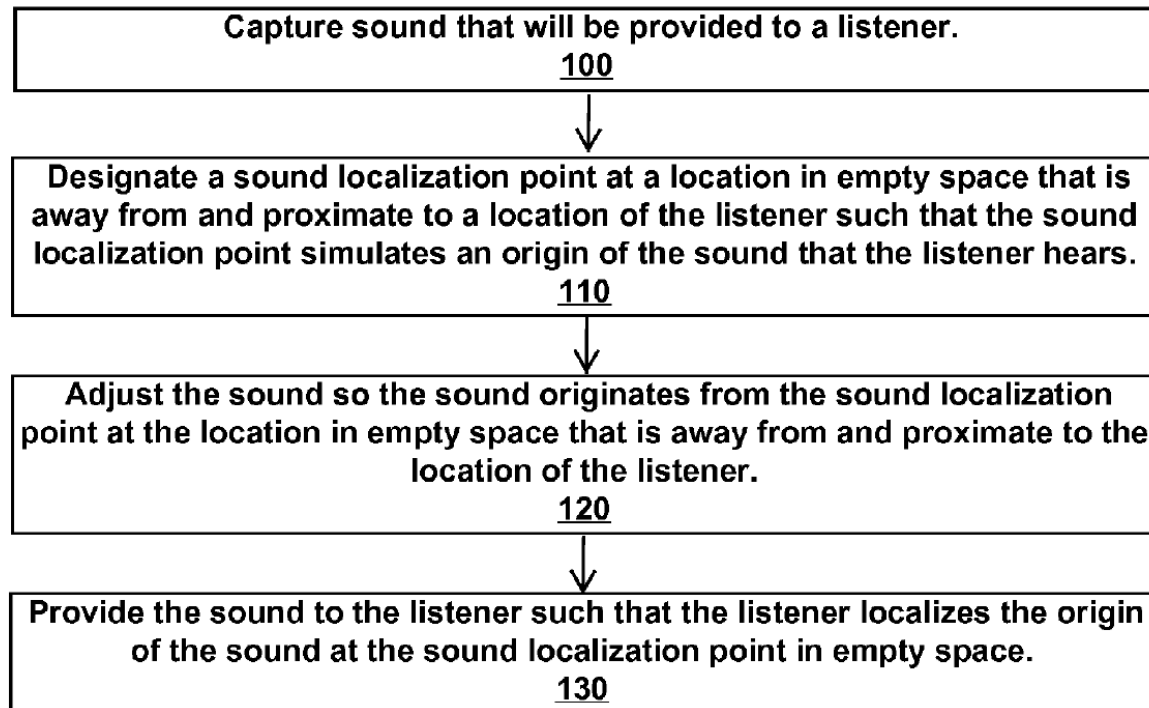


Figure 1

71. In another embodiment, the system maintains the origin of the voice at the SLP even when the head of the listener moves. This is done in a routine way—by adjusting the interaural time difference of the binaural voice. For example, if the listener is facing the sound localization point, as in Figure 13A and 13B, the sound would arrive at the left and right ear at the same time. However, if the listener turns away from the sound localization point, as in Figure 13C and 13D, the sound “arrives at the left ear first and later in time at the right ear.” Ex. 1001, 27:20-21.

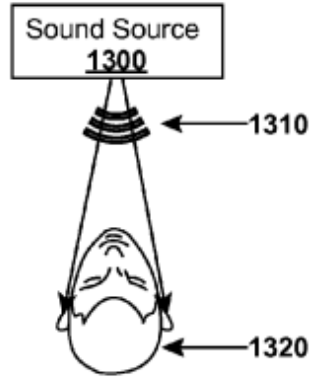


Figure 13A

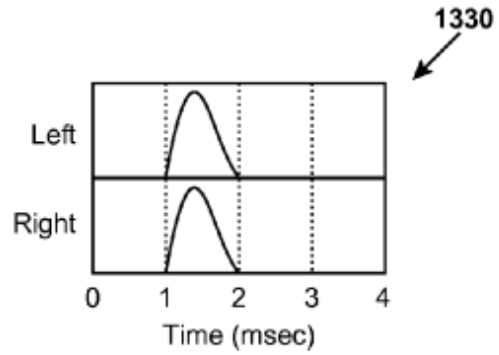


Figure 13B

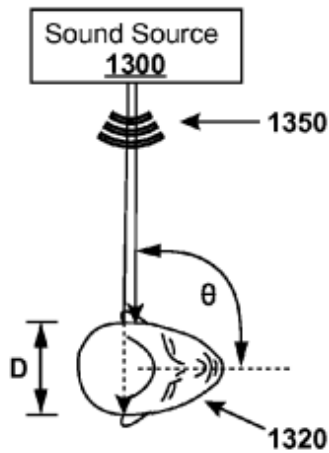


Figure 13C

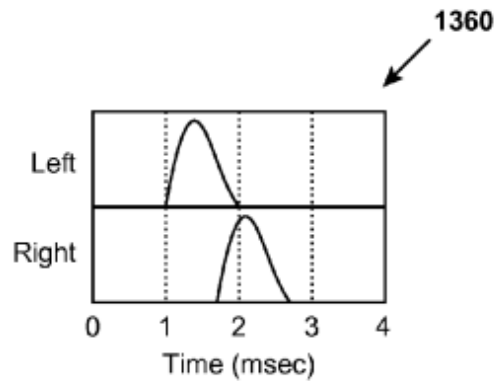


Figure 13D

C. Prosecution History

72. I have reviewed the prosecution history of the '090 Patent, and based on the prosecution history, the application leading to the '090 patent was filed on June 23, 2014, with twenty original claims (Ex. 1002, 1-110). During prosecution, the examiner issued no rejection. *See generally id.*

73. On November 2, 2015 the Examiner issued a Notice of Allowance and Fees Due, in which he allowed all twenty claims. *Id.*, 115-121. The Examiner's statement of reasons for allowance referred to the following element:

... adjusting, with the computer system, the binaural sound captured at the earphone of the talking person so the binaural sound originates during the telephone call from the sound localization point in empty space that is away from and proximate to the listening person; and

providing, with an electronic earphone located at a head of the listening person, the binaural sound to the listening person during the telephone call such that the origin of the binaural sound for the listening person occurs at the sound localization point in empty space that is away from and proximate to the listening person.

Id. The Examiner further stated: "These features, in combination with the other features of Claims 1,8 and 15 are not anticipated by, nor made obvious over, the prior art of record." *Id.*

74. On November 18, 2015, the applicant filed a post-allowance communication, stating "Applicant respectfully disagrees with some of the Examiner's stated reasons for allowance. The recitations cited above appear in

independent claim 1, but not in independent claims 8 and 15. Applicant submits that no words or terms should be ‘read into’ a claim based on the Examiner.” *Id.*, 134.

75. The ’090 patent issued on December 29, 2015.

D. Challenged Claims of the ’090 Patent

76. I have been instructed to assume that the challenged claims are entitled to an effective filing date of no earlier than June 23, 2014.

77. The ’090 patent has 20 claims, 3 independent claims and 17 dependent claims.

78. Independent claim 1 is a method claim, reciting a four-step process:

- (1) capture binaural sound at a talking person;
- (2) designate an SLP to a listening person;
- (3) adjust the binaural sound to originate from the SLP; and
- (4) provide the binaural sound to occur at the SLP.

79. Independent claim 8 is a claim directed at a storage medium storing instructions to execute a four-step process:

- (1) commence a call between two individuals;
- (2) designate an SLP for the first individual;
- (3) execute the call such that the origin of the voice of the second individual appears to originate from the SLP while the first individual moves and changes location and head orientation;

- (4) maintain the origin at the SLP but alter the voice's ITD in response to the first individual moving the head orientation.

80. Independent claim 15 is a system claim with four components that implement a similar process:

- (1) a first wearable electronic device;
- (2) a second wearable electronic device that captures a binaural voice;
- (3) a memory; and
- (4) a processor that adjusts the binaural voice to position an SLP.

E. Effective Priority Date

81. I understand that the '090 patent was filed on June 23, 2014. For the purpose of my analysis, I have been asked to assume that the challenged claims of the '090 patent are entitled to a priority date of June 23, 2014.

VII. CLAIM CONSTRUCTION

82. I understand that Patent Owner contends certain claim elements in the '090 patent should have their plain and ordinary meaning. Specifically, I understand that 8KHZ's position is that if construed, the below terms in the '090 patent should have the following meanings:

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

- *electronic call* in cls. 8, 14 and *electronic phone call* in cl. 15 should mean “voice communication over an electronic network between at least two parties”
- *away from and proximate to* in cls. 8, 14, and 15 should mean “a distance from”
- *a binaural voice* in cl. 15 should mean “a binaural sound”
- *wearable electronic device / WED* in cl. 15 should mean “portable wireless electronic device that is worn on or attached to a person”

I have applied the plain meaning of the terms and phrases in the claims in my analysis, as a POSA would have understood those terms and phrases at the relevant time (around 2014), and where applicable, have also applied 8KHZ’s positions as I understand them. However, it is my opinion that the prior art relied on in the grounds set forth below disclose the subject matter of the challenged claims under any reasonable construction, including their plain meaning.

VIII. THE PRIOR ART

83. In this section, I provide an overview of the five prior art references relied upon in the obviousness combinations detailed herein.

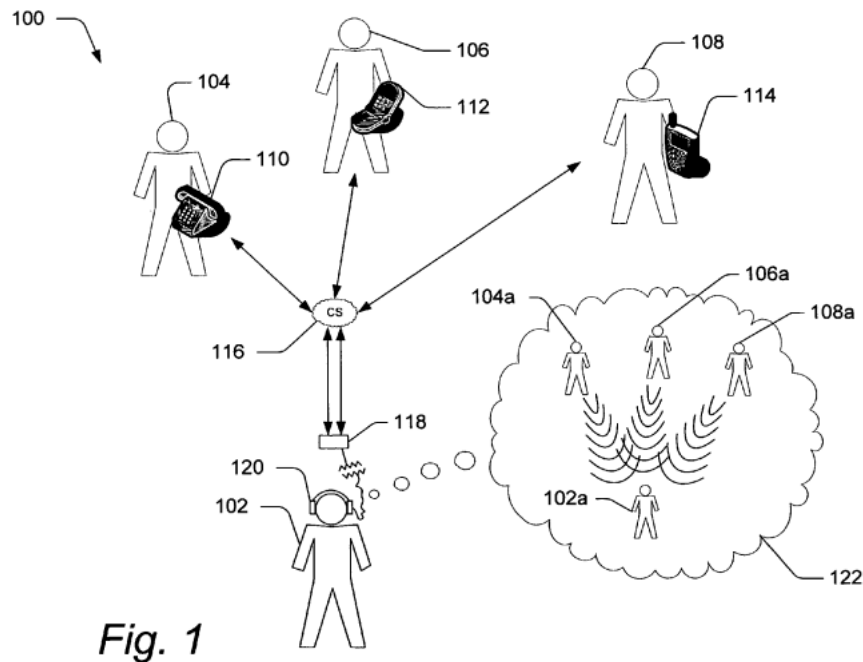
A. Bedingfield

84. U.S. Patent No. 8,406,439 (“Bedingfield”), published on March 26, 2013, is prior art under 35 U.S.C. §§ 102(a)(1) and (a)(2). Bedingfield generally

discloses all the limitations of the independent challenged claims, except for *captur[ing] binaural sound or voice*, and maintaining the voice's origin at a fixed SLP *while the user moves and changes head orientation*. Bedingfield was not cited or discussed during the '090 patent prosecution.

85. Bedingfield is generally directed to synthetic audio placement and rendering a sound field of different audio sources. Ex. 1005, Abstract, 1:15-17, 1:50-2:16, 4:18-19.

86. Like the '090 patent, Bedingfield provides the sound field over headphones and the listener of Bedingfield's conferencing system perceives the other participants as if they "were in the same location and speaking to each other directly." Ex. 1005, 4:31-5:3, Fig. 1. Bedingfield provides a synthetic sound field to wearable speakers such as headphones or a headset (*id.*, 7:43-46, 13:34-43), which sense the movements or orientation of the listener's head (*id.*, 13:36-57).



B. Li

87. U.S. Patent No. 9,131,305 (“Li”), filed January 17, 2013 and published on September 8, 2015, is prior art under 35 U.S.C. §102(a)(2). Li generally discloses all the limitations of the independent challenged claims, except for *captur[ing] binaural sound or voice*, and maintaining the voice’s origin at a fixed SLP *while the user moves and changes head orientation*. Li was not cited or discussed during the ’090 patent prosecution.

88. Li is generally directed to a 3D sound system that allows users to “experienc[e] music, movies, [and] video games.” Ex. 1006, 7:21-22, 15:8, 16:4-7, 16:63-66, 19:39-20:52, Figs. 9, 16, 33.

89. Like the ’090 patent, in an application of a conferencing system, Li “generate[s] configurable 3D sounds,” which involves processing the speaker’s

voice to position “different speaking users at different spots.” *Id.*, 33:14-18, 33:32-51, 33:62-66, 34:4-8, 34:20-51, 35:58-36:2. To generate a 3D sound, Li discloses processing sound with head related transfer functions (HRTFs), which model cues such as ILD and ITD to help humans localize sound sources in three dimensions. *Id.*, 1:65-2:21, 12:51-56, 22:49-52. Li also discloses a user interface that allows the user to provide commands that give the user the freedom to “move the sound source up and down, etc., on the configurable sound field.” *Id.*, 20:20-36, 21:57-22:5.

C. Andrea

90. U.S. Patent No. 8,818,000 (“Andrea”), filed October 29, 2010 and published on August 26, 2014, is prior art under at least 35 U.S.C. §102(a)(2). Andrea discloses a wearable device that captures *binaural sound* and *voice*. Andrea was not cited or discussed during the ’090 patent prosecution.

91. Like the ’090 patent, Andrea captures a voice at a wearable device and manipulates that sound to improve the quality of the recorded voice. Ex. 1007, 1:44-51, 6:13-14, 11:6-27. Andrea’s wearable device (earbuds) includes a microphone at each ear (Ex. 1007, 3:28-31), which capture the wearer’s voice to create a binaural recording. *Id.*, 1:44-51, 14:55-58. As Andrea describes, the usage of two microphones is beneficial for noise cancellation, employed for “directional listening.” *Id.*, 1:44-51.

D. Ballas

92. U.S. Patent Pub. No. 2003/0059070 (“Ballas”), published on October 31, 2005 is prior art under at least 35 U.S.C. §§102(a)(1) and (a)(2). Ballas discloses the use of head tracking and maintaining an SLP even as the user’s head moves and changes location and orientation. Ballas was not cited or discussed during the ’090 patent prosecution.

93. Like the ’090 patent, “allow[s] the listener to change locations and head orientation without degradation of the audio signal.” *Id.*, ¶¶17-18, 25, cl. 1, 7, 13. Specifically, Ballas’s head tracking system provides the location and orientation of the wearer’s head to a computer system that “processes the audio signals using a head related transfer function (HRTF) filter 106 thus producing spatialized audio.” *Id.*, ¶¶17-18, 25-26. The HRTF filter “maintains the virtual location of the audio signals/sound” even as the user’s head moves. *Id.*, Abstract, ¶17.

E. Sandgren

94. U.S. Patent Pub. No. 2008/0273708 to Sandgren, published November 6, 2008 is prior art under at least 35 U.S.C. §§102(a)(1) and (a)(2). Sandgren discloses changing reverberation and attenuation to simulate 3D audio. Ex. [[Sandgren]], ¶¶6, 12, 77. Sandgren was not cited or discussed during the ’090 patent prosecution.

95. Like the '090 patent, Sandgren simulates sound-reflecting objects and contemplates that when sound waves are absorbed and reflected by objects within a room. *Id.*, Abstract, ¶¶2, 12, 16, 70, Fig. 1. To make the sound field as realistic as possible, Sandgren updates parameters for a reverberation generator and attenuation filters to reflect the movement of such objects. *Id.*, ¶¶31-33, 69.

**IX. GROUND 1: BEDINGFIELD IN VIEW OF ANDREA RENDERS
OBVIOUS CLAIMS 1-3, 6, 15-17**

A. Motivation to Combine

96. It is my opinion that it would have been obvious to a POSA implementing the communication device for participants of the Bedingfield communication system to use a wearable device that captures *binaural* sound as taught by Andrea.

97. It is my opinion that a POSA would have been motivated to implement this modification for the following reasons.

98. First, a POSA would have recognized that capturing binaural sound, such as binaural voice, and providing it to a listener would further advance Bedingfield's stated goal to improve sound quality. Specifically, an objective of Bedingfield is to "allow a person to hear the speaker better" during an electronic call, like the exemplified conference call. *See* Ex. 1005, 1:44-46. A person wearing a wearable electronic communication device during an electronic call may be at a

location with a lot of ambient noise. Ambient noise is the background or environmental sound that is present in a recording or listening situation. When a microphone on a wearable device picks up the person's voice, the background or environmental sound is also captured by the microphone. This can often interfere with the intelligibility of the intended sound source and reduce the quality and clarity of a person's voice.

99. Ambient noise can also interfere with the simulation of a face-to-face conversations. For example, if the speaker is at a noisy location and the listener is at a quiet location, it will be difficult for the listener to imagine themselves in the same room as the speaker if the listener can hear ambient noise incongruent with the listener's location.

100. During the relevant period (and well before) a known solution to this problem was noise cancelation as taught in Andrea.

101. Andrea discloses a noise canceling solution that may be employed for electronic communications, including for "directional listening." Ex. 1007, 1:44-51. Andrea teaches that the noise canceling solution requires two or more microphones that capture binaural sound. Ex. 1007, 6:13-14 ("In the simplest embodiment the array comprises two microphones"), 11:25-27 ("The two elements or microphones 250 of the stereo headset-microphone array device may be mounted on the left and right earphones of any size/type of headphone."). It was well-known that "binaural

sound” or binaural audio was commonly used to refer to “systems or techniques that capture or simulate ‘head-related’ signals containing natural acoustic differences between the two ears.” Ex. 1015, 1. Therefore, it is my opinion that Andrea’s disclosure of two microphones mounted on the left and right earphones captured binaural sound, including the binaural voice of the speaker.

102. By comparing the microphone outputs (i.e., the binaural sound), the computer can separate the user’s voice from the ambient sound, and generate a new signal that eliminates the ambient noise. Specifically, Andrea teaches that the microphones are placed equidistant from the sound source—the user’s mouth. Thus, the voice signal will arrive at the same time and at the same intensity, whereas other sound will not. “The stereo microphone spacing allows for determining different time of arrival and direction of the acoustic signals to the microphones. . . . [T]he voice signal 210 will . . . arrive in-phase at same time with equal amplitude at both the microphones, while noise from the sides will arrive at each microphone in different phase/time and be cancelled by the adaptive processing of the algorithm.” Ex. 1007, 11:6-13.

103. By eliminating the ambient noise from the microphone outputs, the listener will be able to hear the speaker more clearly. Thus, a POSA would have recognized that incorporating Andrea’s capture of *binaural* sound into Bedingfield’s

communication system would have reduced noise, improved sound quality, and enhanced the experience of those participating in the electronic call.

104. Therefore, by capturing binaural sound, as Andrea teaches, the POSA would have known this makes it easier for the listener to better distinguish between the speaker and noise. As the POSA would have further understood, capturing binaural sound would have improved call quality by allowing the listener to perceive a spatialized voice, allowing for a more life-like conversation.

105. Second, the modification would have also been a natural extension of Bedingfield's wearable electronic device. A POSA implementing the communication system of Bedingfield would have been motivated to improve the system by implementing known advancements in wearable devices. Bedingfield already contemplates the use of a "combined microphone input and speaker output" (i.e. headset). Ex. 1005, 11:5-6, Fig. 11, 7:43-44. A POSA would have been motivated to use other headsets that would advance Bedingfield's goals.

106. In particular, Bedingfield's system is designed to improve communication of a simulated conference. Ex. 1005, 3:25-27, 1:44-46. The ability to capture a call participant's binaural voice in order to eliminate noise, especially background and environmental noise, would improve a simulated conference. Not only would it by improve the sound quality, it would have also allowed Bedingfield's system to create sound localization points with increased fidelity.

B. Reasonable Expectation of Success

107. It is my opinion that a POSA would have had a reasonable expectation of success in using Andrea's wearable device that captures binaural sound in Bedingfield's communication system.

108. Bedingfield and Andrea are in the same field as each other. Each utilizes a wearable device to capture a user's voice using a wearable device and manipulate that sound to improve its communication.

109. Both Bedingfield and Andrea also concern electronic communications, including voice over internet protocol (VoIP) communications and computer-based communications. Ex. 1007, 1:49-51, Ex. 1005, 7:47-50. Like Bedingfield, Andrea "relates to audio transmitting/receiving devices such as headsets with microphones." Ex. 1007, 1:44-45.

110. It would have been simple for a POSA to modifying the system of Bedingfield to include a wearable device that implements Andrea's second microphone for noise cancellation. Andrea states that the devices it teaches "may be used in a multitude of different applications, including . . . communications such as voice over internet protocol ('VoIP'), PC to PC communications, PC to telephone communications." Ex. 1007, 1:49-51.

111. Further, it would have been straightforward and beneficial for a POSA to modify Bedingfield's system to include a second microphone, as doing so would

have enabled the system to include call quality improvement techniques such as noise cancelling algorithms, which were well known as of the priority date of '090 patent. Ex. 1007. Thus, a POSA would have had a reasonable expectation of success using Andrea's *capture[] of a binaural voice* of a call participant, in Bedingfield's wearable electronic device to improve sound quality and to produce more reliable sound localization points.

112. A POSA would have had a reasonable expectation of success in combining the teachings of Bedingfield and Andrea, because the proposed modification involves nothing more than a combination of known prior art elements (microphones) according to known methods and known techniques (noise cancelation) to yield predictable results (capture binaural sound), and involves use of a known technique (noise cancelation) to improve a similar device (wearable device) in the same way (adding a microphone).

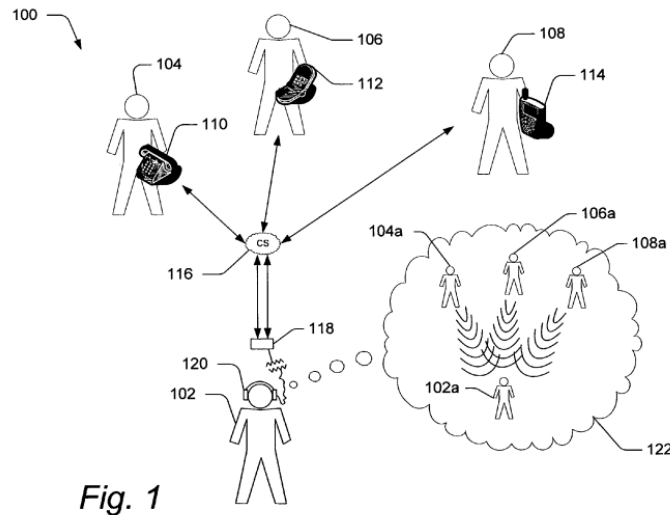
C. Independent Claims 1

1. 1[pre]: “A method, comprising:”

113. To the extent the preamble is limiting, it is my opinion that Bedingfield discloses it.

114. Bedingfield discloses “*methods* and systems for providing synthetic audio placement.” Ex. 1005, 1:15-17 (emphasis added). Specifically, Bedingfield teaches a system that implements methods to simulate sound sources, particularly

for a conference call, as shown in Figure 1. These methods “more closely simulate an in-person conference setting.” *Id.*, 3:25-27.



115. Bedingfield teaches “receiving an incoming signal over an audio connection, the incoming signal including audio input from a plurality of audio sources.” *Id.*, 1:51-54. The Bedingfield method creates “a sound field having a synthetic position associated with each of the plurality of audio sources relative to the listener. *Id.*, 1:54-59. This is done by assigning “a virtual position within [the] virtual sound field” for each audio source so that “[e]ach participant is heard by the listener in multi-channel audio based on the participant’s virtual location in the sound field.” *Id.*, 3:24-34.

2. **1[a]:** “*capturing, with an electronic earphone located at a head of a talking person, binaural sound that will be provided to a listening person during a telephone call;*”

116. It is my opinion that Bedingfield in view of Andrea discloses this limitation.

117. Bedingfield teaches *capturing, with* communication devices 110, 112, and 114 including phones and microphones (*an electronic earphone*) located at a head of participants 104, 106, 108 (*a talking person*), sound that will be provided to listener 102 (*a listening person*) during a telephone call.

118. Figure 1 of Bedingfield shows “an overall environment 100 for providing synthetic audio placement.” Ex. 1005, 4:31-33. This environment includes various participants, including listener 102 (*a listening person during a telephone call*) as well as participants 104, 106, and 108 (each of whom is *a talking person*). *Id.*, 4:19-23. The environment also includes a communication system 116 and a multi-channel processing device 118. *Id.*, 4:31-33.

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

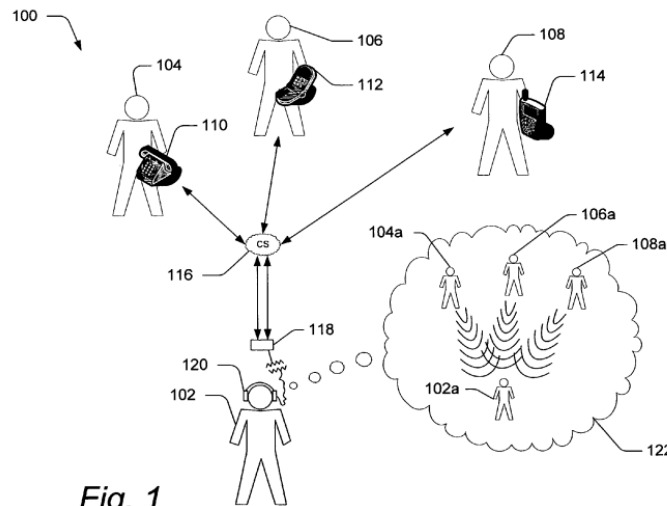


Fig. 1

119. Bedingfield also discloses “communications devices 110, 112, and 114” that the various participants 104, 106, and 108 (each of whom is *a talking person*) may have, such as “telephones, wireless phones, mobile phones, two-way voice devices, walkie-talkies, voice-over-IP (VoIP) devices, microphones, or other communication devices that permit multi-directional communication over a communication system.” *Id.* 4:23-33. It was common in 2014 for *electronic earphones* such as earbuds or headphones to connect to such devices, such that the POSA would have found it obvious that these devices 110, 112, and 114 include an *electronic earphone located at a head of a talking person* because the *talking person* holds the device at the *head*. These devices also *captur[e] sound* because they include microphones that pick up audio signals representing the participant’s voice. *Id.*

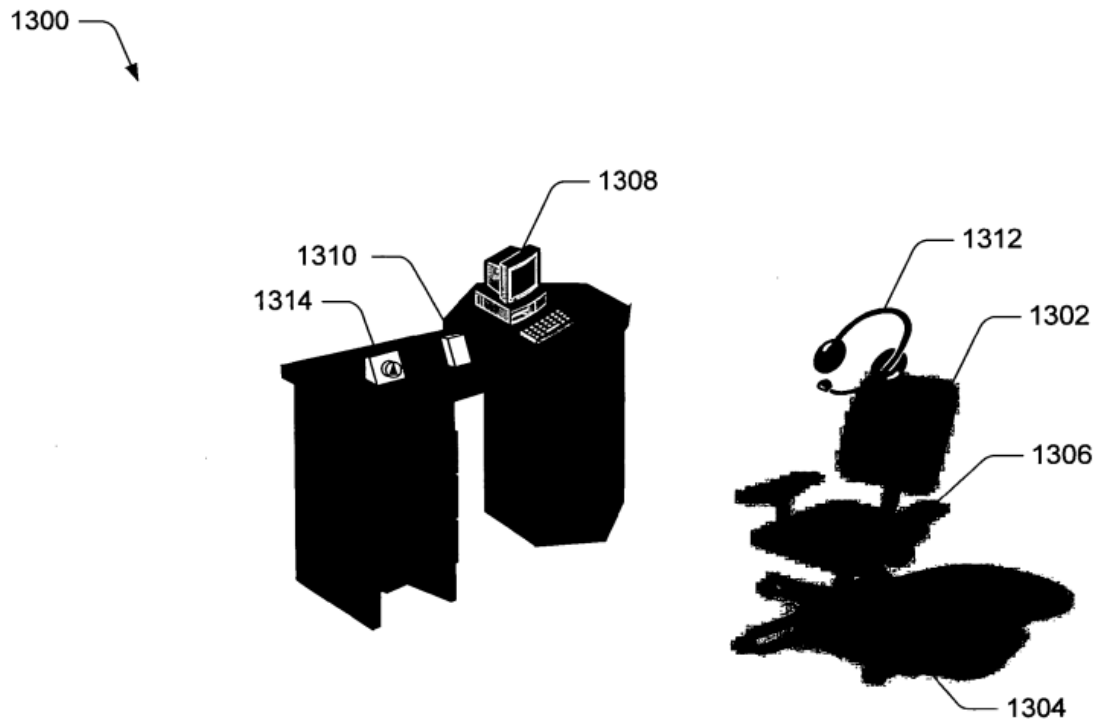


Fig. 13

120. Bedingfield further provides an example of the device used by the participant 104 (a *talking person*) is the “wearable speakers 1312” such as “stereo headphones,” which include an *electronic earphone* that is *located at a head of a talking person*. Ex. 1005, 7:43-46, 13:34-37 (“The environment 1300 may also include wearable speakers 1312. In some embodiments, the wearable speakers 1312 may be stereo headphones. In other embodiments, the speakers may be integrated in other forms as wearable speakers.”), *see also* Figs. 14-15 (shown below depicting each participant in a conference call utilizing wearable speakers), 14:11-23, 13:34-47.

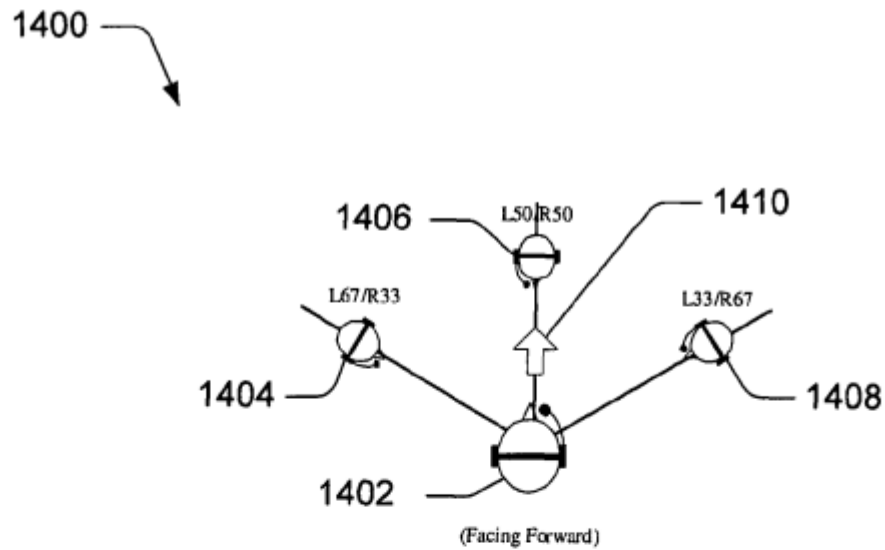
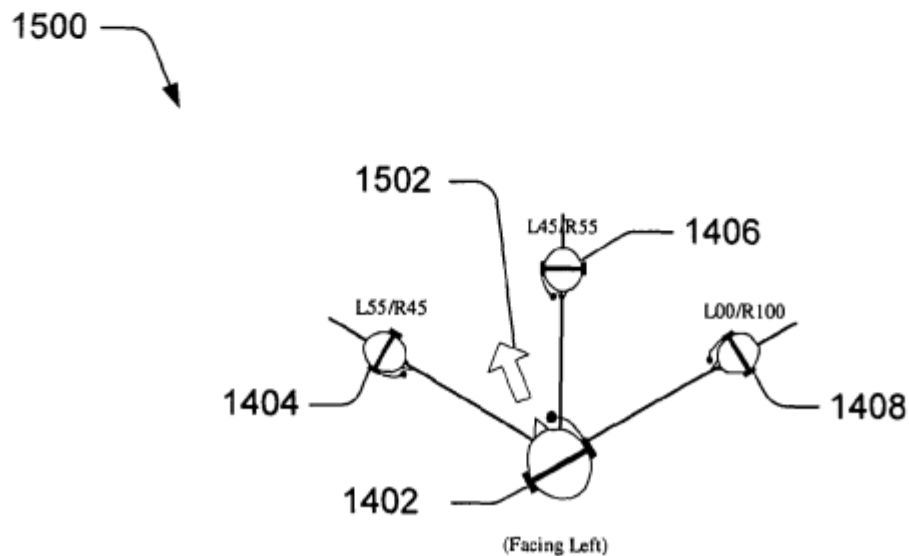


Fig. 14



121. Bedingfield further discloses that *sound will be provided to a listening person during a telephone call*. These steps are performed as part of “[a]n improved

telephone or network-based communication system.” *Id.*, at 3:25-26. Bedingfield processes the recorded sound from the participant 104 through the multi-channel processing device 118 and outputs it (*sound will be provided*) over a multi-channel speaker system 120 that “enable[s] the listener 102 [*listening person*] to hear the speaking participant [*talking person*] in the sound field 122” as left and right sounds *during a telephone call*. *Id.*, 4:18-19, 4:49-51. Because the “communication devices 110, 112, 114 may be *telephones*, *wireless phones*, *mobile phones* ... *voice-over-IP (VoIP) devices* ... or other communication devices that permit multi-directional communication, and are connected to the communication system 116, Bedingfield describes a *telephone call* in reference to Figure 1. *See id.* 4:25-30.

122. Bedingfield discloses that “the communication system 116 may provide a two-channel, or multi-channel signal, to the multi-channel processing device 118,” and therefore Bedingfield contemplates the *sound* has two channels. *Id.* 4:45-47. I note that 8KHZ’s expert in the parallel litigation has testified that it was known before 2014 to provide binaural sound through earphones. Ex. 1032:176:3-5. Bedingfield does not expressly teach *capturing a binaural sound*, but this would have been obvious in view of Andrea.

123. Andrea discloses earbuds (*electronic earphone*) worn by a user that have “audio receiving means on each earbud [that] may pick up the same sound emanating from a single sound source (i.e., the respective audio receiving means

may create a binaural recording).” Ex. 1007, 14:55-58, *see also id.*, 11:66–12:3, 3:28-31, 1:44-51, 14:55-58. “Because of the spatial discrepancy between each of the audio receiving means, a distinct audio signal may be produced in each of the channels corresponding to the same sound.” *Id.*, 14:59-61. By “creat[ing] a binaural recording,” Andrea discloses that its earbuds *captur[e] binaural sound*, which refers to the method of audio recording replicating how a listener would hear sound in the real-world, by recording how sound is received differently at each ear. For example, Andrea explains that because of the difference in the right and left ear audio signal “a user wearing the dual-earbud device will be able to perceive the location from which the sound was originally produced during the playback through the audio outputting means (e.g. speakers).” Ex. 1007, 15:5-8; *see also* 14:47-50 (“Positioning audio receiving means on each earbud creates a dual-channel audio reception device that may be used to create desirable audio effects.”). According to Andrea, the benefit of the “dual-channel earbud embodiment” is that it “allows for the application of audio enhancement techniques, such as active noise reduction,” which “refers to a technique for reducing unwanted sound.” Ex. 1007, 15:18-31, *see also id.*, 11:8-14.

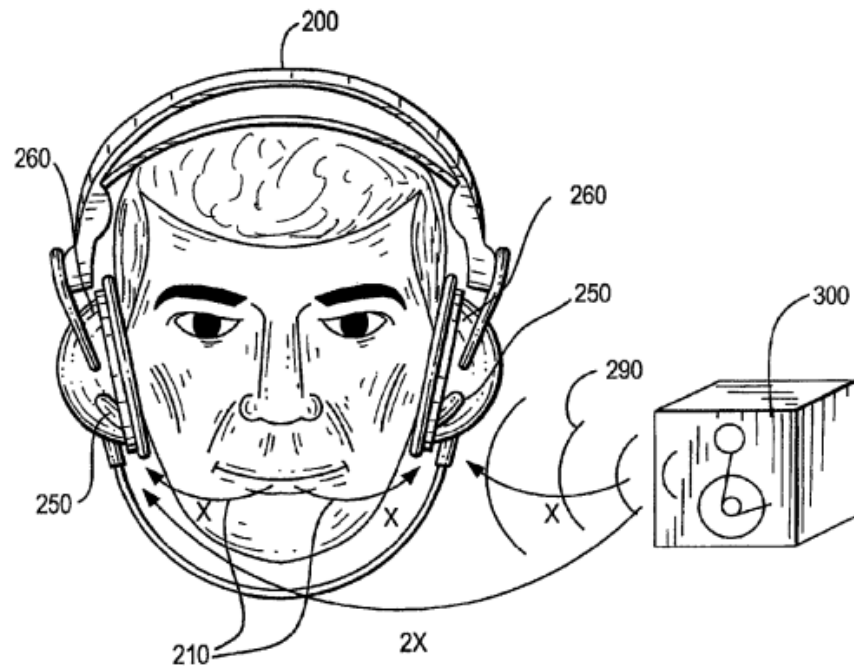


FIG. 10

124. Similarly, Andrea teaches a headset with microphones equidistant from the user's mouth. *Id.*, Fig. 10. This headset is able to cancel noise because the voice signal will travel the same distance to each microphone where as other noise will vary in distance travelled. “[N]oise coming from a speaker 300 on one side of the user is cancelled due to varying distances (X, 2X) of the sound waves 290 from either microphone 250.” *Id.*, 11:14-17. In contrast, “the voice signal 210 travels an

equidistance (Y) to both microphones 250, thus providing for a high fidelity far field noise canceling microphone that possesses good background noise cancellation and that may be used in any type of noisy environment, especially in environments where a lot of music and speech maybe present as background noise (as in a game arena or internet café).” *Id.*, 11:17-2.4

125. A POSA implementing Bedingfield would have been motivated to *capture binaural sound* taught by Andrea because doing so would have enhanced the sound quality by reducing noise. *See* Section IX.A (Ground 1, Motivation to Combine). A POSA would have reasonably expected to successfully modify the Bedingfield communication system by adding an additional microphone to Bedingfield’s *electronic earphone* and a software module to process the *binaural sound* as taught by Andrea. *See* Section IX.A (Ground 1, Motivation to Combine).

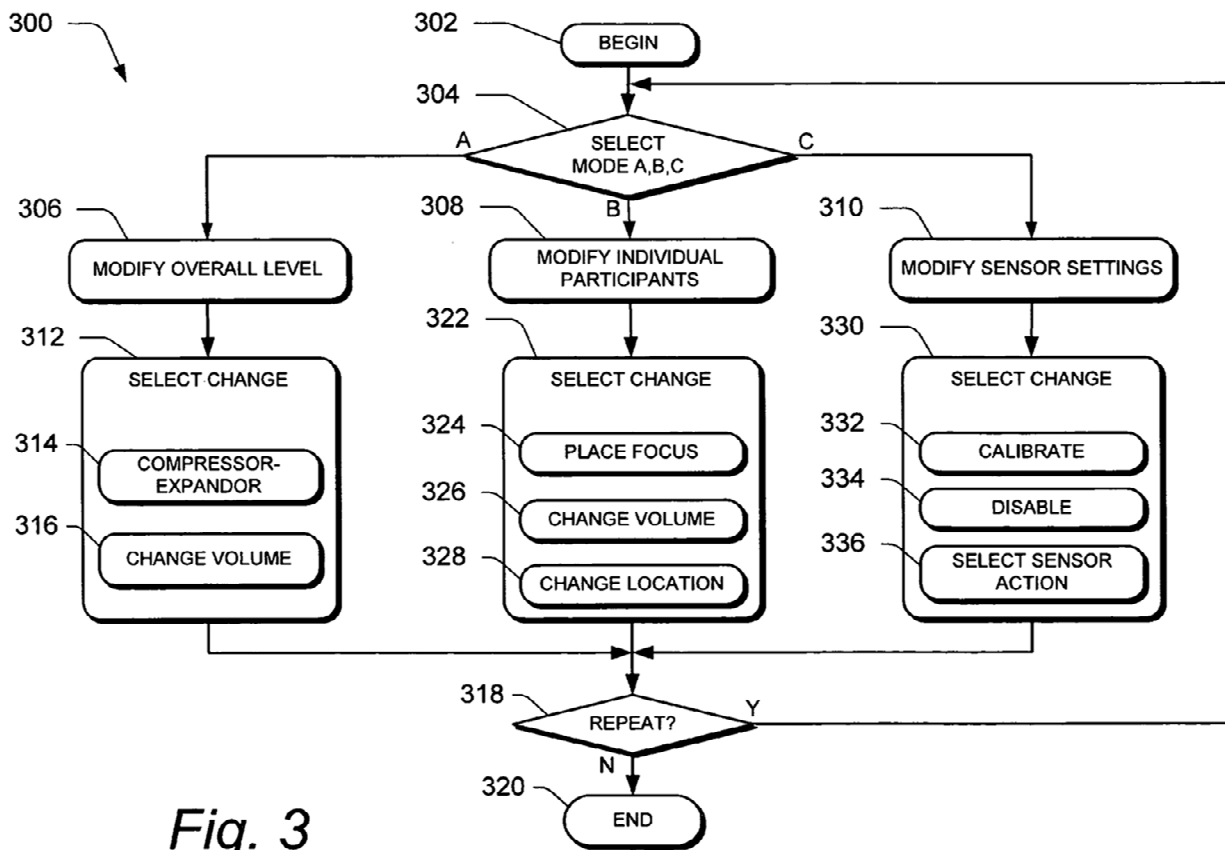
3. 1[b]: “*designating, with a computer system, a sound localization point in empty space that is away from and proximate to the listening person such that the sound localization point simulates an origin of the binaural sound at the empty space that the listening person hears during the telephone call;*”

126. It is my opinion that Bedingfield in view of Andrea discloses this limitation.

127. Bedingfield discloses *a computer system* in the form of a “multi-channel processing device 118 configured with the communication system 116,” which together “creates the sound field 122 that is substantially similar... to an actual

sound field where the listener 102 is present between all of participants 104, 106, 108.” Ex. 1005, 4:66-5:3. In my opinion, this means that the *binaural sound* captured from, for example, participant 104 appears to have an *origin* at the location 104a (*empty space*). As Bedingfield further describes, the *computer system* simulates the sound field 122 so that the listener’s perception of the call is as if the listener and participants “were in the same location and speaking to each other directly.” *See id.*, 4:52-61.

128. In order to achieve this effect, Bedingfield discloses with reference to Figure 3 “a block diagram of a method 300 of how synthetic audio placement may be organized” using multi-channel processing device 118 (part of the *computer system*). *Id.*, 5:51-52.



129. According to method 300, the listener (*the listening person*) provides input to the multi-channel processing device 118 to “select the mode for modifying individual participants via route B at the block 308.” *Id.* 6:27-29. For example, the listener selects to “change a participant’s location,” depicted as locations 104a, 106a, and 108a in Fig. 1 (*designating an SLP*). *Ex.* 1005, 6:29-33, *see also id.*, 6:43-45.

130. A POSA implementing Bedingfield would have found it obvious to position the *SLP in empty space*, or at a location where no physical object exists. *See Ex.* 1005, 4:40-5:3, *see also* 9:60-64 (describing a voice “occupy[ing] a ‘vacant’

location”). For example, Bedingfield describes various sound field configurations, such as the circular configuration in Fig. 19a, and Bedingfield also describes a placement of participants as if they were “around a circular conference table.” Ex. 1005, 9:20-23, 16:36-48.

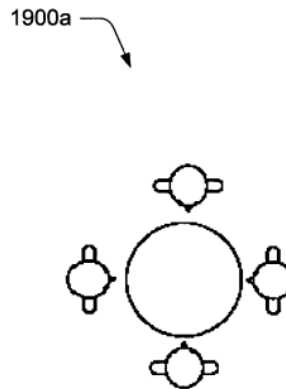


Fig. 19a

131. The POSA would have found it obvious to seat the listener (*listening person*) at a physical conference table, where an empty seat at the table (*empty space*) is allocated for a virtual participant, as was commonly done in the field.

132. In modern working environments, teleconferences are a common form of group communication that allow professionals to communicate and interact from home, in the office, and around the world. Ex. 1037, 1, 4. Teleconferences often involve a mix of participants participating by phone as well as in person. *Id.* 9-10. In such circumstance, it can aid in communication to provide a physical

representation of those participants participating by phone. *Id.* One example of a physical representation is found in Vertesi's Fig. 2, which shows empty chairs representing phone participants:



Vertesi, Fig. 2

133. Similar to Vertesi, Torrejon involves a hybrid conference environment with participants (students in this case) communicating in person and remotely. Ex. 1038, 1, 5. Torrejon introduces an added element: augmented reality. *Id.*, 1. “Augmented Reality . . . allowed students to interact with objects fellow students and the lecturer over great distances.” *Id.* Those participating by phone or other remote communication means select a seat that is not occupied by a physically present student (i.e. an empty chair). *Id.*, 5, 6. The phone participant's image is then projected in the location of the selected seat. *Id.*, 6, 7. During the call, voices of remote participants are produced such that they appear to originate from the projected image, which is located in the physically empty chair. *Id.*, 6-8.

134. In order to mimic an in-person conference, a POSA would place users in locations where conference attendees would normally occupy, like a chair around a conference table. If that chair is already occupied by a person physically present or by another SLP, a conference participant would become confused as to who is speaking. Therefore, each SLP would be placed at a location devoid of people or other SLPs. In addition, some applications attempting to mimic a conference room environment may include a visual representation of a participant at the same location of that participant's SLP. Visual representations would preclude SLP locations already occupied by other objects, and thus, a POSA would place the SLP at an empty space.

135. Positioning the virtual participant's *SLP* at the empty seat at the physical table (*empty space*) would have had the advantage of ensuring that the virtual teleconference environment achieves Bedingfield's stated goal to be as realistic as possible. Moreover, because each of Bedingfield's participants has its own location in the sound field that is separate from the listener 102's location and at a distance, each participant's *SLP* is *in empty space that is away from and proximate to* (under 8KHZ's position) *the listener 102 (listening person)*.

136. Bedingfield's simulated sound field 122 provides audio to the listener that is substantially similar, for example in "specific balance, volume, and timbre for each speaker" (Ex. 1005, 4:52-5:3), "to an actual sound field where the listener 102

is present between all of participants 104, 106, 108” (*simulates an origin of the sound at the empty space that the listening person hears*). *Id.* As discussed for 1[a], Bedingfield discloses the conference call is a *telephone call*, and Bedingfield in view of Andrea discloses the sound is a *binaural sound*.

4. 1[c]: “*adjusting, with the computer system, the binaural sound captured at the earphone of the talking person so the binaural sound originates during the telephone call from the sound localization point in empty space that is away from and proximate to the listening person; and*”

137. It is my opinion that Bedingfield in view of Andrea discloses this limitation.

138. Bedingfield’s multi-channel processing device 118 (part of *the computer system*) receives the audio signals (*sound*) and “process[es] the signals” (*adjust[s] the sound*) to “creat[e] a sound field 122” so that “the listener 102 [] hear[s] the speaking participant in the sound field 122.” Ex. 1005, 4:33-51. Specifically, as explained I explained above in the Background (Section VI), it was commonplace in the field to process sounds using HRTFs so that the source of the sound would appear as if it originated outside the listener’s head, and thereby create Bedingfield’s sound field. This processing is *adjusting* the binaural sound, because it involves transforming the sound so that it appears to originate from a specific location. As explained for 1[a], those audio signals (*sound*) are *captured at the earphone of the talking person* and occur (and are provided) *during a telephone call*,

and Bedingfield in view of Andrea discloses the sound is a *binaural sound*. As explained for 1[b], Bedingfield's sound field is created to simulate the sound to *originate[] from the SLP in empty space that is away from and proximate to the listening person*.

5. 1[d]: “*providing, with an electronic earphone located at a head of the listening person, the binaural sound to the listening person during the telephone call such that the origin of the binaural sound for the listening person occurs at the sound localization point in empty space that is away from and proximate to the listening person.*”

139. It is my opinion that Bedingfield in view of Andrea discloses this limitation.

140. Bedingfield discloses the multi-channel speaker system 120 as a wearable headset or headphones (*electronic earphone located at a head of the listening person*). *Id.* Fig. 1. As explained for 1[a], the audio signals (*sound*) are *provid[ed] to the listening person during the telephone call*, and Bedingfield in view of Andrea discloses the *sound is binaural sound*. As explained for 1[b], the *sound is provided such that the SLP simulates the origin of the sound for the listening person occurs at the SLP in empty space that is away from and proximate to the listening person*.

D. Dependent Claims 2, 3, 6¹

1. **Claim 2[a]:** “*The method of claim 1 further comprising: tracking a head orientation of the listening person with respect to the sound localization point;*”

141. It is my opinion that Bedingfield discloses this limitation.

142. Bedingfield discloses “a number of sensors to detect [*track*] the movement of... listener 104 in FIG. 1.” Ex. 1005, 12:62-67. Specifically, with reference to Fig. 13, Bedingfield describes “wearable speakers 1312” (depicted as a headset) that are worn on the *head* of the listener 104 (*the listening person*) and “detect [*track*] the orientation of [the] listener 104.” *Id.*, 13:38-39.

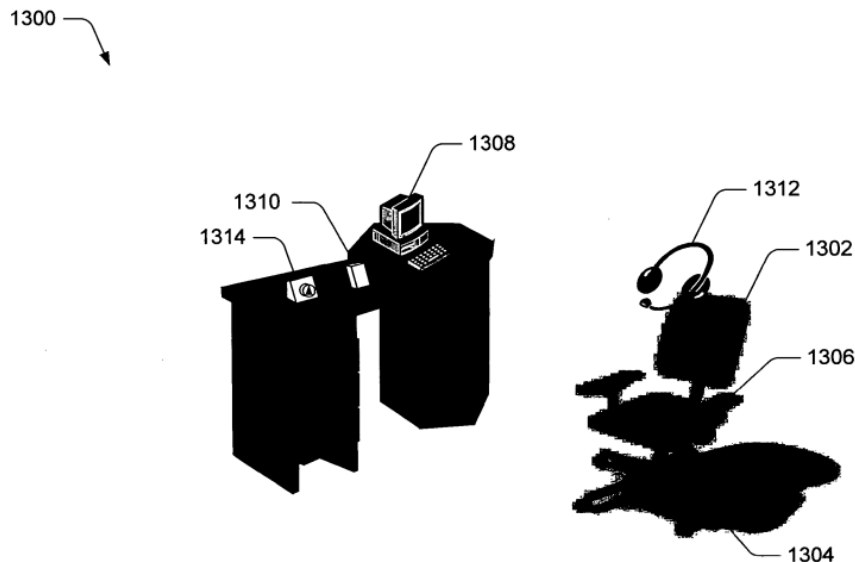


Fig. 13

¹ All dependent claims incorporate the analysis for the claims from which they depend.

143. Bedingfield also describes the wearable speakers 1312 as “sens[ing] movements of the listener’s 104 head” and the direction the listener is facing via “motion detectors, gravity based sensors, or other movement detection devices, including physical, electronic, and image based detection devices.” *Id.*, 13:39-57. By being worn on the head and detecting orientation and movement of the listener’s head, Bedingfield discloses *tracking a head orientation of the listening person*.

144. Bedingfield’s *tracking* of the *head orientation* is also *with respect to the SLP*. Specifically, as explained for 1[b], Bedingfield discloses that the sound field 122 has perceived locations (*SLP*), such that the listener 102 perceives the sound “as if the listener was present between all participants.” Ex. 1005, 4:55-66. In Bedingfield’s system, the perceived locations (*SLP*) are fixed, such that the orientation and movement of the listener’s head is *with respect to the SLP*. *Id.* 13:35-52 (“[I]f the listener is facing the receiver 1310, then the listener may be orientated in a first position within the sound field 122. If the listener moves to a second position (such as by turning his head), as detected by the receiver 1310 from movement (e.g., rotation) of the wearable speakers 1312, then the listener 104 may be orientated in a second position within the sound field 122.), *see also* Figs. 14-15, depicting a listener moving to change focus direction on different participants while the sound field changes “to reflect the new sound field orientation” for the listener (*id.*, 14:20-43).

2. **Claim 2[b]:** “*adjusting an interaural time difference of the sound that the listening person hears in response to the listening person changing the head orientation with respect to the sound localization point.*”

145. It is my opinion that Bedingfield discloses this limitation.

146. As explained for 2[a], Bedingfield orients the listener (*listening person*) differently within the sound field 122 depending on (*in response to*) the listener’s head turning from a first to second position (*the listening person changing the head orientation with respect to the SLP*).

147. Bedingfield also discloses *adjusting an ITD of the sound that the listening person hears* in order to orient the listener within the sound field 112. Bedingfield uses “various audio techniques” such as “delay” to place the different participants in different “configurations within the sound field.” Ex. 1005, 5:35-41. Bedingfield further describes applying “changes to delay” to enhance the “impression that a sound source has moved” relative to the listener. *Id.*, 3:58-59. A fixed SLP and a moving listener is functionally equivalent to a stationary listener and a moving SLP. Furthermore, a POSA would have readily understood that the “delay” in this context is an *ITD*, which is the delay in arrival time of a sound between two ears and is commonly used to manipulate the perceived origin of a sound source in an audio scene. Thus, Bedingfield discloses *adjusting an ITD of the sound* delivered to the listener.

3. **Claim 3:** “*The method of claim 1 further comprising: receiving, from the listening person, a location that specifies a distance from the listening person where to position the sound localization point in empty space that is proximate to and away from the listening person.*”

148. It is my opinion that Bedingfield discloses this limitation.

149. As discussed above for 1[b], the listener (*listening person*) provides input to “change a participant’s location” in the sound field 122, depicted as locations 104a, 106a, and 108a in Fig. 1. See Ex. 1005, 6:29-45. Thus, Bedingfield’s system *receiv[es], from the listening person, a location for where to position the SLP in empty space that is proximate to and away from the listening person*. By tracking the orientation and relative location of the *listening person* and the *listening person* selecting a location for a participant, Bedingfield’s *listening person* provides a location that specifies a distance from the listening person within the sound field. See Ex. 1005, 6:29-45;

4. **Claim 6[a]:** “*The method of claim 1 further comprising: capturing binaural sound from a second talking person;*”

150. It is my opinion that Bedingfield in view of Andrea discloses this limitation.

151. As explained for 1[a], each of Bedingfield’s participants 104, 106, and 108 is a *talking person*, and participant 104 is a first *talking person*. For the same reasons as explained for 1[a], participant 106 is a *second talking person*, and it would

have been obvious to implement Bedingfield to *captur[e] binaural sound* from the participant 106 (*second talking person*).

5. **Claim 6[b]:** “*designating a second sound localization point in empty space that is away from and proximate to the listening person such that the second sound localization point simulates an origin of the binaural sound from the second talking person that the listener hears;*”

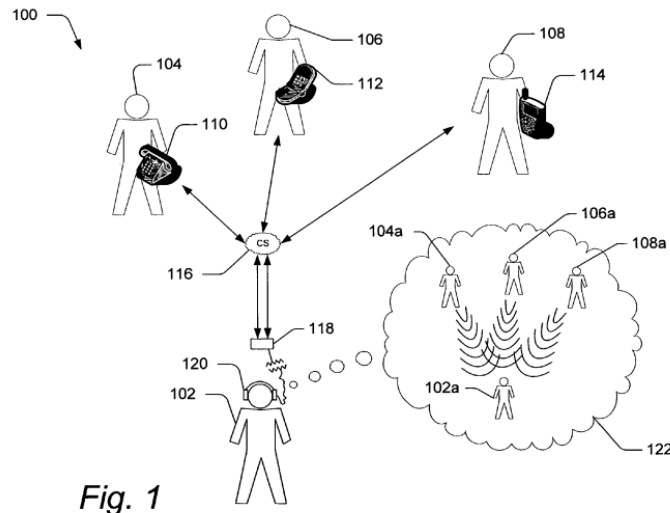
152. For the same reasons as explained for 1[b], Bedingfield *designates* location 106a (*a second SLP in empty space that is away from and proximate to the listening person*) that *simulates and origin of the binaural sound from the participant 106 (second talking person) that the listener hears*.

6. **Claim 6[c]:** “*positioning the second sound localization point adjacent to the sound localization point such that the binaural sound captured from the talking person appears to originate from the sound localization point and the binaural sound captured from the second talking person appears to originate from the second sound localization point.*”

153. It is my opinion that Bedingfield discloses this limitation.

154. As Bedingfield depicts in Fig. 1, the location 106a (*second SLP*) is *position[ed] adjacent to* the location 104a (*SLP*). As explained for 1[b], Bedingfield’s sound field is simulated to provide the perception “as if the listener was present between all participants 104, 106, and 108.” Ex. 1005, 4:52-66. That is, the *binaural sound captured from participant 104 (the talking person) appears to originate from the location 104a (SLP) and the binaural sound captured from*

participant 106 (*the second talking person*) appears to originate from the location 106a (*second SLP*).



E. Independent Claim 15

1. 15 ``` [pre]: “A computer system, comprising:” ```

155. To the extent limiting, Bedingfield discloses the preamble.

156. Bedingfield discloses *a computer system* as a communication system 116, a multi-channel processing device 118, communication devices 110, 112, and 114, and multi-channel speaker system 120. Ex. 1005, 4:31-51, Fig. 1; *see also* Section IX.C.3.

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

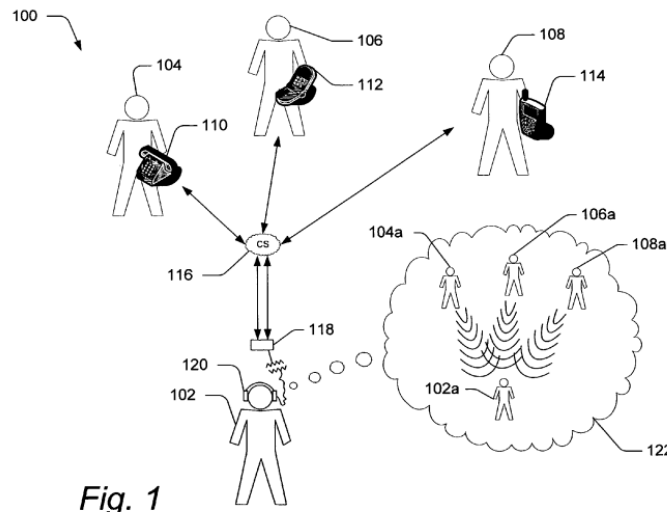


Fig. 1

2. **15[a]:** “a first wearable electronic device that a first person wears;”

157. It is my opinion that Bedingfield discloses this limitation.

158. Bedingfield’s *computer system* includes multi-channel speaker system 120, which as explained for 1[d], is a wearable headset or headphones (*a first wearable electronic device*) that the listener 102 (*a first person*) wears. Ex. 1005, 4:18-19, 4:31-33, 4:49-51, 7:43-46, 13:34-37, Fig. 1; *see also* Section IV.C.5.

3. **15[b]:** “a second wearable electronic device that a second person wears and that captures a binaural voice of the second person during an electronic phone call with the first wearable electronic device that the first person wears;”

159. It is my opinion that Bedingfield in view of Andrea discloses this limitation.

160. Bedingfield's *computer system* includes communication device 114 (*a second wearable electronic device*) that participant 108 (*a second person*) wears, and that, as explained for 1[a], *captures* a sound of participant 108 (*second person*) during a telephone call (*an electronic phone call*) with the multi-channel speaker system 120 (*the first wearable electronic device*) that the listener 102 (*first person*) wears. As further explained for 1[a], it would have been obvious from the teachings of Bedingfield and Andrea to implement Bedingfield's communication device 114 (*second wearable electronic device*) to *capture[] binaural* sound from participant 108 (*second person*), where that sound is the participant's *voice*. See Ex. 1005, Fig. 1 (reproduced below with my annotations in red and blue).

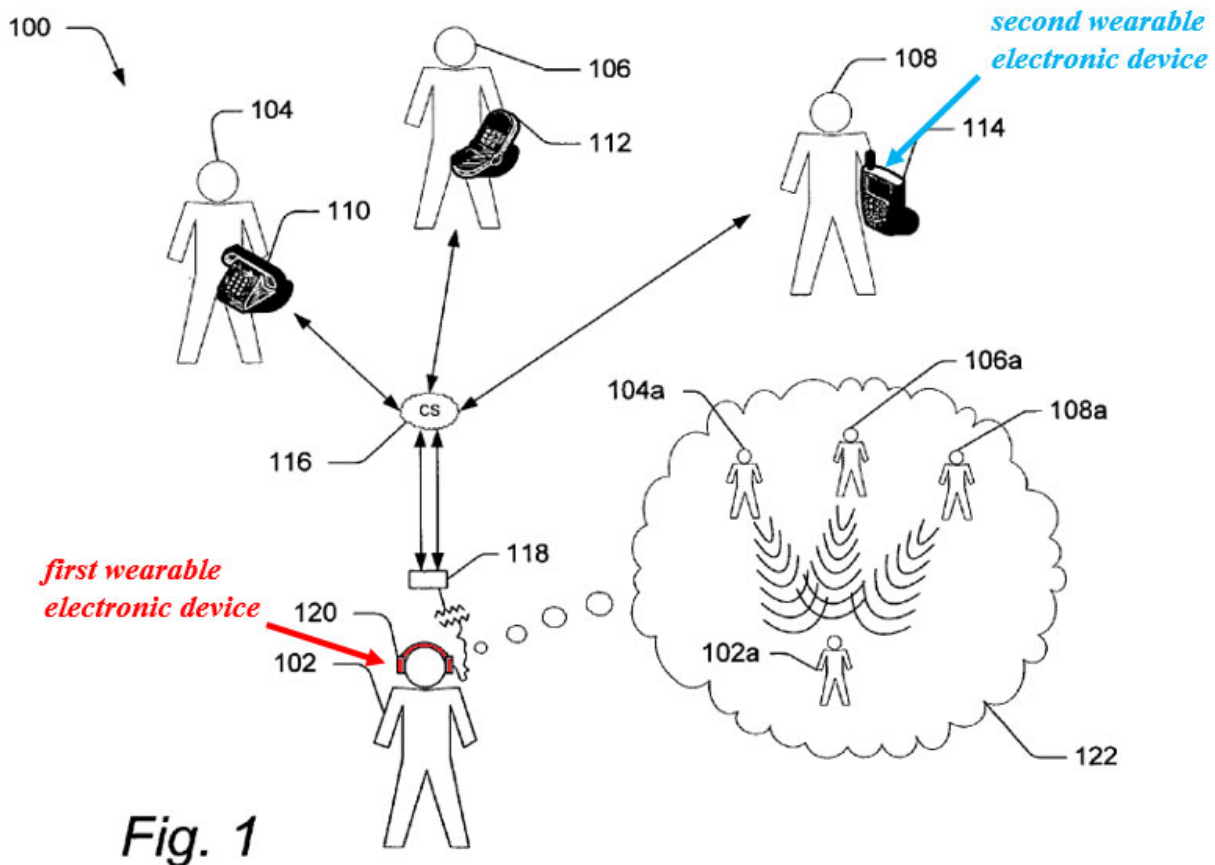


Fig. 1

4. 15[c]: “a memory that stores instructions; and”

161. It is my opinion that Bedingfield discloses this limitation. Bedingfield states that “[g]enerally, any of the functions described herein can be implemented using software,” for which “program code can be *stored* in one or more computer readable *memory* devices.” Ex. 1005, 11:8-22, *see also* 11:23-36 (describing “computer executable *instructions*” as being stored in “local and remote computer storage media, including *memory* storage devices”), 12:5-44 (describing a processor that reads “executable *instructions* from... the storage medium”), Fig. 12. Such

functionality includes Bedingfield's description of rendering a sound field and providing synthetic audio placement within that sound field. Ex. 1005, 1:15-17, 5:51-52, Fig. 3. Bedingfield therefore contemplates that the components that provide that functionality, including communication system 116, the multi-channel processing device 118, communication devices 110, 112, and 114, and the multi-channel speaker system 120 (together, *the computer system*) would have included *a memory that stores instructions*.

5. 15[d]: “a processor that executes the instructions to...”²

162. It is my opinion that Bedingfield discloses this limitation. As explained for 15[c], “any of the functions” in Bedingfield, including rendering of a sound field and providing synthetic audio placement, “can be implemented using software.” Ex. 1005, 11:8-22, *see also* 11:23-36, 12:5-44, Fig. 12. That software takes the form of “program code that performs specified tasks when executed on *processor(s)* (e.g., any of microprocessors, controllers, and the like)” and “may be implemented on a variety of commercial computing platforms having a variety of *processors*.” *Id.*, 11:13-22; *see also* 11:60-12:16 (describing processors coupled to user interfaces and reading executable instructions from memory), 12:25-43 (describing processors

² I note that 8KHZ's expert in the parallel litigation has admitted that the POSA would have known how to implement 15[d], 15[d1], and 15[d2]. Ex. 1032, 148:22-149:14.

executing software modules). Therefore, the components that provide the functionality of sound field rendering and synthetic audio placement, namely Bedingfield's communication system 116, multi-channel processing device 118, communication devices 110, 112, and 114, and multi-channel speaker system 120 (together, *the computer system*), include *a processor that executes the instructions*.

6. **15[d1]:** *“adjust the binaural voice of the second person during the electronic phone call in order to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person,”*

163. Bedingfield discloses processing (*adjust*) the audio signals (*voice*) of the speaking participant (*second person*) during the conference call (*electronic phone call*) in order to position a perceived location 104a, 106a, 108a (*SLP*) of the audio signals (*voice*) to a location that is away from and proximate to the listener (*first person*). As explained for 15[b], that *voice* is a *binaural* voice, as taught by Andrea.

7. **15[d2]:** *“wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.”*

164. It is my opinion that Bedingfield in view of Andrea discloses this limitation.

165. As explained for 1[a], 1[d], 15[b], and 15[d1], Bedingfield's locations 104a, 106a, and 108a are each an *SLP* that *represents an origin of the binaural* sound

of the participant (second person) to the listener (first person), and that binaural sound is the binaural voice from the participant.

F. Dependent Claims 16-17

- 1. Claim 16:** *“The computer system of claim 15, wherein the processor further executes the instructions to determine a location of an empty chair that is proximate to the first person and position the sound localization point at the empty chair.”*

166. In my opinion, Bedingfield renders obvious this limitation.

167. As I already explained above for 15[d]-15[d1] in this ground, it is my opinion that the Bedingfield system discloses *a processor that executes instructions... to position the SLP* at a location *that is proximate to the first person*, and Bedingfield’s locations 104a, 106a, and 108a are *SLPs*. See Sections IX.C.5-6 (Ground 1, 15[d]-15[d1]). Moreover, as I already explained above for 1[b] in this ground, the POSA would have known to position Bedingfield’s *SLP* for speaking participants as if they were at a physical table, where an *empty seat or chair* is allocated for the virtual participant. See Section IX.C.3 (Ground 1, 1[b]); see also Ex. 1005, 9:20-23, 16:36-48, Fig. 19a (reproduced below).

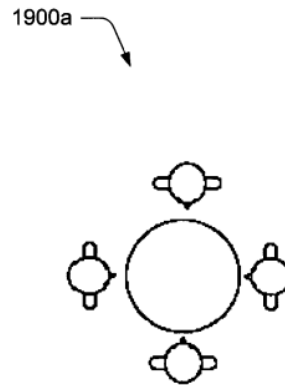


Fig. 19a

168. Not only was it well-known to seat participants in *chair[s]* at a conference table, but Bedingfield even states the sound field is provided to remote participants as if they were arranged “around a circular conference table.” Ex. 1005, 9:20-23, 16:36-48, Fig. 19a (reproduced above); *see also* teachings in Ex. 1037 and 1038 for providing spatialized audio in hybrid in-person and remote environments, including where remote participants are assigned an empty chair in the space. As the POSA would have readily understood, conference participants are typically seated in chairs around conference tables, and it would have been obvious to position the *SLP* for a remote participant at an *empty chair*, and therefore *determine a location of an empty chair*, and *position the SLP at the empty chair*.

169. Furthermore, Bedingfield explains that a user may enter a user input to select a location of for a participant *SLP*. EX1005, 6:43-45, 11:66–12:4. When, as

in Bedingfield's Figs 9a and 9b, Bedingfield's computer system receives a user input indicating a request to place the SLP at an *empty chair*, Bedingfield's computer system then *determine[s]* the *location* of the *empty chair* as the target for the *SLP* and then *position[s]* the SLP at the *determine[d]* *empty chair location*. *Id.* 1:62-63 (“*determin[e]* the synthetic position associated with the providing audio source”). Therefore, in my opinion, a POSA would have known that a user selection would result in Bedingfield's *computer system execute[ing] instructions to determine a location of an empty chair*.

2. **Claim 17:** “*The computer system of claim 15, wherein the processor further executes the instructions to change an interaural time difference (ITD), an interaural level difference (ILD), and a head-related transfer function (HRTF) in order to position the sound localization point of the binaural voice to the location that is away from and proximate to the first person.*”

170. Bedingfield renders obvious this limitation.

171. As explained in 15[d], Bedingfield and Andrea discloses a *processor that executes instructions... to position the SLP of the binaural voice to the location that is away from and proximate to the first person*. Section IV.E.5-7.

172. As explained for claim 2, Bedingfield further discloses that *ITD* is a cue that provides directionality and places a sound source in the sound field. Bedingfield discloses another such cue is an *ILD*. As Bedingfield explains in relation to Fig. 6, to provide directionality in the sound field, the system applies a level difference

between left and right ears, which is an *ILD*. Ex. 1005, 9:48-57, *see also id.*, 3:39-40 (“A sound whose volume is stronger in the left ear appears to be coming from a position to the left of the listener.”), Table 1, 14:35-43 (describing providing balanced volume between the left and right ears to reflect a new sound field orientation).

173. Bedingfield further discloses *HRTF*, by describing a “way the human ear detects the position of a signal is a change in frequencies heard as a sound moves from the front of a listener’s ear, where most frequencies are received with equal strength, to a position behind the listener’s ear, where higher frequencies are attenuated and the lower frequencies seem more apparent.” Ex. 1005, 3:49-54. A POSA would have understood that Bedingfield’s reference to using relative strength and attenuation of different frequencies (for each direction) of a sound for sound localization is the same as an *HRTF*, which was well-known in the art.

174. Thus, based on Bedingfield’s teachings and the general knowledge in the art, a POSA would have known that the *ILD*, *ITD*, and *HRTF* are all audio techniques that would be *changed* in Bedingfield to *position the SLP* at the desired location.

**X. GROUND 2: BEDINGFIELD IN VIEW OF ANDREA AND BALLAS
RENDERS OBVIOUS CLAIM 18**

A. Motivation to Combine Bedingfield and Andrea with Ballas

175. It is my opinion that it would have been obvious to a POSA to modify the system resulting from the combination Bedingfield and Andrea to track the user's head orientation and adjust the sound so that the sound appears fixed even as the user's head moves, as taught in Ballas.

176. It is my opinion that a POSA would have been motivated to implement this modification for the following reasons.

177. First, adjusting the sound to be fixed even as the user's head move was a natural extension of the Bedingfield-Andrea system. Bedingfield already recognized that the sound field should be adjusted based on the orientation of the listener. Specifically, Bedingfield already discloses "detect[ing] the orientation of a listener 104" and that "the sound field may correspondingly change to reflect the new sound field orientation." Ex. 1005, 13:38-39, 14:35-38. In fact, in Bedingfield's system, the listener provides manual input to change the orientation, which causes the sound field to adjust so that it appears fixed to the listener. *Id.*, 14:11-43; Figs. 14, 15.

178. In my opinion, the POSA would have been motivated to combine the teachings of Bedingfield-Andrea and Ballas because a POSA would have recognized

that each of these references has a shared goal of providing a realistic experience for the user. To achieve that goal, the POSA would have known to track the user's head orientation and movements and provide Bedingfield's sound field in a manner that accounts for the user's orientation and movements, using conventional head tracking and spatialization techniques expressly taught in Ballas. Therefore, in my opinion, a POSA would have been motivated to perform such an implementation. Doing so would have had the benefits I describe below.

179. Bedingfield's primary goal is to provide a sound field that "closely simulate[s] an in-person conference setting." Ex. 1005, 3:25-27. Ballas shares the same goal. Ex. 1008, ¶¶2, 9, 10. In Bedingfield, the system already detects the listener's orientation and changes the sound field accordingly. Ex. 1005, 1:56-59, 13:38-39, 13:58-64, 14:32-38, 18:49-52. As the POSA would have appreciated, it would have been key for simulating an in-person conference setting, to track the listener's head and change the sound field so that the sound sources appear fixed to the listener, even while the listener's head moves. Modifying Bedingfield-Andrea in accordance with Ballas's teachings of head tracking and adjusting the sound field to appear fixed to the user would have had the benefit of allowing the user to move while staying connected to the event. This would have achieved Bedingfield's express purpose to simulate an in-person conference setting.

180. Bedingfield already discloses certain hardware devices to detect the listener's orientation, such as motion detectors, sensors, and other similar devices. Ex. 1005, 13:54-56. Therefore, the modification would have been a natural extension of Bedingfield. Specifically, Ballas discloses head tracking and "maintaining virtual location of the audio signals." Ex. 1008, Abstract, ¶18, *see also id.*, ¶¶17, 19-20, 25-26, claims 1, 7. Using Ballas's head tracking system would have allowed the user to change head location and orientation while maintaining fixed virtual sound locations, and this would have made the simulation of an in-person conference setting even more realistic and immersive. Therefore, implementing Ballas's head tracking and maintaining of fixed virtual sound locations would have been desirable to advance the common goals of both Bedingfield and Ballas to provide a realistic environment in a teleconferencing application.

B. Reasonable Expectation of Success

181. In my opinion, the POSA would have had a reasonable expectation of success in modifying the Bedingfield-Andrea system in accordance with the teachings of Ballas. Both Bedingfield and Ballas track the listener's movement and orientation and change the sound field accordingly. Moreover, the hardware in Bedingfield and Ballas is similar in that both include wearable speakers. Specifically, Bedingfield's user wears "wearable speakers" in the form of stereo headphones. Ex. 1005, 13:35-36. Similarly, Ballas's speakers are attached to a

wearable head frame. Ex. 1008, ¶25. Accordingly, it is my opinion that a POSA would have recognized that the Bedingfield-Andrea system and Ballas's system are structurally and functionally similar.

182. In my opinion, a POSA implementing the Bedingfield-Andrea system would have readily recognized that Bedingfield's headphones would have been attached to Ballas's frame. Such an implementation would have merely involved a combination of known prior art elements (headset and head tracking system) according to known methods and techniques (accounting for head movement) to yield predictable results (producing spatialized sound that appears fixed as the user moves), and involves use of a known technique (HRTF) to improve a similar device (wearable communication device) in the same way (filtering sounds to produce a sound field that appears fixed).

C. Dependent Claim 18

1. **Claim 18:** *“The computer system of claim 15, wherein the processor further executes the instructions to change an interaural time difference, an interaural level difference, and a head-related transfer function of the binaural sound in order to maintain the sound localization point fixed at the location during the electronic phone call while the first person physically moves away from the location and changes a head orientation with respect to the location.”*

183. In my opinion, this limitation is obvious from Bedingfield and Andrea, in view of Ballas.

184. As I explained above in Ground 1 for 15[d] and claim 17, Bedingfield and Andrea render obvious *a processor that executes instructions to change an ITD, an ILD, and an HRTF to position the SLP of the binaural voice to a location that is away from and proximate to the first person*, and as explained for claim 17, Bedingfield renders obvious *chang[ing] an ITD, an ILD, and an HRTF of the binaural sound*. Section IX.E.5-7, IX.F.1.

185. Additionally, Bedingfield discloses “wearable speakers 1312 may ... sense movements of the listener’s 104 head....If the listener moves to a second position (such as by turning his head), as detected by the receiver 1310 from movement (e.g., rotation) of the wearable speakers 1312, then the listener 104 may be orientated in a second position within the sound field 122.” Ex. 1005, 13:34-52. Moreover, Bedingfield discloses “adjust[ing] the focus” of the sound field by “chang[ing] the orientation of the sound field 122.” Ex. 1005, 13:58-64, *see also id.*, 6:57-7:2, 4:7-12. Specifically, as Bedingfield describes for Figures 14 and 15 (which I have reproduced below with my added annotations in blue and green), the listener provides a manual input to “change[] the focus [direction], [and] the sound field [] correspondingly change[s] to reflect the new sound field orientation.” *Id.*, 14:32-38.

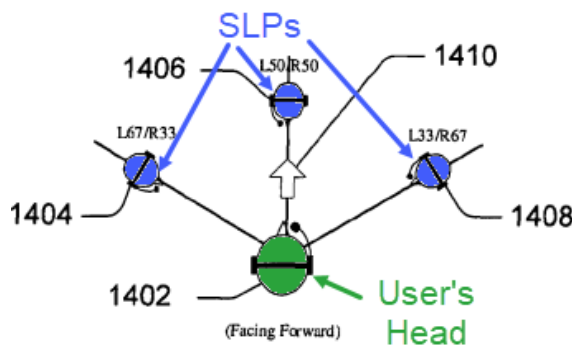


Fig. 14

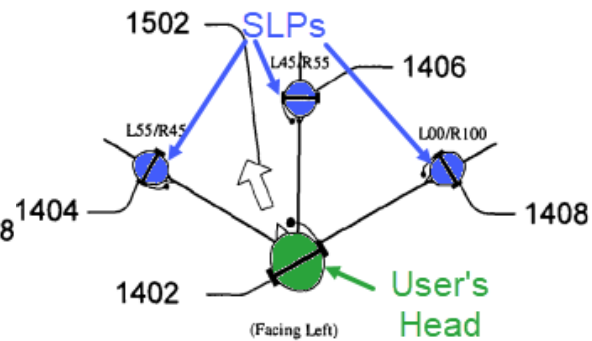


Fig. 15

186. Thus, Bedingfield changes the sound field to reflect the new sound field orientation input by the user, and the sound field appears fixed while the user's head moves.

187. As I explained above, Ballas generally relates to “producing virtual sound sources that are externally perceived and positioned at any orientation in azimuth and elevation from a listener.” Ex. 1008, Title, Abstract, ¶1. Ballas's system is “a virtual reality head mounted visual display system,” such as an eyeglass frame or helmet. *Id.*, ¶25, *see also* Figs. 3, 4.



Figure 3. Eyeglass frame mount prototype

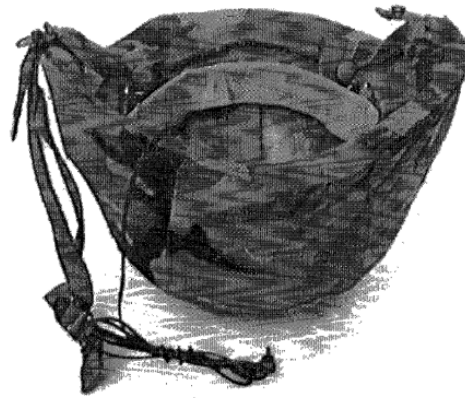


Figure 4. Helmet mount prototype

188. As Ballas specifically states, mounted to Ballas's frame is a "head tracking system 104 [that] measures the location and orientation of a user's head and provides the measured information to the computer system 102." Ex. 1008, ¶25.

189. In Ballas's system, "a pair of head related transfer functions (HRTFs)" are "for a particular location" (*the SLP*). *Id.*, ¶16. More particularly, Ballas discloses that its HRTF filters "maintain[] [the] virtual location of the audio signals thereby allowing the user to change location and head orientation without degradation of the virtual location of audio signals." *Id.*, Abstract, ¶¶17-20, claims 1, 7, 13. In other words, while the user moves and changes location and head orientation, Ballas ensures the perceived source location of the audio signals is maintained, meaning it is fixed in space despite the user's movement. Therefore, by allowing the user to move and tracking the user's head movements and orientation, while maintaining the virtual location of the audio signals, Ballas discloses *maintain[ing] the SLP fixed at the location during the electronic phone call while the first person physically*

moves away from the location and changes a head orientation with respect to the location.

190. In my opinion, the POSA would have been motivated to implement Ballas's wearable frame with its head tracking system and HRTF processing as part of Bedingfield-Andrea's headphones. Such an implementation would have provided a more realistic virtual experience for the user, by allowing the user to move while fixing the perceived origin of the sound sources. *See* Section X.A (Ground 2, Motivation to Combine Bedingfield-Andrea with Ballas).

191. Moreover, I note that 8KHZ's expert in the parallel litigation admitted that it was known in the 1980s for "head-track virtual acoustic systems [to] keep a sound source fixed in a location in virtual space independent of head movement"). Ex. 1032, 250:2-9; *see also* Ex. 1036, 12, 99, 132-34, Figs. 4.37-4.38.

XI. GROUND 3: BEDINGFIELD IN VIEW OF ANDREA AND SANDGREN RENDERS OBVIOUS CLAIM 19

A. Motivation to Combine

192. In my opinion, the POSA would have found it obvious to modify the Bedingfield-Andrea combination to adjust attenuation and reverberation of the sound, so as to add artificial cues of virtual objects that do not actually exist but are perceived to exist, as expressly taught in Sandgren. It is my opinion that a POSA would have been motivated to combine the teachings of Bedingfield-Andrea and

Sandgren, at least because a POSA would have recognized that these references have a shared goal of providing a realistic and immersive experience for the user. In order to achieve this goal, the POSA would have understood to use Sandgren's conventional sound enhancement techniques to simulate sound-reflecting objects that do not actually exist but are perceived to exist in the virtual environment.

193. Sandgren shares the same goal of enhancing realism. Specifically, Sandgren discloses the “electronic creation of virtual three-dimensional (3D) audio scenes” and “increasing the externalization of virtual sound sources presented through earphones.” Ex. 1017, ¶1; *see also id.*, Abstract. To achieve this goal, Sandgren states “it is not enough to concentrate on the direct sound.” Ex. 1017, ¶6. Instead, Sandgren teaches that other cues, including reverberation and attenuation, are important to fully simulate the 3D audio. *Id.*, ¶¶6, 12.

194. Sandgren appreciates that “[a] room’s reverberation depends on... the types of objects present in the room,” and “it is not the main reflections from the floor, ceiling, and walls of a room that result in externalization. Instead, *other objects in the room, e.g., tables and chairs, that scatter sound waves are essential* for good externalization.” Ex. 1017, ¶5-6, 19 (emphasis added). As Sandgren further contemplates, a realistic sound simulator includes attenuation filters combined with reverberation generators. *Id.*, ¶¶62-63, Fig. 5C.

195. This implementation would have enhanced the realism and immersion for the user in the virtual environment, as sound-reflecting objects in the physical world would have affected the sound signals that reached the listener. Thus, to improve realism of the virtual environment, a POSA would have been motivated to implement the Bedingfield-Andrea system to change attenuation and reverberation of the binaural sound to add artificial audial cues of virtual objects.

196. Moreover, in my opinion, the combination would have been a natural extension of Bedingfield-Andrea's system already providing a sound field of spatialized sounds over headphones. Like Bedingfield, Sandgren "simulat[es] an audio scene or creat[e] a virtual audio scene" over headphones. Ex. 1017, ¶¶9-10; *see also* ¶¶11, 15-16, 19, 22, 46, 48-64, 70-100, Figs. 5A-C. Specifically, Sandgren describes "simulated sound-reflecting objects are simulated by processing a direct-sound signal with at least one [HRTF]...and generating simulated early-reflection signals." *Id.*, Abstract. It is my opinion that a POSA would have recognized that the teachings of Sandgren would have been obvious to implement with Bedingfield-Andrea, as they would have enhanced Bedingfield's goal of an immersive experience by introducing attenuation filters and reverberation generators indicative of virtual objects in the environment. This implementation would have advanced the common goals of both Bedingfield-Andrea and Sandgren to provide a realistic and immersive experience for the user.

B. Reasonable Expectation of Success

197. It is my opinion that a POSA would have had a reasonable expectation of success in modifying Bedingfield-Andrea in accordance with the teachings of Sandgren. Both Bedingfield and Sandgren simulate realistic auditory scenes having spatialized binaural sound and provide such sounds over headphones. Sandgren creates virtual 3D audio scenes to externalize virtual sound sources. Accordingly, Bedingfield and Sandgren are structurally and functionally similar.

198. In my opinion, the POSA implementing the Bedingfield-Andrea system would have known that the spatialization of sound would have been accompanied with Sandgren's attenuation filters and reverberation generators, as this would have artificially simulated the presence of virtual objects and enhanced externalization and realism of the virtual environment. Such an implementation would have merely involved a combination of known prior art elements (localization cues with attenuation filters and reverberation generators) according to known methods and techniques (simulate sound-reflecting virtual objects) to yield predictable results (producing sound so that the user perceives the virtual objects), and involves use of a known technique (simulating sound reflections) to improve a similar device (wearable communication device) in the same way (filtering sounds to produce a realistic and immersive sound field).

C. Dependent Claim 19

1. **Claim 19[a]:** “*The computer system of claim 15, wherein the processor further executes the instructions to change attenuation and reverberation of the binaural sound in order to add artificial audial cues of virtual objects near the second person,*”

199. It is my opinion that Bedingfield-Andrea in view of Sandgren discloses this limitation.

200. In my opinion, it would have been obvious to use Bedingfield’s *processor* to perform this limitation, in view of Sandgren.

201. As I explained above in Section XI.A-B, Sandgren relates to simulating sound-reflecting objects (*virtual objects*) to enhance the “[e]xternalization of the simulated sound source.” Ex. 1017, Abstract. As Sandgren makes clear, “[w]hen an object in a room produces sound, a sound wave expands outward from the source and impinges on walls, desks, chairs, and other objects that absorb and reflect different amounts of the sound energy.” *Id.*, ¶2, *see also id.*, ¶¶12, 16, 70, Fig. 1. Therefore, Sandgren discloses that these reflections and absorptions affect the sound wave.

202. In my opinion, Sandgren discloses *chang[ing] reverberation*. In Sandgren, “[r]eflected sound energy reaching the listener is generally called *reverberation*,” and “early *reverberation*” refers to early reflections that are from the simulated sound-reflecting object. Ex. 1017, ¶4, 47, *see also* ¶¶23-25, 48-66, 68-74,

100. “When simulating a dynamic 3D audio scene with moving objects and a moving listener, the parameters used by the [] early-reverberation generators 502, 502’ must be updated continuously in order to simulate the reflection paths accurately.”³ *Id.*, ¶69 (emphasis added), Figs. 5A-C, *see also id.*, 74 (describing updating early reflections when simulating a dynamic scene with moving objects). Furthermore, Sandgren explains that changes to reverberation provide *artificial audial cues*. EX1017, ¶¶47, 82-84. Sandgren further contemplates *reverberation* from objects near a listener (*second person*). *Id.* ¶4. By updating parameters for a reverberation generator to simulate virtual moving objects in a dynamic 3D audio scene, Sandgren discloses *change... reverberation of the... sound in order to add artificial audial cues of virtual objects near the person.*” Sandgren further contemplates the *sound* is a *binaural sound*. *See* Ex. 1017, ¶11 (describing binaural impulse responses for a reverberant room).

203. In my opinion, Sandgren also discloses *chang[ing] attenuation*. Specifically, Sandgren depicts in Figs. 5A-C audio simulators “having HRTF processors and an early-reflection generator,” and each of these figures additionally depicts attenuation filters 504, 506, 520, and 522. Ex. 1017, ¶¶31-33. According to

³ Sandgren refers to element 502 as an “early-reflection generator” as well as an “early-reverberation generator” and uses these names interchangeably. I therefore refer to Sandgren’s element 502 as an “early-reverberation generator.”

Sandgren, these attenuation filters represent a gain or amplification applied to the direct sound (*e.g.*, 504, 506) or early reflections (*e.g.*, 520, 522) for the left and right ears. *Id.*, ¶¶49, 62, Figs. 5A-C. In my opinion, the POSA would have readily appreciated that the attenuation applied to both direct sound and early reflections (which include $a_k(n)$ representing the attenuation filter for the k -th reflection) *change* as Sandgren's *virtual objects* move and change location. Ex. 1017, ¶¶74-75, 78-80, 89, Fig. 1. Indeed, as a virtual object moves and changes location, the direct sound and early reflections must change, and thus the attenuation applied to the relevant sounds would have also changed to reflect the new location. Furthermore, Sandgren explains that changes to attenuation provide *artificial audial cues*. EX1017, ¶¶47, 82-84. Sandgren further contemplates *attenuation* from objects near a listener (*second person*). *Id.* ¶4. Therefore, it is my opinion that Sandgren discloses *change... attenuation of the binaural sound in order to add artificial audial cues of virtual objects near the second person*.

2. **Claim 19[b]:** “*wherein the virtual objects do not actually exist near the second person but are perceived to exist by the first person from the artificial audial cues.*”

204. It is my opinion that Bedingfield in view of Sandgren discloses this limitation.

205. As I explained above for 19[a], the *virtual objects* in Sandgren are simulated sound-reflecting objects that are useful for enhancing the

“[e]xternalization of the simulated sound source.” Ex. 1017, Abstract. These *virtual objects* in Sandgren are “simulated” and therefore *do not actually exist near the person but* are simulated in Sandgren’s system, meaning they *are perceived to exist by the person from the artificial audial cues*. Ex. 1017, Abstract, ¶¶6, 23-25, 46-48, claims 1, 10, 19;

XII. GROUND 4: BEDINGFIELD IN VIEW OF BALLAS RENDERS OBVIOUS CLAIMS 8-10, 12, 14

A. Motivation to Combine

206. For the same reasons that I already explained above for the Bedingfield-Andrea and Ballas combination in Ground 2, it is my opinion that it would have been obvious to a POSA to track the user’s head orientation and adjust the sound so that the sound appears fixed even as the user’s head moves, as expressly taught in Ballas. *See* Section X.A.

B. Reasonable Expectation of Success

207. For the same reasons that I already explained above for the Bedingfield-Andrea and Ballas combination in Ground 2, it is my opinion that a POSA would have had a reasonable expectation of success in making the same modification to Bedingfield. *See* Section X.B (Ground 2, Reasonable Expectation of Success).

C. Independent Claim 8

1. **8[Pre]:** “*A non-transitory computer readable storage medium storing instructions that cause a computer system to execute a method, comprising:*”

208. To the extent limiting, it is my opinion that Bedingfield discloses the preamble.

209. As I already explained above for 15[pre], 15[c], and 15[d] in Ground 1, Bedingfield discloses *a computer system* that has a memory (*a non-transitory computer readable storage medium*) storing instructions that cause a processor within the *computer system to execute a method*. Section IX.E.1, 4-5.

2. **8[a]:** “*commence an electronic call between a first individual wearing an electronic earphone and a second individual that are remote from each other;*”

210. It is my opinion that Bedingfield discloses this limitation.

211. As I already explained above for 1[a] in Ground 1, Bedingfield discloses a telephone call (*electronic call* under 8KHZ’s position) that takes place (and is therefore *commence[d]* by Bedingfield’s *computer system*) between listener 102 (*first individual*, depicted in Bedingfield’s Fig. 1 as *wearing* headphones, or an *electronic earphone*) and participants 104, 106, and 108 (each of which is a *second individual* who is *remote from* the listener 102). Section IX.C.2.

3. **8[b]:** “*designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first*

individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;”

212. It is my opinion that Bedingfield discloses this limitation.

213. As I already explained above for 1[b] in Ground 1, Bedingfield describes its *computer system as designat[ing] ... an SLP ... in empty space ... that is located away from and proximate to the listener 102 (first individual)*. See Section IX.C.3. In Bedingfield, that *designat[ion]* is *for the first individual* because the designated sound is delivered to the listener 102 (*first individual*) through the headphones (*electronic earphone*). See Section IX.C.3. Moreover, in Bedingfield, the *SLP simulates a location ... from where an origin of a voice of any of participants 104, 106, 108 (second individual) occurs for the listener 102 (first individual)*. See Section IX.C.3 (Ground 1, 1[b]).

4. **8[c]:** “*execute the electronic call such that the origin of the voice of the second individual appears to the first individual to originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and*”

214. It is my opinion that Bedingfield alone or in view of Ballas discloses this limitation.

215. As I already explained above for 1[d] in Ground 1, Bedingfield’s computer system execute[s] the telephone call (electronic call) such that the origin

of the voice of the participant (second individual) appears to the listener (first individual) to originate from... the SLP in empty space. Section IX.C.5. As I already explained above for claim 18 in Ground 2, the combination of Bedingfield and Ballas teaches that the SLP is a fixed location... while the first individual moves and changes a location and a head orientation with respect to the SLP. Section X.A-C.

5. **8[d]:** *“maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first individual hears in response to the first individual moving the head orientation with respect to the sound localization point.”*

216. It is my opinion that Bedingfield alone or in view of Ballas discloses this limitation.

217. As I already explained above for 8[c] in Ground 4, Bedingfield and Ballas each discloses fixing (maintaining) the origin of the voice of the second individual at the SLP in response to the first individual moving the head orientation with respect to the SLP. See Section XII.C.4.

218. Furthermore, as I already explained above for 2[b] in Ground 1, claim 17 in Ground 1, and claim 18 in Ground 2, Bedingfield discloses *alter[ing] an ITD of the voice that the first individual hears*. See Sections IX.D.2, IX.F.1, X.A-C. As the POSA would have appreciated, and based on my discussion in the Background

(Section VI), it was commonplace to alter ITDs to maintain the origin of a sound while the user moves head orientation.

D. Dependent Claims 9, 10, 12, 14

1. **Claim 9:** “*The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: maintain the origin of the voice of the second individual at the sound localization point but alter an interaural level difference of the voice that the first individual hears in response to the first individual moving closer to and away from the sound localization point.*”

219. It is my opinion that Bedingfield alone or in view of Ballas discloses this limitation.

220. As I already explained above for 8[c] and 8[d] in Ground 4, Bedingfield alone or in view of Ballas discloses *maintaining the origin of the voice of the second individual at the SLP... in response to the first individual moving closer to and away from the sound localization point. See Sections XII.C.4-5 (Ground 4, 8[c]-8[d]).*

221. Furthermore, as explained for claim 17 in Ground 1 and claim 18 in Ground 2, Bedingfield discloses *alter[ing] an ILD of the voice that the first individual hears. See Sections IX.F.1, X.A-C.* As the POSA would have appreciated, and as I already referenced in the Background (Section VI) it was commonplace to alter ILDs to maintain the origin of a sound while the user moves nearer and further away from the SLP.

2. **Claim 10[a]:** “*The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: monitor movements of the head orientation of the first individual in response to the first individual localizing the voice of the second individual during the electronic call;*”

222. It is my opinion that Bedingfield alone or in view of Ballas discloses this limitation.

223. As I already explained for 2[a] in Ground 1, as Bedingfield’s listener’s head moves and changes orientation, for example to focus direction on different participants (*in response to the first individual localizing the voice of the second individual during the electronic call*), Bedingfield senses and detects those movements (*monitor movements of the head orientation of the first individual*). Section IX.D.1-2. Additionally, as I already explained above for claim 18 in Ground 2, Ballas discloses a head tracking system 104, which is used to *monitor movements of the head orientation of the first individual in response to the first individual localizing the voice of the second individual during the electronic call*. Section X.A-C.

3. **Claim 10[b]:** “*adjust head-related transfer functions (HRTFs) of the first individual to more accurately simulate real HRTFs of the first individual in response to the movements of the head*

orientation of the first individual localizing the voice of the second individual during the electronic call.”

224. It is my opinion that Bedingfield in view of Ballas discloses this limitation.

225. As I already explained above for claim 18 in Ground 2 and claim 10[a] in Ground XII, both Bedingfield and Ballas disclose changing (*adjust*) *HRTFs* to process audio signals and maintain a fixed location for a sound’s SLP *in response to the movements of the head orientation of the first individual localizing the voice of the second individual during the electronic call.* Section X.A-C, XII.D.2. Moreover, it is my opinion that a POSA would have understood that changing HRTFs in this context involves *adjust[ing] HRTFs of the first individual to more accurately simulate real HRTFs of the first individual*, because head tracking and adjusting HRTFs would have provided a more realistic virtual experience for the user, by allowing the user to move while fixing the perceived origin of sound sources.

- 4. Claim 12[a]:** *“The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: add a third individual to the electronic call;”*

226. It is my opinion that Bedingfield discloses this limitation.

227. As I already explained above for 1[a] in Ground 1, Bedingfield’s *electronic call* includes the listener 102 (*first individual*) and three additional

participants 104, 106, 108 (where any two participants are the *second and third individual*, for example where participant 104 is a *second individual* and participant 106 is a *third individual*). Section IX.C.2; *see* Ex. 1005, Figs. 4-9, 11, 14-15, 17-19, 4:19-23 (“any number of participants may be in connection with the listener.”). In my opinion, by disclosing an *electronic call* with at least three individuals on the call, Bedingfield discloses *add[ing] a third individual to the electronic call*.

5. **Claim 12[b]:** “*designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the third individual occurs for the first individual through the electronic earphone such that the sound localization point of the third individual is adjacent to the sound localization point of the second individual.*”

228. It is my opinion that Bedingfield discloses this limitation.

229. For the same reasons as I already explained for 1[b] in Ground 1, 6[b] in Ground 1, and 8[b] in Ground 4, Bedingfield teaches *designat[ing], for the first individual, an SLP that simulates a location 106a (location) in empty space from where an origin of a voice of the third individual occurs for the first individual through the electronic earphone*. Sections IX.C.3, IX.D.5, XII.C.3. Moreover, as Bedingfield depicts in Fig. 1, Bedingfield’s location 106a (*SLP of the third individual*) is adjacent to the location 104a (*SLP of the second individual*).

6. **Claim 14:** “*The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: receive, from the first individual, commands that move the sound*

localization point from a first point in empty space that is away from and proximate to the first individual to a second point in empty space that is away from and proximate to the first individual.”

230. It is my opinion that Bedingfield discloses this limitation.

231. Specifically, Bedingfield explicitly describes the listener (*first individual*) as “mak[ing] selections on the multi-channel processing device 118.”

Ex. 1005, 5:51-61, Fig. 3 (reproduced below).

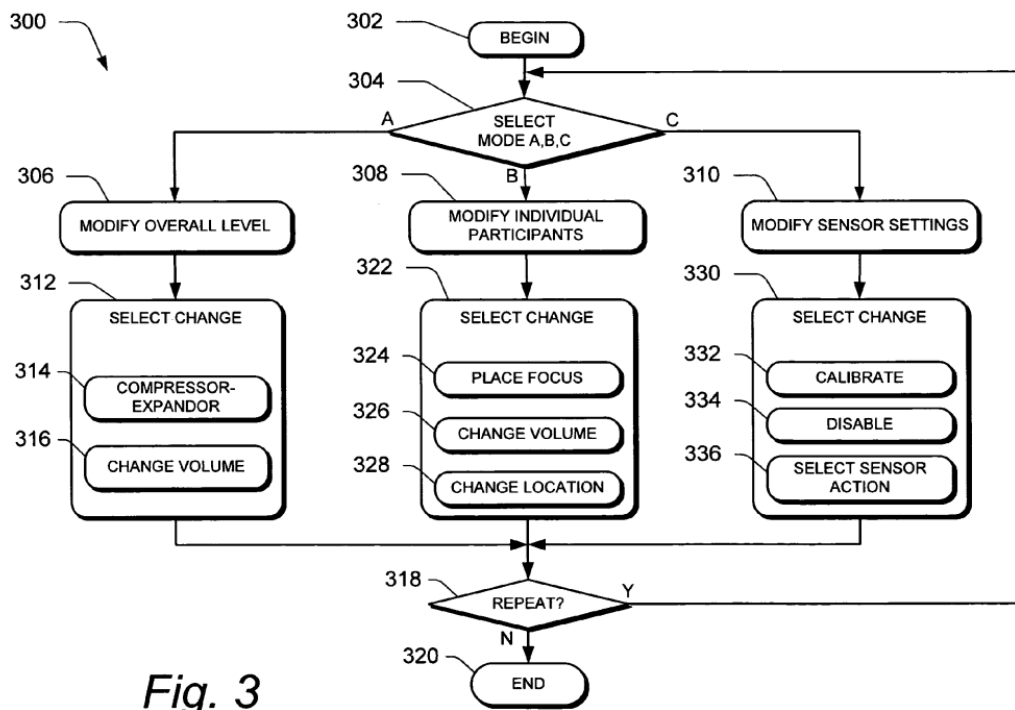


Fig. 3

232. According to Bedingfield, these selections from the listener include “at the block 328, the listener selects a participant and then selects a new location for the participant within the sound field.” Ex. 1005, 6:43-45, *see also id.*, 6:29-32, 9:60-64, 14:7-10, 16:33-35. As the POSA would have understood, by allowing the

listener (*first individual*) to select a participant of the conference call and select a new location for that participant at a distance from the listener, Bedingfield discloses *receive, from the first individual, commands that move the SLP from a first point in empty space that is away from and proximate to the first individual to a second point in empty space that is away from and proximate to the first individual* (under 8KHZ's position for *away from and proximate to*).

XIII. GROUND 5: LI IN VIEW OF ANDREA RENDERS OBVIOUS CLAIMS 1, 3, 6, 15-17

A. Motivation to Combine

233. In my opinion, it would have been obvious to a POSA use a wearable electronic device that *captures binaural sound* as disclosed in Andrea in Li's configurable 3D sound system. A POSA would have been motivated to implement this modification for the following reasons, as well as for the reasons explained in Ground 1.

234. The POSA would have readily recognized the advantages of incorporating Andrea's teachings of *capturing binaural voice* of a participant in an electronic phone call. For example, during such a call, a participant may be in a noisy location, and it would have been desirable to implement techniques and hardware for reducing that noise. For the same reasons as explained in Ground 1 (Section IX.A), the POSA would have recognized that incorporating Andrea's

capture of *binaural* sound into Li's 3D sound system would have reduced noise, improved sound quality, and enhanced the experience of those participating in the electronic call.

235. Moreover, the POSA would have readily understood that the modification would have been a natural extension of Li's disclosure of "[b]inaural recordings" that "sound closer to what humans hear in the real world." Ex. 1006, 2:22-45, *see also id.*, 19:56-59 (describing generating 3D binaural sound from a microphone array or multiple microphones). As Li also contemplates, two audio recordings (e.g., one at each ear) would have included localization cues, such as ITD and ILD. Ex. 1006, 1:65-2:3. By implementing Li to include Andrea's noise reduction techniques would have allowed Li's system to capture a participant's binaural voice in order to eliminate noise. Doing so would have improved the simulated conference by improving the sound quality and would have allowed Li's 3D sound system to create sound localization points with increased fidelity.

B. Reasonable Expectation of Success

236. It is my opinion that a POSA would have had a reasonable expectation of success in using Andrea's capture of a binaural voice in Li's 3D sound system.

237. Specifically, Li and Andrea are analogous art to each other and to the '090 patent. All three describe methods and systems for capturing a user's voice using at a wearable device and manipulating that sound to improve the

communication of the recorded voice. Specifically, Andrea is like Li in that it “relates to audio transmitting/receiving devices such as headsets with microphones.” Ex. 1008, 1:44-45. Further similar to Li, Andrea is also directed to applications including “communications such as voice over internet protocol (“VoIP”), PC to PC communications, PC to telephone communications.” Ex. 1008, 1:49-51. It is my opinion that, given the similarity between the two references, a POSA would have had a reasonable expectation of success modifying Li’s wearable device to *capture binaural voice* as taught in Andrea, as doing so would have improved sound quality and to produce more reliable localization cues.

238. Moreover, as Li describes, its 3D sound system is compatible with input from a multi-microphone array, including an array of two microphones, which the POSA would have understood to include Andrea’s microphones. *See* Ex. 1006, Fig. 9, 17:60-63. Thus, it would have been straightforward and beneficial to include a second microphone on the wearable device, and the modification to Li would have been a combination of known prior art elements according to known methods and known techniques to yield predictable results, and involves use of a known technique to improve a similar device in the same way.

C. Independent Claim 1

1. 1[pre]: “*A method, comprising:*”

239. To the extent this is limiting, it is my opinion that Li discloses the preamble.

240. In my opinion, Li discloses a *method* and configurable 3D sound system that generates a configurable three-dimensional sound. Ex. 1006, 3:39-44.

2. 1[a]: “*capturing, with an electronic earphone located at a head of a talking person, binaural sound that will be provided to a listening person during a telephone call*”

241. It is my opinion that Li in view of Andrea discloses this limitation.

242. Li describes a “configurable 3D sound system 900.” Ex. 1006, 16:58-17:14, 19:39-20:52, 21:60-22:5, 22:53-67, 23:21-43, 28:9-13, Figs. 9 (reproduced below), 16, 33.

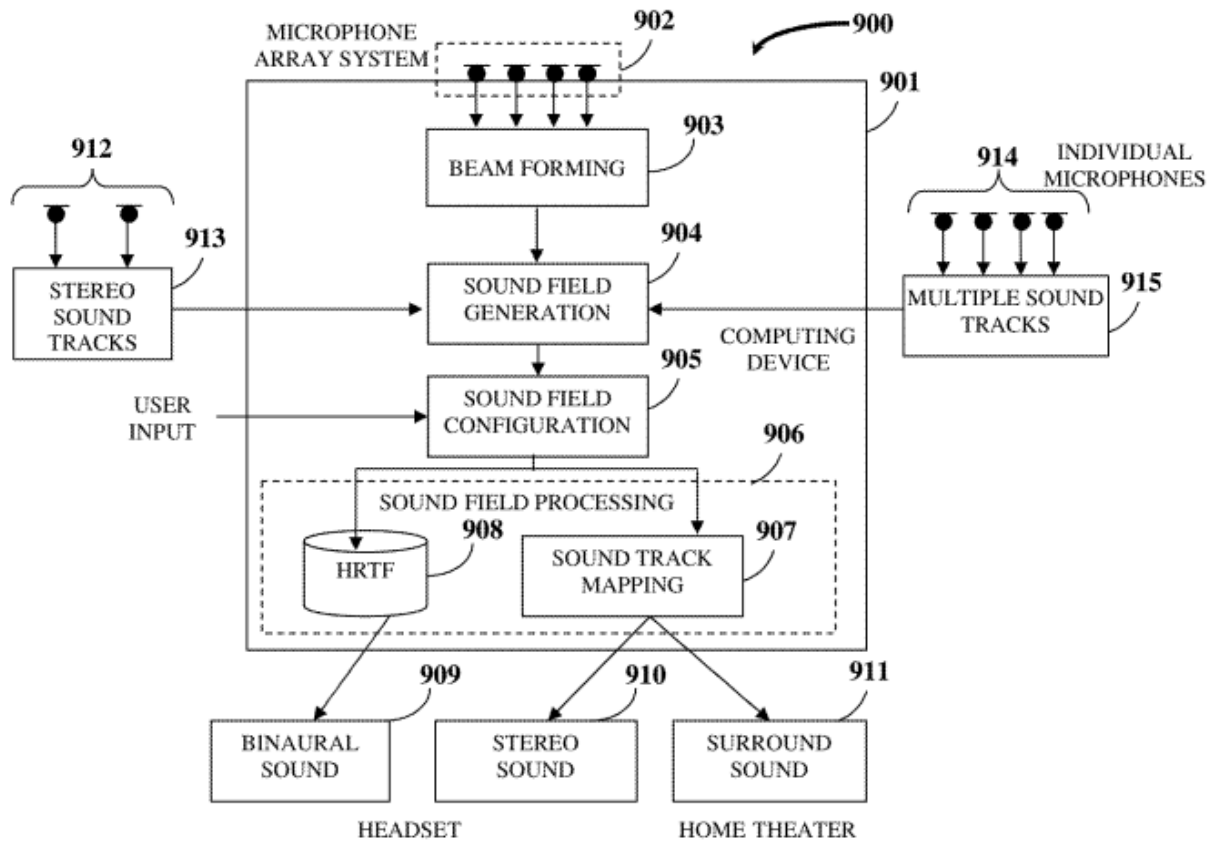


FIG. 9

243. Li describes Fig. 9 as a configurable 3D sound system 900 that provides a call amongst users with 3D sound. Specifically, Li's system 900 includes headphones or a headset, a microphone array system, and individual microphones, which together enable an electronic phone call (*a telephone call*) between users. Ex. 1006, 17:1-4, 17:22-37, 6:63-66; 7:41-8; 10:5-7, 22:1-5 (“[T]he configurable 3D sound system 900 provides a method of communication among multiple people located remotely from one another, for example, in a conference call, by placing different speaking users at different spots to mimic a real conference room

environment.”), According to Li, that electronic phone call (*telephone call*) is what allows a first user (*listening person*) and a second user (*a talking person*) to communicate using, for example, headphones and headsets. *See also id.*, 3:6-8 (“consumers often use headphones for audio rendering on a mobile device.”), 6:68-7:4, 7:33-37, 10:5-11, 14:14-15, 16:7, 17:35-37, 20:9-12, 20:44-48.

244. As was very common and routine in 2014, a headset would have included microphones, such that the POSA would have readily understood that Li’s headset in Fig. 9 would have included microphones, and that Li’s microphones in Fig. 9 would have been included on a headset. Thus, in my opinion, in Li’s system each of the two users (*listening person* and *talking person*) would have had their own headset or headphones.

245. Moreover, in Li’s system, because the headset (*electronic earphone*) has microphones, it is not only *located at the head of* the second (speaking) user (*talking person*), but it also *captures sound*, including the user’s voice, which *will be provided* to the first user (*listening person*). Ex. 1006, 4:58-63 (“‘Sound sources’ refers to... a person’s voice.”), 18:16-21 (“The 3D sound processing application records both audio and spatial information such that the recorded sound can be processed and reproduced to 3D sound.”). Li even contemplates *binaural* recording of a sound. Specifically, Li describes “placing two microphones in inner ear locations,” which the POSA would have understood to capture *binaural sound*. Ex.

1006, 2:22-32, *see also id.*, 4:6-7. Moreover, Li's system 900 enables simultaneous recording of binaural sound using a microphone array. Ex. 1006, 34:52-61; *see also id.*, 7:33-37 (describing individuals recording 3D sound), 2:22-32, 4:6-7, 34:52-61; Ex. 1032:176:3-5.

246. To the extent this is not explicit in Li, Andrea discloses a headset, which would have been used in Li. Ex. 1007, 14:55-58; *see also id.*, 11:66–12:3, 3:28-31, 1:44-51, 14:55-58.

247. I note that 8KHZ's expert in the parallel litigation has testified that it was known before 2014 to provide binaural sound through earphones. Ex. 1032:176:3-5.

248. To the extent Li does not expressly disclose *capture binaural sound*, it is my opinion that a POSA would have found it obvious to implement Li's *electronic earphone to capture[] binaural sound*, in view of Andrea, for the same reasons as explained in Ground 1. *See* Section IX.C.2 (Ground 1, 1[a]). A POSA implementing Li would have therefore been motivated to *capture binaural sound* taught by Andrea.

3. **1[b]:** “*designating, with a computer system, a sound localization point in empty space that is away from and proximate to the listening person such that the sound localization point simulates an origin of the binaural sound at the empty space that the listening person hears during the telephone call*”

249. It is my opinion that Li in view of Andrea discloses this limitation.

250. Li's *computer system* is a "configurable 3D sound system 900 for simultaneously generating configurable three-dimensional sounds" that is used for communication among multiple users, which include the *listening person* on a *telephone call with the talking person*. Ex. 1006, 16:58-17:14, 19:39-20:52, 21:60-22:5, 22:1-5, 22:53-67, 23:21-43, 28:9-13, Figs. 9, 16.

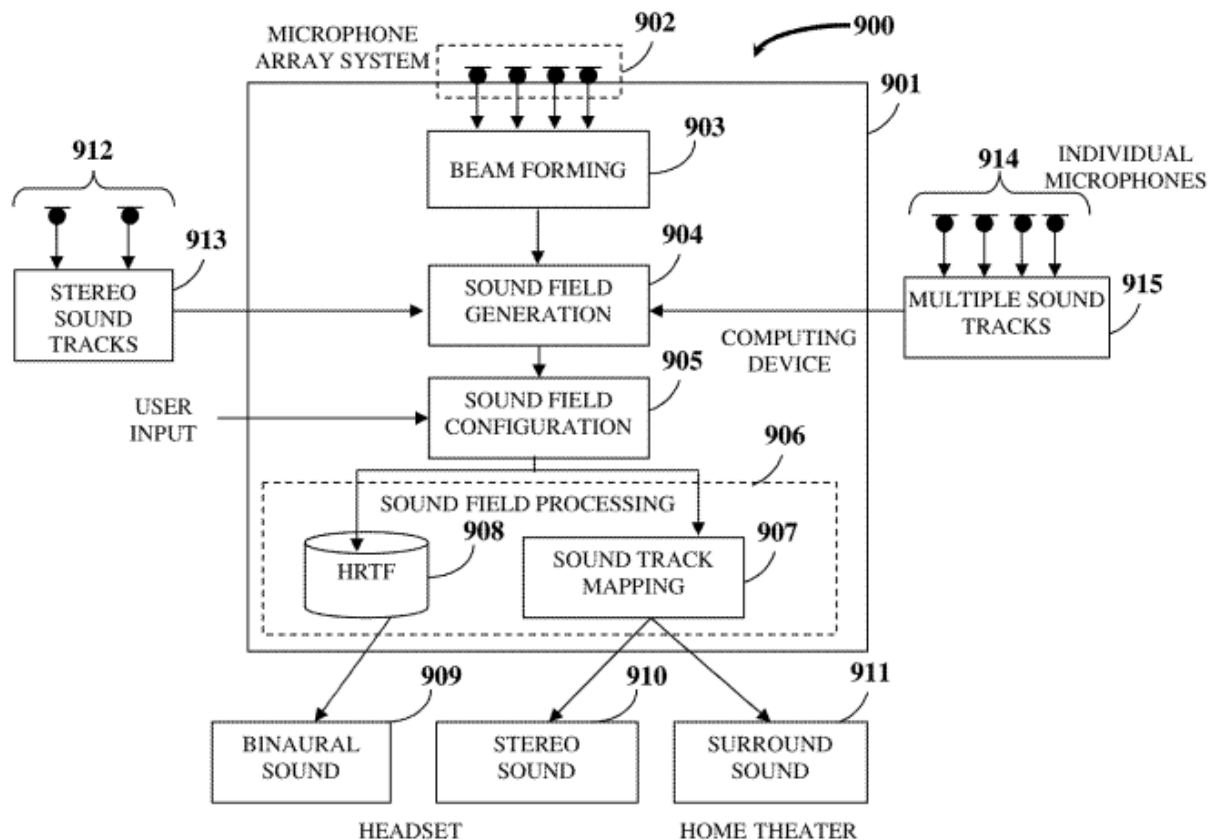


FIG. 9

251. In Li, the *computer system designat[es] an SLP* based on inputs received by a user of the 3D sound system 900. As Li depicts and explained in relation to Fig. 9 (which I have reproduced above), "the user configures the

placements and movements of any available sound sources as the inputs in order to obtain a virtual reality scene,” and the resulting 3D sound is provided to a headset. *Id.*, 22:53-67; *see also id.*, 21:57-66 (describing system 900 as “allowing the user to move the sound source around the user 1703, move the sound source up and down, etc., on the configurable sound field 1700.”), 20:37-40 (describing system 900 allowing the user the freedom to set the sound source locations), 22:39-42 (describing acquiring user selection of a location), *see also* Abstract, 2:42-45, 5:7-30, 15:40-42, 15:61–16:8, Fig. 9.

252. I read Li to disclose that the *SLP* is localized where no physical object exists that is at a distance from the user (*in empty space that is away from and proximate to the user*, under 8KHZ’s position). Ex. 1006, 20:50-53 and Fig. 17. Furthermore, I consider the term empty space to mean either space with no object at that location or a location that is unoccupied by an *SLP*. Under either understanding, my conclusion is the same – that it is obvious from the prior art. For example, Li describes “placing different speaking users at difference spots to mimic a real conference environment.” Ex. 1006, 22:1-5.

253. The POSA would have found it obvious to seat the listener (*listening person*) at a physical conference table, where an empty seat at the table (*empty space*) is allocated for a virtual participant, as was commonly done in the field. Such an implementation would have positioned the virtual participant’s *SLP* at the empty

seat at the physical table (*empty space*), with the advantage of ensuring that the virtual teleconference environment achieves Li's stated goal to "match a vivid 3D vision field." Ex. 1006, 3:11-16.

254. As I already explained for Ground 1, 1[b], teleconferences were commonplace, and it was known to designate empty chairs representing phone participants:



Ex. 1037, Fig. 2

255. Similarly, it was also known to have hybrid conference environments with participants (e.g., students) communicating in person and remotely, where those participating remotely select a seat not occupied by a physically present student (empty chair), and the phone participant's image is projected in the location of the selected set. Ex. 1038, 1, 5-7. During the call, voices of remote participants are produced such that they appear to originate from the projected image, which is located in the physically empty chair. *Id.*, 6-8.

256. In order to mimic an in-person conference, a POSA would have known to place users in locations where conference attendees would normally occupy, like a chair around a conference table. If that chair is already occupied by a person physically present or by another SLP, a conference participant would become confused as to who is speaking. Therefore, each SLP would be placed at a location devoid of people or other SLPs. In addition, some applications attempting to mimic a conference room environment may include a visual representation of a participant at the same location of that participant's SLP. Visual representations would preclude SLP locations already occupied by other objects, and thus, a POSA would place the SLP at an empty space.

257. Positioning the virtual participant's *SLP* at the empty seat at the physical table (*empty space*) would have had the advantage of ensuring that the virtual teleconference environment achieves Li's stated goal to "match a vivid 3D vision field." Ex. 1006, 3:11-16.

258. Furthermore, Li describes with respect to Fig. 1, measuring HRTFs "in a quiet sound treated room free of impulsive background noise." Ex. 1006, 10:39-54. As Li states, these measured HRTFs are used to generate binaural sound 909. *Id.*, 19:62-67, 17:27-37, 19:27-36. In my opinion, the POSA would understand that because HRTFs are measured in a "quiet sound treated room" to generate the binaural sound, the *SLP* would be placed in a similarly object-free space. Moreover,

because the sound sources are placed “at different spots to mimic a real conference room environment,” such as seats around a conference table (Ex. 1006, 21:66-22:5, 20:9-14), the *SLP at a first location is away from and proximate to the user*.

259. Moreover, Li’s first user (*listening person*) perceives (*hears*) the binaural sound to originate at the SLP *during the telephone call* because Li’s 3D sound system 900 is used to provide “a conference call” that “plac[es] different speaking users at different spots to mimic a real conference room environment.” Ex. 1006, 22:1-5.

260. Furthermore, to the extent that Li does not expressly disclose *binaural sound*, it is my opinion that the POSA would have found it obvious to implement Li’s system to adjust *binaural sound* as taught by Andrea. *See* Section XIII.C.2 (Ground 5, 1[a]).

4. 1[c]: “*adjusting, with the computer system, the binaural sound captured at the earphone of the talking person so the binaural sound originates during the telephone call from the sound localization point in empty space that is away from and proximate to the listening person; and*”

261. In my opinion, Li in view of Andrea renders obvious this limitation.

262. As I already explained above for 1[b] in this ground, Li’s configurable 3D sound system 900 (*computer system*) designates an SLP and processes sound using HRTFs (*adjusting the binaural sound*) to generate localized sound to the first user (*listening person*). *See* Section XIII.C.3 (Ground 5, 1[b]). Specifically, Li’s

configurable 3D sound system 900 includes a computing device 901 on which there is a “3D sound processing application” that reproduces 3D binaural sound 909 using HRTFs. Ex. 1006, 17:1-4, 17:22-37, 17:45-49, 18:11-13, 20:26-36, Fig. 16; *see also id.*, 5:22-28 (“The 3D sound processing application dynamically processes the recorded sound... to generate a configurable 3D binaural sound.”).

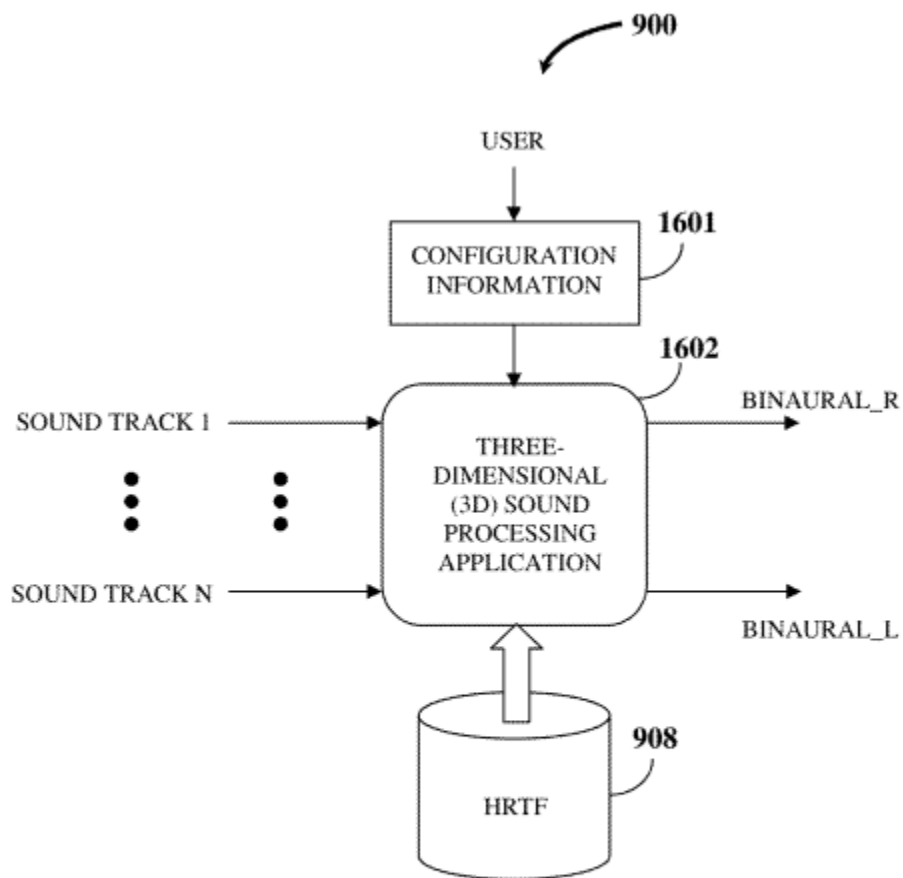


FIG. 16

263. As I already explained above for 1[b] in this ground, localized sound is at an *SLP in empty space that is away from and proximate to the listening person* and the *SLP* is a location from which *the binaural sound originates during the telephone call*. See Section XIII.C.3 (Ground 5, 1[b]). As I already explained above for 1[a] in this ground, Li in view of Andrea renders obvious that *the binaural sound is captured*. See Section XIII.C.2 (Ground 5, 1[a]).

5. 1[d]: “*providing, with an electronic earphone located at a head of the listening person, the binaural sound to the listening person during the telephone call such that the origin of the binaural sound for the listening person occurs at the sound localization point in empty space that is away from and proximate to the listening person.*”

264. In my opinion, Li renders obvious this limitation.

265. As I already explained above for 1[a] in this ground, Li contemplates that the first user (*listening person*) also has a headset (*an electronic earphone located at a head of the listening person*). See Section XIII.C.2 (Ground 5, 1[a]). Moreover, as I already explained above for 1[b] and 1[c] in this ground, Li discloses places the sound sources at “different spots to mimic a real conference room environment” over headsets, which discloses this limitation. See Sections XIII.C.3-4 (Ground 5, 1[b]-1[c]).

D. Dependent Claims 3, 6

1. **Claim 3:** “*The method of claim 1 further comprising: receiving, from the listening person, a location that specifies a distance*

from the listening person where to position the sound localization point in empty space that is proximate to and away from the listening person.”

266. It is my opinion that Li discloses this limitation.

267. As I already discussed above for 1[b] in this ground, Li’s first user (*listening person*) “configures the placements” of the sound sources. *See* Section XIII.C.3 (Ground 5, 1[b]). Thus, the POSA would have known that Li’s system *receiv[es], from the listening person, a location for where to position the SLP in empty space that is proximate to and away from the listening person.* Moreover, by configuring placement and allowing the user to move the sound source within the sound field 1700, Li’s *listening person* provides a location that specifies a distance *from the listening person* within the sound field. *See* Section XIII.C.3 (Ground 5, 1[b]).

2. Claim 6[a]: *“The method of claim 1 further comprising: capturing binaural sound from a second talking person;”*

268. It is my opinion that Li in view of Andrea discloses this limitation.

269. As I already explained above for 1[a] in this ground, each of Li’s conference call participants would have had a headset or headphones. *See* Section XIII.C.2 (Ground 5, 1[a]). As the POSA would have further appreciated, it would have been obvious to modify each headset or headphone by Andrea to *capture binaural sound* from each respective user. *See* Section XIII.C.1 (Ground 5, 1[pre]).

Moreover, Li discloses a conference call, which the POSA would have readily understood is capable of including more than two participants, such that a third user (*second talking person*) participates in the conference call.

3. **Claim 6[b]:** “*designating a second sound localization point in empty space that is away from and proximate to the listening person such that the second sound localization point simulates an origin of the binaural sound from the second talking person that the listener hears;*”

270. For the same reasons that I already explained for 1[b] and 6[a] in this ground, Li designates a placement (*a second SLP in empty space that is away from and proximate to the listening person*) that simulates an origin of the binaural sound from the third user (*second talking person*) that the listener hears. Sections XIII.C.3, XIII.D.2.

4. **Claim 6[c]:** “*positioning the second sound localization point adjacent to the sound localization point such that the binaural sound captured from the talking person appears to originate from the sound localization point and the binaural sound captured from the second talking person appears to originate from the second sound localization point.*”

271. It is my opinion that Li discloses this limitation.

272. As I already explained above for 1[b] in this ground, Li describes placing different speaking users at different spots. *See* Section XIII.C.3 (Ground 5, 1[b]). Moreover, as the POSA would have readily appreciated, Li’s goal to mimic

a real conference room environment would have meant that the different placements (*SLPs*) for the *talking persons* would have been positioned adjacent to each other.

E. Independent Claim 15:

1. 15[Pre]: “*A computer system, comprising:*”

273. To the extent this is limiting, Li discloses the preamble. *See* Section XIII.C.3-4 (Ground 5, 1[b]-1[c]).

2. 15[a]: “*a first wearable electronic device that a first person wears;*”

274. It is my opinion that Li discloses this limitation. As I already explained above for 1[a], Li discloses headsets and headphones, and each of two users would have had their own headset or headphones. *See* Section XIII.C.2 (Ground 5, 1[a]). Thus, in my opinion, Li discloses a headset (*first wearable electronic device*) that a first user (*first person*) wears.

3. 15[b]: “*a second wearable electronic device that a second person wears and that captures a binaural voice of the second person during an electronic phone call with the first wearable electronic device that the first person wears;*”

275. It is my opinion that Li in view of Andrea discloses this limitation.

276. As I already explained above for 15[a] in this ground, Li discloses a headset or headphones (*second wearable electronic device*) that a second user (*second person*) wears. *See* Section XIII.E.2 (Ground 5, 15[a]). For the same reasons that I already explained for 1[a] in this ground, the POSA would have found

it obvious to modify Li's headset or headphones to incorporate Andrea's *capture of a binaural voice of the second person during an electronic phone call with the first wearable electronic device that the first person wears*. See Section XIII.C.2 (Ground 5, 1[a]).

4. 15[c]: “a memory that stores instructions; and”

277. It is my opinion that Li discloses this limitation. As I've explained, Li's configurable 3D sound system 900 is a *computer system* (see 15[pre]), which “employs the architecture of the computer system 3400” including “a *memory* unit 3402 for *storing* programs and data” as well as “*instructions* [that are] fetched by the processor 3401.” Ex. 1006, 30:4-35, 31:18-35, Fig. 34, *see also* 30:65-31:1, 35:24-50. Li's *computer system* therefore includes a memory unit 3402 (*memory*) that *stores instructions*.

5. 15[d]: “a processor that executes the instructions to...”⁴

278. It is my opinion that Li discloses this limitation. Li's configurable 3D sound system 900 is a *computer system* (see 15[pre]), which “employs the architecture of the computer system 3400” including a *processor* 3401 that “*executes* computer programs” stored in the memory unit 3402. (Ex. 1006, 30:17-19, 30:26-

⁴ I note that 8KHZ's expert in the parallel litigation has admitted that the POSA would have known how to implement 15[d], 15[d1], and 15[d2]. Ex. 1032, 148:22-149:14.

29 (“The memory unit 3402... stores... *instructions* for execution by the processor 3401.”), *see also* 31:18-35, 32:32-33:31, 34:36-51, 35:17-50, 36:37-41). Therefore, in my opinion, Li’s *computer system* includes a processor 3401 (*processor*) that *executes the instructions*.

6. **15[d1]:** “*adjust the binaural voice of the second person during the electronic phone call in order to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person,*”

279. It is my opinion that Li and Andrea render obvious this limitation.

280. As I already explained above for 1[c] in this ground, Li processes sound by applying HRTFs to (*adjust*) the sound from the speaking user (*the binaural sound of the second person*) during the electronic phone call in order to position the localized sound at an *SLP that is away from and proximate to the listening user (first person)*. *See* Section XIII.C.4 (Ground 5, 1[c]). As I already explained for 15[b] in this ground, that sound is a *binaural voice*, as taught by Andrea. *See* Section XIII.E.3 (Ground 5, 15[b]).

7. **15[d2]:** “*wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.*”

281. It is my opinion that Li in view of Andrea discloses this limitation. As I already explained for 1[a], 1[d], 15[b], and 15[d1] in this ground, Li places speaking users at “different spots,” each of which is an *SLP that represents an origin*

of the binaural sound of the speaking user (second person) to the listening user (first person), and that binaural sound is the binaural voice from the speaking user.

Sections XIII.C.2, XIII.C.5, XIII.E.3, XIII.E.6.

F. Dependent Claims 16-17

1. **Claim 16:** “*The computer system of claim 15, wherein the processor further executes the instructions to determine a location of an empty chair that is proximate to the first person and position the sound localization point at the empty chair.*”

282. In my opinion, Li renders obvious this limitation.

283. As I already explained above for 1[b], 1[c], 15[d]-15[d1] in this ground, Li-Andrea discloses *a processor that executes instructions... to position the SLP at a location that is proximate to the first person*, and Li’s placement of sound source locations are *SLPs*. See Sections XIII.C.3-6 (Ground 5, 1[b]-1[c], 15[d]-15[d1]). Moreover, as I already explained above for 1[b] in this ground, the POSA would have known to position Li’s *SLP* at an *empty chair* at a conference table, where an *empty seat or chair* is allocated for the virtual participant. See Section XIII.C.3 (Ground 5, 1[b]); *see also* Ex. 1006, 22:1-5; *see also* teachings in Ex. 1037 and 1038 for providing spatialized audio in hybrid in-person and remote environments, including where remote participants are assigned an empty chair in the space.

284. In my opinion, the POSA would have readily known that conference participants are typically seated in chairs around conference tables. Thus, the POSA

would have known that to simulate a real conference environment as realistically as possible, it would have been obvious to position the *SLP* for a remote participant at an *empty chair*, and therefore *determine a location of an empty chair*, and *position the SLP at the empty chair*.

285. Moreover, Li explains that a user may enter a user input to select a location of for a participant *SLP*. EX1006, 15:44-45, 27:16-22. When Bedingfield's computer system receives a user input indicating a request to place the *SLP* at an *empty chair*, which a *POSA* would recognize as a common request in a teleconference setting, Li's computer system then *determine[s]* the *location of the empty chair* as the target for the *SLP* and then *position[s]* the *SLP* at the *determine[d]* *empty chair location*. EX1006, 11:58-62, 20:32-36. Therefore, in my opinion, a *POSA* would have known that a user selection would result in Bedingfield's computer system *execute[ing] instructions to determine a location of an empty chair*.

2. **Claim 17:** “*The computer system of claim 15, wherein the processor further executes the instructions to change an interaural time difference (ITD), an interaural level difference (ILD), and a head-related transfer function (HRTF) in order to position the sound localization point of the binaural voice to the location that is away from and proximate to the first person.*”

286. It is my opinion that Li renders obvious this limitation.

287. As I've already explained above in 15[d], 15[d1], 15[d2] for this ground, Li's processor executes instructions that include applying HRTFs to position

an SLP of the binaural voice to a location that is away from and proximate to the first person. Section XIII.E.5-7.

288. Li further explains that “human hearing localizes sound sources in a three-dimensional (3D) spatial field, mainly by three cues: *an interaural time difference* (ITD) cue, *an interaural level difference* (ILD) cue, and a spectral cue.” Ex. 1006, 1:65-2:3. Li also gives the well-known definitions for ITD and ILD: defining *ITD* as “the difference of arrival times of transmitted sound between the two ears” and *ILD* as “the difference in level and/or intensity of transmitted sound received between the two ears.” *Id.*, 1:65-2:4, *see also id.*, 12:54-51-57.

289. In my opinion, Li also discloses *changing* each of the *ILD* and *ITD* as part of its bank of HRTF filters. Specifically, according to Li, the “combination of these factors [including ILD and ITD] is modeled by two sets of filters... [t]he transfer functions of [which] are called head related transfer functions. Since different effects are caused by different locations of the sound source, the HRTFs are a bank by positions.” Ex. 1006, 2:6-21, *see also id.*, 12:57-65. In my opinion, the POSA would have understood that Li applies one HRTF (from the bank of HRTFs) to produce a spatialized 3D sound for a specific position, and applies a different HRTF (*change the HRTF*) to position the 3D sound at a new position (*SLP*). *See id.*, *see also id.*, 5:26-30, 10:55-11:9, 12:37-13:7, 15:53-16:4, 20:37-52, 22:44-52. As each HRTF models a different combination of ILD and ITD cues, Li’s

selection of different HRTFs for different 3D sound positioning discloses *changing ILD and ITD*.

290. Thus, based on Li's teachings and the general knowledge in the art, it is my opinion that a POSA would have known that the *ILD*, *ITD*, and *HRTF* are all sound processing techniques that would be *changed* in Li in order to *position the SLP* at the desired location.

XIV. GROUND 6: LI IN VIEW OF ANDREA AND BALLAS RENDERS OBVIOUS CLAIMS 2, 18

A. Motivation to Combine Li and Andrea with Ballas

291. It is my opinion that it would have been obvious to a POSA to modify Li-Andrea to track the user's head orientation and adjust the sound so that the sound appears fixed even as the user's head moves, as taught in Ballas. Li already recognizes that to determine a sound source, humans rely on various cues including head motion. Ex.1005, 1:38-42. Based on this teaching, the POSA would have been motivated to look to head tracking references, including Ballas, to monitor the head motion. A POSA would have been motivated to combine the teachings in Li with those in Ballas because a POSA would have recognized that both references have a shared goal of providing a realistic experience for the user. To achieve this goal, the POSA would have known to track the user's head orientation and movements and provide that information to process Li's sound field. That processing would have

accounted for the user's orientation and movements, using conventional head tracking and spatialization techniques taught by Ballas.

292. As Li explains, its primary goal is to provide a sound field that closely resembles “what humans hear in the real world” and “mimic a real conference room environment.” Ex. 1006, 2:34-36, 22:1-5; *see also id.* 2:22-24, 13:57-60. Ballas shares this stated goal of providing “a realistic experience” by providing “spatialized sound” for applications such as “teleconferencing and telepresence” and “virtual and augmented reality environments.” Ex. 1008, ¶¶2, 9, 10. Li's method is for potentially lengthy stretches of time, such as for conferences, concerts, or radio dramas. Ex. 1006, 22:1-5. For such applications, head tracking is key to make the experience as realistic as possible for the user. As already explained for Ground 2, above, head tracking allows the user to freely change the orientation of the user's head and adjust the sound field to account for the user's head movements so that the origins of the sound sources appear fixed and therefore realistic to the user. Modifying Li-Andrea with Ballas's teachings of head tracking and adjusting the sound field would have been beneficial as it would have allowed the user to move while staying connected to the event, and maintaining the sound field to appear fixed as the user moves and changes head orientation.

293. Moreover, it is my opinion that the modification would have been a natural extension of Li's use of HRTFs to spatialize sound and generate a sound

field. Ballas is similar to Li in that it also discloses an “[HRTF] filter to produce spatialized audio.” Ex. 1008, Abstract; *see also id.*, ¶¶16-17, 25. Ballas further discloses a “head tracking system determines the location and orientation of the listener’s head and provides the measurements to a computer which processes audio signals” using an “[HRTF] filter for maintaining virtual location of the audio signals thereby allowing the user to change location and head orientation without degradation of the virtual location of audio signals.” *Id.*, Abstract, ¶18, *see also id.*, ¶¶17, 19-20, 25-26, claims 1, 7. By using head tracking to adjust HRTFs would have allowed the user to change head location and orientation while maintaining fixed virtual sound locations. This implementation would have been obvious to the POSA as it would have advanced the common goals of both Li and Ballas to provide a realistic environment in a teleconferencing application.

B. Reasonable Expectation of Success

294. In my opinion, a POSA would have had a reasonable expectation of success in modifying Li-Andrea with the teachings of Ballas. Both Li and Ballas use HRTFs to generate spatialized binaural sound and provide such sound to a user over headphones on wearable devices. In Li-Andrea, the wearable device includes headphones or headset worn on the user’s head. Ex. 1006, Fig. 9.

295. Ballas’s system is consistent and compatible with the Li-Andrea system, as it includes speakers 110 that are attached to a “frame [that] is mounted on

the user's head and moves as the head moves.” Ex. 1008, ¶25. The sound that is provided over the speakers to the user is “binaural in nature” and is spatialized audio because it is processed using HRTF filters 106. *Id.* Accordingly, it is my opinion that a POSA would have recognized that Li-Andrea and Ballas are structurally and functionally similar.

296. Moreover, the POSA implementing the Li-Andrea system would have recognized that the Li-Andrea headset would be attached to Ballas's frame. Also mounted on Ballas's frame is a head tracking system 104 that “measures the location and orientation of a user's head and provides the measured information to the computer system 102 which processes the audio signals using a head related transfer function (HRTF) filter 106 thus producing spatialized audio,” which is “fed to speakers 110.” Ex. 1008, ¶25. This implementation would have merely involved a combination of known prior art elements (headset and head tracking system) according to known methods and techniques (accounting for head movement) to yield predictable results (producing spatialized sound that appears fixed as the user moves), and involves use of a known technique (HRTF) to improve a similar device (wearable communication device) in the same way (filtering sounds to produce a sound field that appears fixed).

C. Dependent Claim 2

1. **Claim 2[a]:** “*The method of claim 1 further comprising: tracking a head orientation of the listening person with respect to the sound localization point;*”

297. It is my opinion that Li in view of Ballas discloses this limitation.

298. For the same reasons that I explained above for claim 18 in Ground 2 (Section X.C.1), it is my opinion that a POSA would have been motivated to implement the Li-Andrea headphones as part of Ballas’s wearable frame having a head tracking system. Doing so would have provided a more realistic virtual experience for the user, by allowing the user to move while fixing the perceived origin of the sound sources. *See also* Section XIV.A (Motivation to Combine Li-Andrea with Ballas).

2. **Claim 2[b]:** “*adjusting an interaural time difference of the sound that the listening person hears in response to the listening person changing the head orientation with respect to the sound localization point.*”

299. It is my opinion that Li in view of Ballas discloses this limitation.

300. As I explained above for claim 18 in Ground 2 (Section X.C.1), Ballas discloses *adjusting... the sound as the listening person chang[es] head orientation with respect to the SLP*. As I explained above for claim 17 in Ground 5 (Section XIII.F), Li further explains that *ITD* is *adjust[ed]* in order to position a 3D sound,

which the POSA would have understood to be performed in order to fix the sound's *SLP* while the user changes *head orientation*.

D. Dependent Claim 18

1. **Claim 18:** *“The computer system of claim 15, wherein the processor further executes the instructions to change an interaural time difference, an interaural level difference, and a head-related transfer function of the binaural sound in order to maintain the sound localization point fixed at the location during the electronic phone call while the first person physically moves away from the location and changes a head orientation with respect to the location.”*

301. It is my opinion that Li in view of Ballas discloses this limitation.

302. As I explained above for 15[d] in Ground 5 and claim 17 in Ground 5, Li and Andrea render obvious a processor that executes instructions to change an ITD, an ILD, and an HRTF to position the SLP of the binaural voice to a location that is away from and proximate to the first person. Moreover, as I explained above for claim 17 in Ground 5, Li renders obvious chang[ing] an ITD, an ILD, and an HRTF of the binaural sound. Section XIII.E.5-7, XIII.F.1.

303. For the same reasons that I explained above for claim 18 in Ground 2, Ballas teaches maintaining the SLP fixed at the location during the electronic phone call while the first person physically moves away from the location and changes a head orientation with respect to the location. Section X.C.1, XIV.C-D. Moreover, I note that 8KHZ's expert in the parallel litigation admitted that it was known in the

1980s for “head-track virtual acoustic systems [to] keep a sound source fixed in a location in virtual space independent of head movement”). Ex. 1032, 250:2-9; see also Ex. 1036, 12, 99, 132-34, Figs. 4.37-4.38.

XV. GROUND 7: LI IN VIEW OF ANDREA AND SANDGREN RENDERS OBVIOUS CLAIM 19

A. Motivation to Combine

304. In my opinion, it would have been obvious to a POSA to modify Li-Andrea to adjust attenuation and reverberation of the sound to add artificial cues of virtual objects that do not actually exist but are perceived to exist, as taught in Sandgren.

305. Each of the references recognizes the importance of sound cues in creating a virtual environment. Li already recognizes that humans rely on various types of cues (including reverberation) to determine a source of a sound. Ex. 1006, 1:38-42 (identifying eight sources of localization cues including reverberation).

306. A POSA would have been motivated to combine these teachings with Sandgren because a POSA would have recognized that the references have a shared goal of providing a realistic and immersive experience for the user. To achieve this goal, and as I explained above for Ground 3, Sandgren states “it is not enough to concentrate on the direct sound,” as other cues, including reverberation and

attenuation, are important to fully simulate the 3D audio. Ex. 1017, ¶¶6, 12; Section XI.A.

307. In my opinion, the POSA would have understood to use conventional sound enhancement techniques taught by Sandgren to simulate sound-reflecting objects that do not actually exist but are perceived to exist in the virtual environment. The primary goal of Li is to provide a sound field that closely resembles “what humans hear in the real world” and “mimic a real conference room environment.” Ex. 1006, 2:34-36, 22:1-5; *see also id.* 2:22-24, 13:57-60. Li recognizes that “[s]ounds are a constant presence in everyday life and offer rich cues about the environment.” *Id.*, 1:24-25. “There is a need for accurately synthesizing such spatial sound to add to the immersiveness of a virtual environment.” *Id.*, 1:46-47. Sandgren shares this goal of “electronic creation of virtual three-dimensional (3D) audio scenes” and “increasing the externalization of virtual sound sources presented through earphones.” Ex. 1017, ¶1, *see also id.*, Abstract.

308. As I explained for Ground 3, Sandgren appreciates that “[a] room’s reverberation depends on... the types of objects present in the room,” and in order to “creat[e] a realistic 3D audio scene... it is not enough to concentrate on the direct sound. Simulating reverberation is also important as the reverberation changes the loudness, timbre, and the spatial characteristics of sounds and can give a listener different kinds of information about a room, e.g., the room’s size and whether it has

hard or soft reflective surfaces.” *Id.*, ¶¶5-6. According to Sandgren, “it is not the main reflections from the floor, ceiling, and walls of a room that result in externalization. Instead, other objects in the room, e.g., tables and chairs, that scatter sound waves are essential for good externalization.” *Id.*, ¶19. As Sandgren further contemplates, a realistic sound simulator includes attenuation filters combined with reverberation generators. *Id.*, ¶¶62-63, Fig. 5C. Modifying Li-Andrea in accordance with Sandgren’s teachings of adjusting attenuation and reverberation to simulate virtual objects would have provided a more realistic and immersive experience for the user.

309. Moreover, in my opinion, the modification would have been a natural extension of the use of HRTFs in Li-Andrea to spatialize sound and generate a sound field. Sandgren is like Li in that it also discloses using HRTFs to “simulat[e] an audio scene or creat[e] a virtual audio scene” over headphones. Ex. 1017, ¶¶9-10; *see also* ¶¶11, 15-16, 19, 22, 46, 48-64, 70-100, Figs. 5A-C. Sandgren further describes “simulated sound-reflecting objects are simulated by processing a direct-sound signal with at least one [HRTF]... and generating simulated early-reflection signals.” *Id.*, Abstract. In my opinion, the POSA would have recognized that the teachings of Sandgren would have enhanced the use of HRTFs in Li-Andrea for an immersive experience by introducing attenuation filters and reverberation generators indicative of virtual objects in the environment, which would have advanced the

common goals of the references to provide a realistic and immersive experience for the user.

B. Reasonable Expectation of Success

310. In my opinion, a POSA would have had a reasonable expectation of success in modifying Li-Andrea in accordance with the teachings of Sandgren. Li-Andrea and Sandgren are similar in that both use HRTFs to generate spatialized binaural sound and provide such sound over headphones. In addition, Sandgren is consistent with the Li-Andrea system, in that it creates virtual 3D audio scenes over headphones so that the virtual sound sources are externalized. Ex. 1017, ¶¶1, 46. The 3D sounds are provided to the user, who perceives the sounds to originate from different SLPs. *Id.* Accordingly, it is my opinion that a POSA would have recognized that Li-Andrea and Sandgren are structurally and functionally similar.

311. Moreover, as the POSA implementing the Li-Andrea system would have recognized, it would have been reasonably simple to implement the HRTF processing that spatializes the sound with Sandgren's attenuation filters and reverberation generators, in order to artificially simulate the presence of virtual objects. Such an implementation would have merely involved a combination of known prior art elements (HRTF and attenuation filters and reverberation generators) according to known methods and techniques (simulate sound-reflecting virtual objects) to yield predictable results (producing sound so that the user

perceives the virtual objects), and involves use of a known technique (simulating sound reflections) to improve a similar device (wearable communication device) in the same way (filtering sounds to produce a realistic and immersive sound field).

C. Dependent Claim 19

1. **Claim 19[a]:** *“The computer system of claim 15, wherein the processor further executes the instructions to change attenuation and reverberation of the binaural sound in order to add artificial audial cues of virtual objects near the second person,”*

312. It is my opinion that Li in view of Sandgren discloses this limitation.

313. Li already discloses that *reverberation* is a “source[] of localization cues.” Ex. 1006, 1:39-42. It would have been obvious to implement Li’s *processor* to perform this limitation, in view of Sandgren, for the same reasons discussed for Ground 3. *See* Section XI.C.1 (Ground 3, 19[a]).

2. **Claim 19[b]:** *“wherein the virtual objects do not actually exist near the second person but are perceived to exist by the first person from the artificial audial cues.”*

314. It is my opinion that Li in view of Sandgren discloses this limitation, for the same reasons discussed for Ground 3. *See* Section XI.C.2 (Ground 3, 19[b]).

XVI. GROUND 8: LI IN VIEW OF BALLAS RENDERS OBVIOUS CLAIMS 8-10, 12, 14

A. Motivation to Combine

315. For the same reasons that I already explained for the Li-Andrea combination with Ballas, it would have been obvious to a POSA to modify the Li

system alone to track the user's head orientation and adjust the sound so that the sound appears fixed even as the user's head moves, as taught in Ballas. *See* Section XIV.A (Ground 6, Motivation to Combine).

B. Reasonable Expectation of Success

316. In my opinion, a POSA would have had a reasonable expectation of success in making the Ballas modification to the Li system alone, for the same reasons as I already explained for the Ballas modification to the Li-Andrea system. *See* Section XIV.B (Ground 6, Reasonable Expectation of Success).

C. Independent Claim 8

1. **8[pre]:** *“A non-transitory computer readable storage medium storing instructions that cause a computer system to execute a method, comprising:”*

317. To the extent this is limiting, it is my opinion that Li discloses the preamble.

318. As I already explained above in Ground 5 for 15[pre], 15[c], and 15[d], Li discloses *a computer system* that has a memory (*a non-transitory computer readable storage medium*) storing instructions that cause a processor within the *computer system to execute a method*. *See* Section XIII.E.1, 4-5.

2. **8[a]:** *“commence an electronic call between a first individual wearing an electronic earphone and a second individual that are remote from each other;”*

319. It is my opinion that Li discloses this limitation.

320. As I already explained for 1[a] in Ground 5, Li discloses an electronic phone call (*electronic call*) that takes place (and is therefore *commence[d]* by Li's *computer system*) between a first user (*first individual wearing an electronic earphone*) and a second user (*second individual who is remote from the first user*). See Section XIII.C.2.

3. **8[b]:** “*designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;*”

321. It is my opinion that Li discloses this limitation.

322. As I already explained above in Ground 5 for 1[b], Li describes its computer system as *designat[ing]. . . an SLP . . . in empty space . . . that is located away from and proximate to the first user (first individual)*. See Section XIII.C.3.

323. Li teaches that *designat[ion]* is *for the first individual* because the designated sound is delivered to the first user (*first individual*) through the headphones (*electronic earphone*). See Section XIII.C.3.

324. Li further teaches that the *SLP simulates a location . . . from where an origin of a voice of the second user (second individual) occurs for the first user (first individual)*. See Section XIII.C.3.

4. **8[c]:** “*execute the electronic call such that the origin of the voice of the second individual appears to the first individual to*

originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and”

325. It is my opinion that Li in view of Ballas discloses this limitation.

326. As I explained above in Ground 5 for 1[d], Li’s *computer system execute[s] the electronic phone call (electronic call) such that the origin of the voice of the second user (second individual) appears to the first user (first individual) to originate from . . . the SLP in empty space. See Section XIII.C.5.*

327. As I explained for claim 18 in Ground 6, the combination of Li and Ballas teaches that the *SLP is a fixed location . . . while the first individual moves and changes a location and a head orientation with respect to the SLP. See Sections XIV.A-B, D.*

5. **8[d]:** *“maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first individual hears in response to the first individual moving the head orientation with respect to the sound localization point.”*

328. It is my opinion that Li in view of Ballas discloses this limitation.

329. As I already explained for 8[c] in this ground, the combination of Li and Ballas discloses fixing (maintaining) *the origin of the voice of the second individual at the SLP in response to the first individual moving the head orientation with respect to the SLP. See Section XVI.C.4.*

330. Furthermore, as I already explained for 2[b] in Ground 6 and claims 17 in Ground 5 and claim 18 in Ground 6, Li discloses *alter[ing] an ITD of the voice that the first individual hears*. See Sections XIII.F.1, XIV.C-D.

331. A POSA would have appreciated that it was commonplace to alter ITDs to maintain the origin of a sound while the user moves head orientation. As I explained in the Technology Overview, when a sound source is positioned directly in front of the human listener, the sound generally arrives at each ear of the human listener at the same time. See Section VI.A. But when the sound is shifted to the left or the right, the sound will arrive at each ear at different times. Li explains that “[i]f the sound sources shifts to the left, the left ear receives the sound earlier . . . [t]his helps humans determine where the sound is being emitted.” Ex. 1006, 2:9-13.

332. Li further teaches that ITD is one of the factors that is modeled in HRTFs. *Id.*, 2:14-17 (“The combination of these factors is modeled by two sets of filters on the left ear and the right ear separately in order to describe the spatial effect which is recognizable by human hearing.”). “The transfer functions of such filters are called head related transfer functions (HRTFs).” *Id.*, 2:7-19. Thus, a POSA modifying the system of Li to include the head tracking of Ballas would have understood that the alteration of the ITD would be done in response to the sound source shifting in relation to the head orientation.

D. Dependent Claims 9, 10, 12, 14

1. **Claim 9:** “*The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: maintain the origin of the voice of the second individual at the sound localization point but alter an interaural level difference of the voice that the first individual hears in response to the first individual moving closer to and away from the sound localization point.*”

333. It is my opinion that Li in view of Ballas discloses this limitation.

334. As I explained above for 8[c] and 8[d] in this ground, Li in view of Ballas discloses *maintaining the origin of the voice of the second individual at the SLP... in response to the first individual moving closer to and away from the sound localization point. See Sections XVI.C.4-5.*

335. Furthermore, as I explained above for claims 17 in Ground 5 and claim 18 in Ground 6, Li discloses *alter[ing] an ILD of the voice that the first individual hears. See Sections XIII.F.1, XIV.A-B, D.*

336. A POSA would have appreciated that it was commonplace to alter ILDs to maintain the origin of a sound while the user moves nearer and further away from the SLP. As I explained in the Technology Overview, when a sound source is positioned closer to one ear than the other, the sound level between the two ears also differs. *See Section VI.A.* When a user moves nearer or further away from the SLP there is a difference in the sound level. This difference in sound level arises between

the two ears because “the listener’s head casts an acoustical shadow giving rise to a lower sound level at the ear farthest from the sound source.” Ex. 1035, 944. Thus, a POSA would have understood that the alteration in the ILD is done in response to the user moving closer to or away from the SLP.

2. Claim 10

a. 10[a]: “*The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: monitor movements of the head orientation of the first individual in response to the first individual localizing the voice of the second individual during the electronic call;*”

337. It is my opinion that Li in view of Ballas discloses this limitation.

338. As I discussed above in Ground 6 for 2[a], Ballas discloses a tracking system for tracking the user’s head orientation and location (*monitor movements of the head orientation of the first individual*). See Section XIV.C.1, Ex. 1008, ¶¶25-26.

339. Additionally, as I explained above in Ground 5 for 1[b], Li’s first user (*listening person*) “configures the placements” of the sound sources (*localizing the voice of the second individual during the electronic call*). See Section XIII.C.3.

b. 10[b]: “*adjust head-related transfer functions (HRTFs) of the first individual to more accurately simulate real HRTFs of the first individual in response to the movements*

of the head orientation of the first individual localizing the voice of the second individual during the electronic call.”

340. It is my opinion that Li in view of Ballas discloses this limitation.

341. As I already explained above for claim 18 in Ground 6 and 10[a] in this ground, Li in view of Ballas discloses changing (*adjust*) *HRTFs* to process audio signals and maintain a fixed location for a sound’s SLP *in response to the movements of the head orientation of the first individual localizing the voice of the second individual during the electronic call*. See Section XIV.A-B, D, XVI.D.2.

342. Moreover, it is my opinion that a POSA would have understood that changing HRTFs in this context involves *adjust[ing] HRTFs of the first individual to more accurately simulate real HRTFs of the first individual*. As Ballas explains, “[a] more realistic experience is observed when listening to spatialized sound than stereo because stereo only varies across one axis.” Ex. 1008, ¶2. Because head tracking and adjusting HRTFs would have provided a more realistic virtual experience for the user, by allowing the user to move while fixing the perceived origin of sound sources. Ballas explains that “[t]he HRTF filter maintains virtual location of the audio signals, thus allowing the listener to changes locations and head orientation without degradation of the audio signal.” *Id.*, ¶17. This “produces virtual sound sources that are externally perceived and positioned at any desired orientation in azimuth and elevation from the listener.” *Id.*

3. Claim 12

- a. 12[a]:** “*The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: add a third individual to the electronic call;*”

343. It is my opinion that Li discloses this limitation.

344. Li teaches that “the configurable 3D sound system 900 provides a method of communication among multiple people, for example, placing different speaking users at different spots to mimic a real conference room environment.” Ex. 1006, 22:1-5. Li’s *electronic call* has the form of a conference call, where the participants include at least a first user who “plac[es] different speaking users” in the sound field. *Id.*, 10:5-6 (“The computing device is, for example, a portable device such as a mobile phone, a smart phone”), 21:57-60 (“the configurable 3D sound system 900 discloses herein . . . also enhances the performance of radio dramas and conference calls.”). Thus, Li discloses *add[ing] a third individual to the electronic call*.

- b. 12[b]:** “*designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the third individual occurs for the first individual through the electronic earphone such that the sound localization point of the third individual is adjacent to the sound localization point of the second individual.*”

345. It is my opinion that Li discloses this limitation.

346. For the same reasons that I explained above for 1[b] in Ground 5, 6[b] in Ground 5, and 8[b] in this ground, Li discloses that a “user can place the sound sources in specific locations” using a GUI that can “mimic a real conference room environment” for the user through headphones or a headset, a microphone array system, and individual microphones, and thus teaches *designat[ing], for the first individual, an SLP that simulates a location in empty space from where an origin of a voice of the third individual occurs for the first individual through the electronic earphone. See Sections XIII.C.3, XIII.D.5, XVI.C.3.*

347. Li further teaches that the SLP of the third individual is adjacent to the SLP of the second individual. Li teaches that “the configurable 3D sound system 900 provides a method of communication among multiple people, for example, placing different speaking users at different spots to mimic a real conference room environment.” Ex. 1006, 22:1-5. In a conference room environment, participants are adjacent to each other in order to speak face-to-face. Thus, a POSA would have understood that to simulate a conference room environment, the sound sources of the participants would have been adjacent to each other.

4. **Claim 14:** *“The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising: receive, from the first individual, commands that move the sound localization point from a first point in empty space that is away from and proximate to the first individual to a second point in*

empty space that is away from and proximate to the first individual.”

348. It is my opinion that Li discloses this limitation.

349. Specifically, Li teaches that its configurable 3D sound system 900 “acquires configuration information 1601 from a user.” This configuration information includes “user selections of configurable parameters, for example, an azimuth, an evaluation, a distance, a trace of movement, etc.” and “give[s] the user the freedom of configuring the positions and movements of multiple sound tracks.” Ex. 1006, 20:20-36, *see also* 15:27-32. By receiving user input to move the sound source, the system of Li will *receive, from the user (first individual) commands that move the SLP.*

350. The configuration information (*commands*) “allow[] the user to move the sound source around the user 1703, move the sound source up and down, etc., on the configurable sound field 1700.” *Id.*, 21:57-22:5. Li explains that “[t]he 3D sound processing provides the GUI to allow the user the freedom to configure the positions and movements of sources, in order to generate customized 3D sound.” *Id.*, 15:29-32. Using the GUI, the user can move a sound source from one location to another. *Id.*, 15:44-47 (“the user can place the sound sources in specific locations, dynamically move the sound sources, focus on or zoom in on one sound source and reduce others, etc., on the generated configurable sound field via the GUI”).

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,226,090

351. Therefore, as Li allows the user to move the sound source, it is my opinion that Li discloses *receiving, from the first individual, configurable parameters and configuration information (commands) to move the position (SLP of the voice) from the first point in an object-free space (empty space) to a second location in the object-free space (empty space) that is next to (proximate) the first location and a distance from (away from and proximate to, under 8KHZ's position) the user.*

XVII. SECONDARY CONSIDERATIONS

352. I am not aware of any secondary considerations of non-obviousness for the claims of the '090 patent. If Patent Owner raises any secondary considerations in its briefing or expert's declaration, I reserve the right to respond to those considerations.

XVIII. CONCLUSION

353. For these reasons, it is my opinion that claims 1-3, 6, 8-10, 12, 14-19 of the '090 patent are obvious.

Executed on June 8, 2023 in Los Angeles, California.



Dr. Chris Kyriakakis