IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re *Inter Partes* Review of:	)
U.S. Patent No. 9,282,196	)
Issued: Mar. 8, 2016	)
Application No.: 14/946,806	)
Filing Date: Nov. 20, 2015	)

For: **Moving a Sound Localization Point of a Computer Program during a Voice Exchange**

FILED VIA P-TACTS

DECLARATION OF CHRIS KYRIAKAKIS IN SUPPORT OF PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,282,196

TABLE OF CONTENTS

I. INTRODUCTION 10

II. EXPERIENCE, QUALIFICATIONS, AND STATUS AS AN INDEPENDENT EXPERT WITNESS 11

III. UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES 16

- A. Claim Construction 16
- B. Obviousness 17

IV. SUMMARY OF MATERIALS REVIEWED AND CONSIDERED 19

V. LEVEL OF SKILL IN THE ART AND PERSPECTIVE APPLIED IN THIS DECLARATION 22

VI. BACKGROUND 23

- A. Overview of the Technology 23
  - 1. Wearable virtual reality 23
  - 2. Binaural sound 27
  - 3. Sound source localization 32
  - 4. Creating binaural sound 36
  - 5. Compensating for head movement 39
- B. The'196 Patent 43
- C. Prosecution History 44
- D. Challenged Claims of the '196 Patent 45
- E. Effective Priority Date 47

VII. CLAIM CONSTRUCTION 47

VIII. THE PRIOR ART 48

- A. Li 48
- B. Bedingfield 49
- C. Ballas 50
- D. Sandgren 51
- E. Woo 52

IX. GROUND 1: LI RENDERS OBVIOUS CLAIMS 1, 8, 15, AND 17 52

A. Motivation to Modify 52
B. Reasonable Expectation of Success 55
C. Independent Claims 1 and 8: 56
1. 1[pre]/8[pre] 57
2. 1[a] /8[a] 63
3. 1[b]/8[b] 69
4. 1[c]/8[c] 70
5. 1[d] /8[d] 71
D. Independent Claims 15: 72
1. 15[Pre]: *An electronic system with one or more electronic devices that move a sound localization point (SLP) of a voice of a computer program during a voice exchange between the computer program and a person, the electronic system comprising.* 72
2. 15[a]: *a wearable electronic device that the person wears and that provides the voice of the computer program to the person at the SLP at a first location in empty space that is away from and proximate to the person;* 73
3. 15[b]: *a memory that stores instructions; and* 73
4. 15[c]: *a processor that executes the instructions to . . .* 74
5. 15[c1]: *receive a command from the person to move the SLP of the voice of the computer program from the first location in empty space to a second location in empty space that is proximate to the first location and away from and proximate to the person,* 74
6. 15[c2]: *move the SLP of the voice of the computer program in response to the command from the first location in empty space to the second location in empty space, and* 75

7. 15[c3]: *provide binaural sound to the person through the wearable electronic device during the voice exchange between the computer program and the person such that an origin of the voice of the computer program occurs at the SLP at the second location in empty space that is proximate to the first location and away from and proximate to the person.* 75

E. Dependent Claim 17: *processor further executes the instructions to change an interaural time difference (ITD) and an interaural level difference (ILD) of the binaural sound in order to move the SLP of the voice of the computer program from the first location in empty space to the second location in empty space* 75

X. GROUND 2: LI IN VIEW OF BALLAS RENDERS OBVIOUS CLAIMS 4, 6, 9-10, 14, 18 77

A. Motivation to Combine Li with Ballas 77

B. Reasonable Expectation of Success 80

C. Dependent Claim 4: 81

1. 4[a]: *tracking head orientation of user as user moves with respect to SLP at the second location* 81

2. 4[b]: *adjusting voice of computer program that user hears through electronic earphone so voice continues to appear to originate from SLP at the second location even as the user moves with respect to the second location.* 83

D. Dependent Claim 6: *changing an amplitude, an interaural time difference, and an interaural level difference of the voice of the computer program as the user moves with respect to the SLP at the second location so a sound of the voice of the computer program continues to appear to originate from the SLP at the second location as the user moves* 84

E. Dependent Claim 9: *maintain the origin of the voice of the computer program at the SLP at the second location in empty space but alter an amplitude of the voice of the computer program that the person hears in response to the person moving closer to or away from the SLP at the second location* 86

F. Dependent Claim 10 86

1. 10[a]: *monitor movements of head orientation of person in response to person moving with respect to SLP at second location in empty space; and* 86
2. 10[b]: *adjust an interaural time difference (ITD) and an amplitude of the voice of the computer program in response to the movements of the head orientation of the person as the person moves* 86

G. Dependent Claim 14: *adjust interaural level difference, interaural time difference, and amplitude for voice person hears so SLP of voice remains fixed at second location in empty space while head orientation of person moves* 87

H. Dependent Claim 18: *processor further executes instructions to change interaural time difference (ITD) and amplitude of binaural sound to maintain SLP fixed at second location in empty space during voice exchange while person physically moves away from second location in empty space and changes head orientation with respect to second location in empty space* 87

XI. GROUND 3: LI IN VIEW OF BALLAS AND SANDGREN RENDERS OBVIOUS CLAIM 5 89

A. Motivation to Combine 89
B. Reasonable Expectation of Success 90
C. Dependent Claim 5: *adjust amplitude of voice of computer program to be inversely proportional to square of distance between user and SLP at second location in empty space as user moves with respect to SLP at second location in empty space* 91

XII. GROUND 4: LI IN VIEW OF SANDGREN RENDERS OBVIOUS CLAIM 19 92

A. Motivation to Combine 92
B. Reasonable Expectation of Success 95
C. Dependent Claim 19: 95

1. 19[a]: *The electronic 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 person,* 95
2. 19[b]: *wherein virtual objects do not actually exist near person but are perceived to exist by person from artificial audial cues.* 97

XIII. GROUND 5: LI AND WOO RENDERS OBVIOUS CLAIM 11 .. 98

A. Motivation to Combine 98
B. Reasonable Expectation of Success 99
C. Dependent Claim 11: 100
1. 11[a]: *determine physical objects located in a room with the person* 100
2. 11[b]: *adjust, based on physical objects located in room, the voice of computer program that person hears to emulate sound that person would hear if voice of computer program actually originated from SLP at second location.* 101

XIV. GROUND 6: BEDINGFIELD RENDERS OBVIOUS CLAIMS 1, 8, 15, and 17 102

A. Motivation to Modify 102
1. *Voice of a computer Program.* 102
2. *HPED.* 103
B. Reasonable Expectation of Success 104
1. *Voice of a computer Program.* 104
2. *HPED.* 105
C. Independent Claims 1 and 8: 105
1. 1[Pre]/8[pre] 105
2. 1[a]/8[a] 110
3. 1[b]/8[b] 113
4. 1[c]/8[c] 115
5. 1[d]/8[d] 115
D. Independent Claims 15: 116

1. 15[Pre]: *An electronic system with one or more electronic devices that move a sound localization point (SLP) of a voice of a computer program during a voice exchange between the computer program and a person, the electronic system comprising.* 116
2. 15[a]: *a wearable electronic device that the person wears and that provides the voice of the computer program to the person at the SLP at a first location in empty space that is away from and proximate to the person;* 117
3. 15[b]: *a memory that stores instructions; and* 117
4. 15[c]: *a processor that executes the instructions to . . .* 118
5. 15[c1]: *receive a command from the person to move the SLP of the voice of the computer program from the first location in empty space to a second location in empty space that is proximate to the first location and away from and proximate to the person,* 118
6. 15[c2]: *move the SLP of the voice of the computer program in response to the command from the first location in empty space to the second location in empty space, and* 118
7. 15[c3]: *provide binaural sound to the person through the wearable electronic device during the voice exchange between the computer program and the person such that an origin of the voice of the computer program occurs at the SLP at the second location in empty space that is proximate to the first location and away from and proximate to the person* 119

E. Dependent Claim 17: *processor further executes the instructions to change an interaural time difference (ITD) and an interaural level difference (ILD) of the binaural sound in order to move the SLP of the voice of the computer program from the first location in empty space to the second location in empty space* 119

XV. GROUND 7: BEDINGFIELD IN VIEW OF BALLAS RENDERS OBVIOUS CLAIMS 4, 6, 9-10, 14, AND 18 120

A. Motivation to Combine Bedingfield with Ballas 120

B. Reasonable Expectation of Success 122
C. Dependent Claim 4 123
1. 4[a]: *tracking head orientation of user as user moves with respect to SLP at the second location* 123
2. 4[b]: *adjusting voice of computer program that user hears through electronic earphone so voice continues to appear to originate from SLP at the second location even as the user moves with respect to the second location.* 124
D. Dependent Claim 6: *changing an amplitude, an interaural time difference, and an interaural level difference of the voice of the computer program as the user moves with respect to the SLP at the second location so a sound of the voice of the computer program continues to appear to originate from the SLP at the second location as the user moves* 125
E. Dependent Claim 9: *maintain the origin of the voice of the computer program at the SLP at the second location in empty space but alter an amplitude of the voice of the computer program that the person hears in response to the person moving closer to or away from the SLP at the second location* 126
F. Dependent Claim 10 126
1. 10[a]: *monitor movements of head orientation of person in response to person moving with respect to SLP at second location in empty space; and*..... 126
2. 10[b]: *adjust an interaural time difference (ITD) and an amplitude of the voice of the computer program in response to the movements of the head orientation of the person as the person moves* 126
G. Dependent Claim 14: *adjust interaural level difference, interaural time difference, and amplitude for voice person hears so SLP of voice remains fixed at second location in empty space while head orientation of person moves* 127

H. Dependent Claim 18: *processor further executes instructions to change interaural time difference (ITD) and amplitude of binaural sound to maintain SLP fixed at second location in empty space during voice exchange while person physically moves away from second location in empty space and changes head orientation with respect to second location in empty space* 127

XVI. GROUND 8: BEDINGFIELD IN VIEW OF BALLAS AND SANDGREN RENDERS OBVIOUS CLAIM 5 128

A. Motivation to Combine 128
B. Reasonable Expectation of Success 129
C. Dependent Claim 5: *adjust amplitude of voice of computer program to be inversely proportional to square of distance between user and SLP at second location in empty space as user moves with respect to SLP at second location in empty space* 130

XVII. GROUND 9: BEDINGFIELD IN VIEW OF SANDGREN RENDERS OBVIOUS CLAIM 19 131

A. Motivation to Combine 131
B. Reasonable Expectation of Success 132
C. Dependent Claim 19 132
1. 19[a]: *The electronic 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 person,* 132
2. 19[b]: *wherein virtual objects do not actually exist near person but are perceived to exist by person from artificial audial cues.* 133

XVIII. GROUND 10: BEDINGFIELD AND WOO RENDERS OBVIOUS CLAIM 11 133

A. Motivation to Combine 133
B. Reasonable Expectation of Success 134
C. Dependent Claim 11 135

1. 11[a]: *determine physical objects located in a room with the person* 135
2. 11[b]: *adjust, based on physical objects located in room, the voice of computer program that person hears to emulate sound that person would hear if voice of computer program actually originated from SLP at second location.* 135

XVI. SECONDARY CONSIDERATIONS 136

XVII. CONCLUSION 136

I. INTRODUCTION

1. I have been retained by Meta Platforms, Inc. ("Petitioner") to provide my opinions concerning the validity of claims 1, 4-6, 8-11, 14-15, and 17-19 of U.S. Patent No. 9,282,196.

2. I am simultaneously engaged by Petitioner to provide declarations for IPR petitions challenging U.S. Patent No. 9,226,090, U.S. Patent No. 9,674,628, U.S. Patent No. 11,290,836, U.S. Patent No. 11,172,316, U.S. Patent No. 10,368,179, U.S. Patent No. 10,448,184, U.S. Patent No. 10,798,509, and U.S. Patent No. 10,917,737.

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, 4-6, 8-11, 14-15, and 17-19 of the '196 patent are invalid based on the following grounds:

Ground 1	Claims 1, 8, 15, and 17 are obvious under 35 U.S.C. § 103 over Li.
Ground 2	Claims 4, 6, 9-10, 14, and 18 are obvious under 35 U.S.C. § 103 over Li in view of Ballas.
Ground 3	Claim 5 is obvious under 35 U.S.C. § 103 over Li and Ballas in view of Sandgren.
Ground 4	Claim 19 is obvious under 35 U.S.C. § 103 over Li in view of Sandgren.
Ground 5	Claim 11 is obvious under 35 U.S.C. § 103 over Li in view of Woo.
Ground 6	Claims 1, 8, 15, and 17 are obvious under 35 U.S.C. § 103 over Bedingfield.
Ground 7	Claims 4, 6, 9-10, 14, and 18 are obvious under 35 U.S.C. § 103 over Bedingfield in view of Ballas.
Ground 8	Claim 5 is obvious under 35 U.S.C. § 103 over Bedingfield and Ballas in view of Sandgren.
Ground 9	Claim 19 is obvious under 35 U.S.C. § 103 over Bedingfield in view of Sandgren
Ground 10	Claim 11 is obvious under 35 U.S.C. § 103 over Bedingfield in view of Woo

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.

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 '196 patent. I have no financial interest in 8KHZ.

III. UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES

A. Claim Construction

19. 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).

20. 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*.

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

B. Obviousness

22. 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.

23. 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.

24. 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.

25. I understand that, when there is a reason to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill 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.

26. 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 '196 patent.

IV. SUMMARY OF MATERIALS REVIEWED AND CONSIDERED

27. All of the opinions contained in this Declaration are based on the documents I reviewed and my knowledge and professional judgment. In forming the opinions expressed in this Declaration, while drawing on my experience in the field of audio signal processing and sound localization (including my teaching, research, and consulting experiences), I reviewed the '196 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. 9,282,196 ("'196 patent")
1002	Prosecution History of U.S. Patent No. 9,282,196 ("'196 FH")

Exhibit	Description
1004	*Curriculum Vitae* 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")
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")
1014	U.S. Patent Pub. 2002/0151996 to Wilcock ("Wilcock")
1015	Rumsey, Francis "Whose head is it anyway? Optimizing binaural audio." J. Audio Eng. Soc., Vol. 59, No. 9, 2011 ("Rumsey")
1016	U.S. Patent Pub. No. 2009/0060235 to Woo ("Woo")
1017	U.S. Patent App. Pub. 2008/0273708 to Sandgren, et al. ("Sandgren")
1021	Comeau, Charles et al. "Headsight TV System." Electronics, 1961, pp. 86-90 ("Comeau")
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")

Exhibit	Description
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 ("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")
1040	Berkley, D.A., et al. "A Multimodal Teleconference System Using Hands-Free Voice Control." 3rd International Conference on Spoken Language Processing, 1994 ("Berkley")
1041	Fagel, Sascha, et al. "An Expandable Web-based Audiovisual Text-to-Speech Synthesis System." 8th European Conference on Speech Communication and Technology, 2003 ("Fagel")
1042	Taylor, Paul, Text-to-Speech Synthesis (Cambridge University Press 2009)
1044	U.S. Patent No. 2,955,156 to Heilig ("Heilig")

28. 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.

29. 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 '196 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

30. 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.

31. The invention of the '196 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, 1:13-16.

32. 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 '196 patent would have had at least a bachelor's degree in computer science or engineering with a focus on audio signal processing and sound localization, 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.

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

34. 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 ’196 patent. I believe that I have a sufficient level of knowledge, experience, and education to provide an expert opinion in the field of the ’196 patent.

35. 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 ’196 patent, the file history of the ’196 patent, and the challenged claims.

VI. BACKGROUND

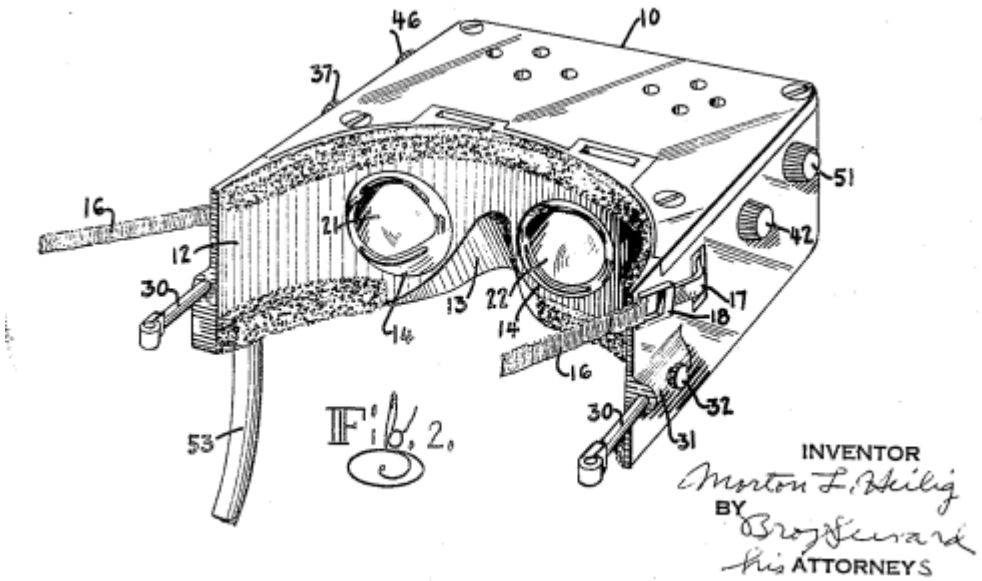
A. Overview of the Technology

1. Wearable virtual reality

36. 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.

37. 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.



38. 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]henever 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.*

39. 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.

40. 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.

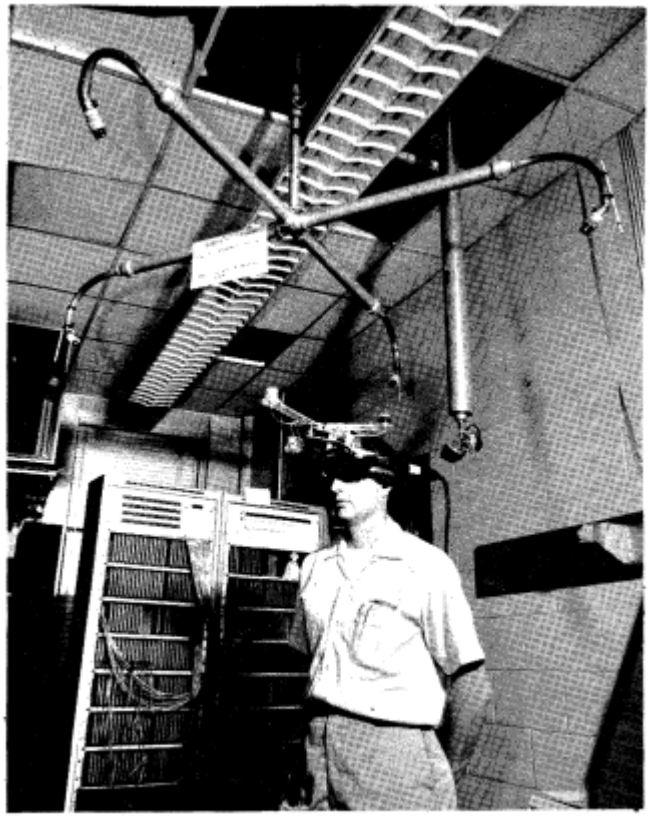


FIGURE 3—The mechanical head position sensor in use

2. Binaural sound

41. 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*.

42. 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).

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

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).

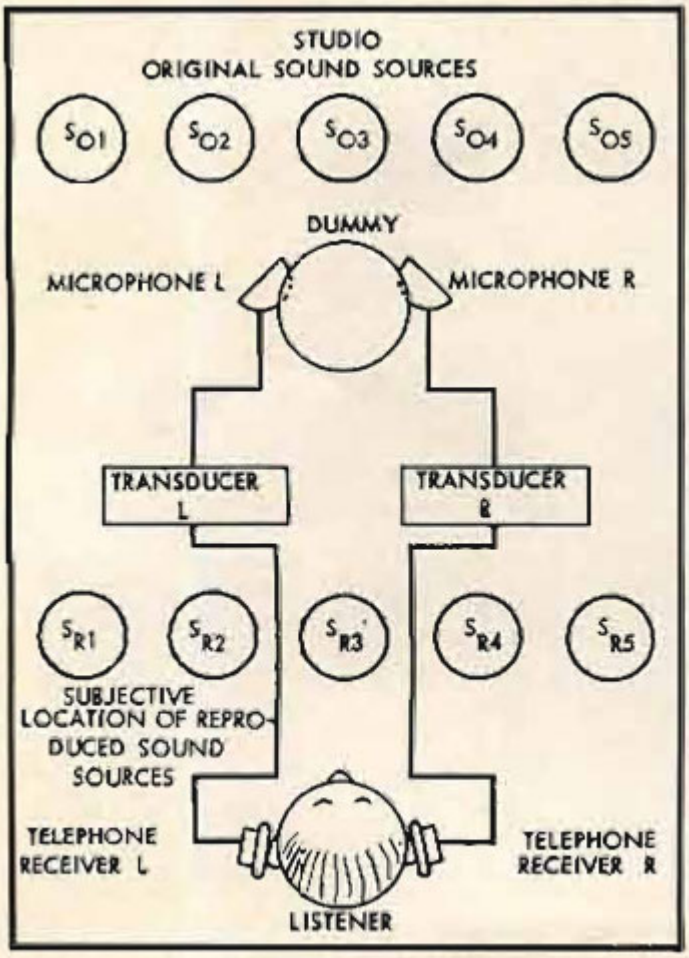


Fig. 2. Binaural.

44. 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.

45. 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.

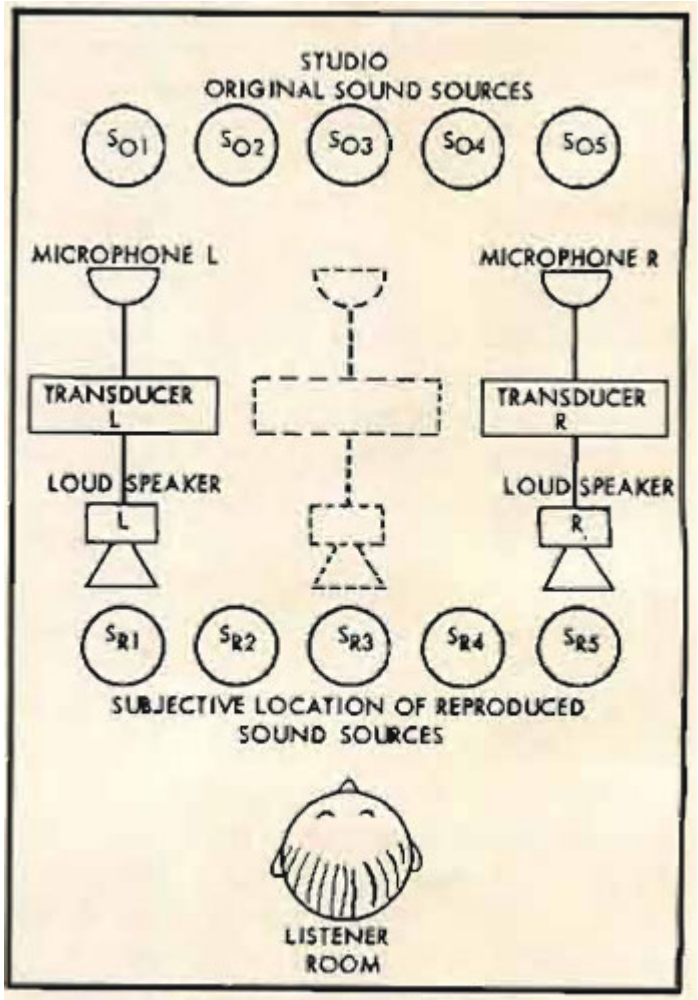


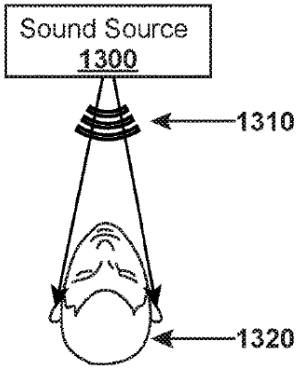
Fig. 4. Stereophonic.

3. Sound source localization

46. 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.

47. 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.*

48. ***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 '196 patent.



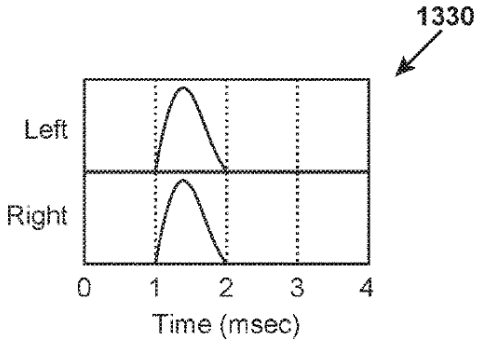
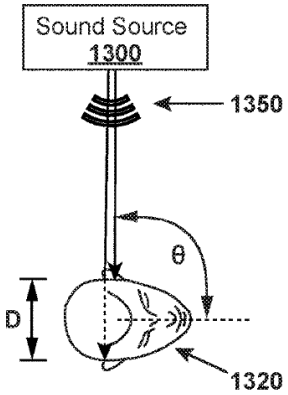


Figure 13A **Figure 13B**

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 '196 patent.



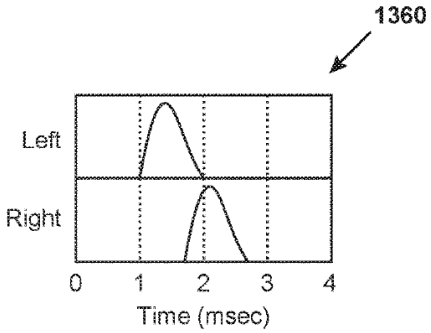


Figure 13C **Figure 13D**

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.

49. ***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.

50. ***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).

51. 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

52. 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.

53. 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, 25; *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, 25. 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, 25.

54. 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.

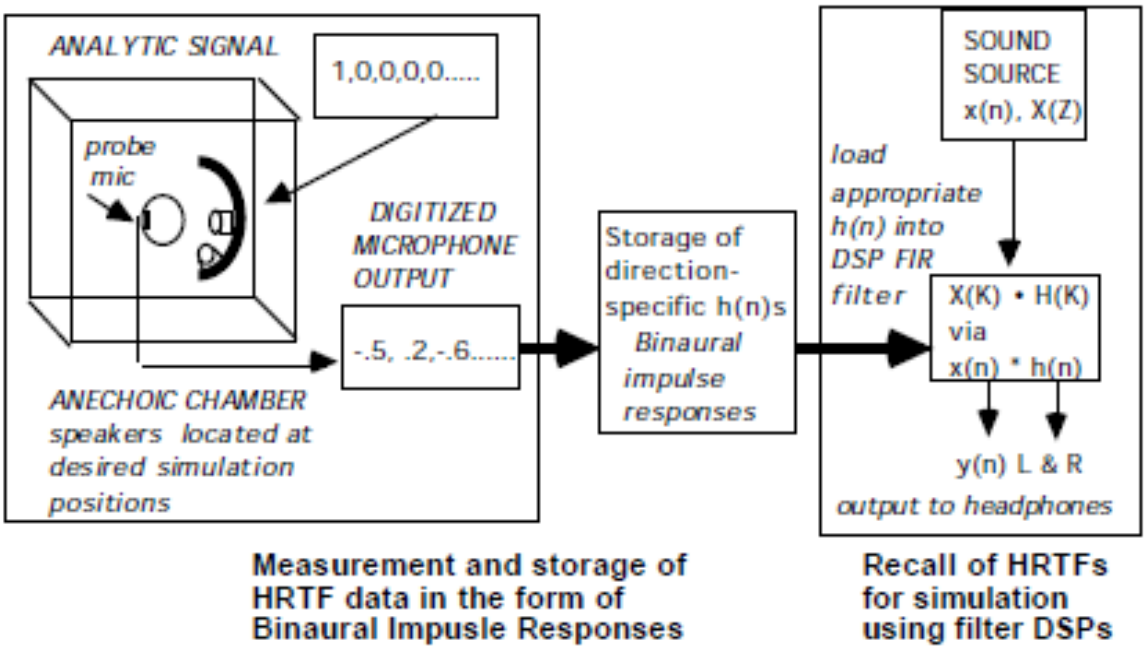
55. It was commonly known to measure HRTFs, by the "[m]easurement of [b]inaural [i]mpulse [r]esponses," Durand Begault described in his April 2000 book "3-D Sound for Virtual Reality and Multimedia":

> "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.

56. 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.



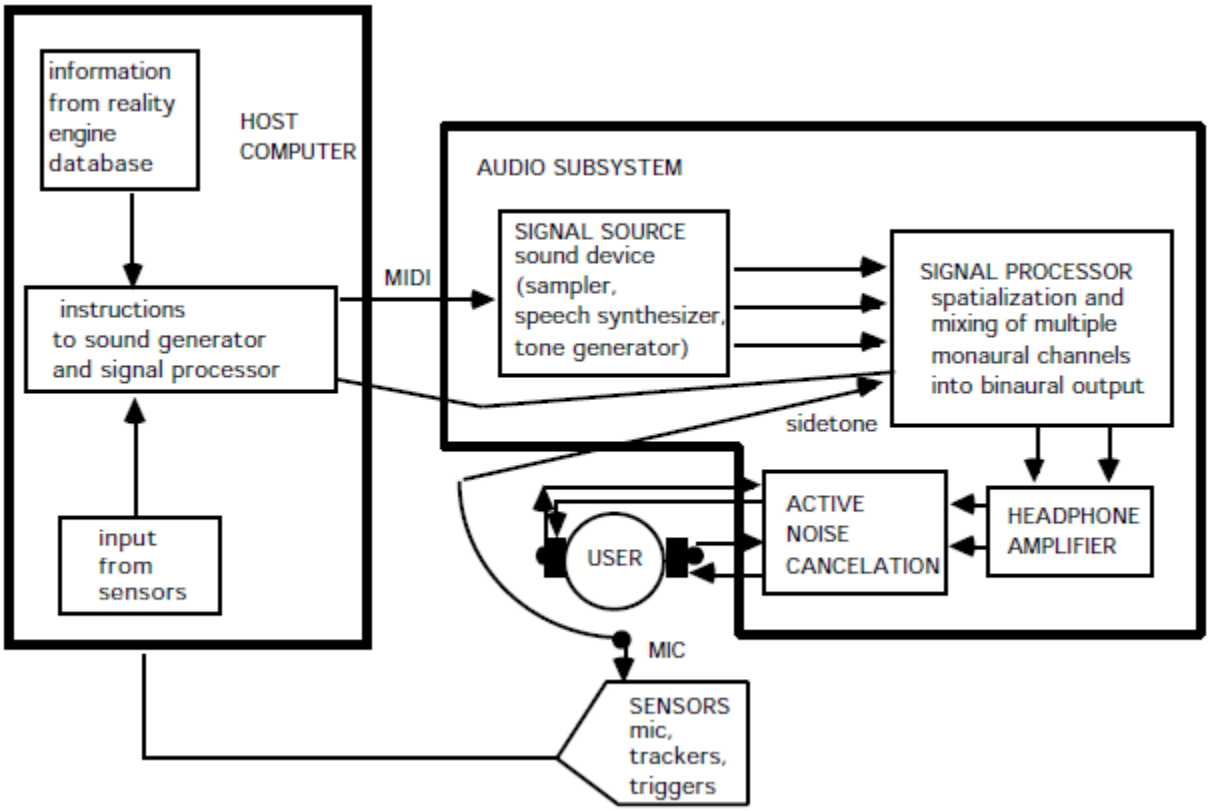
57. Such 3-D sound was useful in various virtual reality environments, such as videoconferencing. *Id.*, 174 (describing arranging spatialized audio signals in

virtual space for multiconversation teleconferencing with broadcast video images of the participants).

5. Compensating for head movement

58. 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.

59. 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).

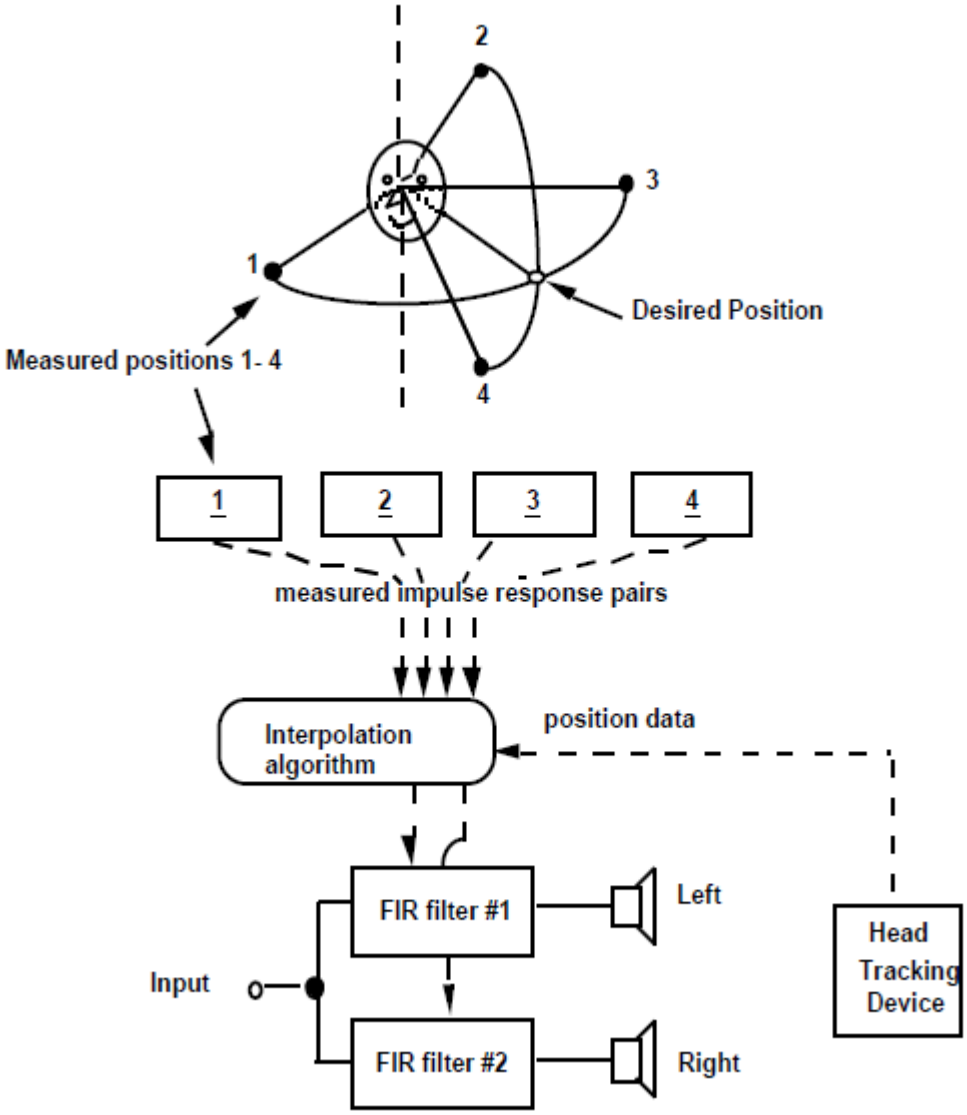


60. These "trackers" were for tracking the head movement of the listener, to adjust the binaural sound delivered to the listener. This was done so that the perceived location of the sound source remains fixed in virtual space, regardless of the listener's head movement:

> "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.

61. 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.

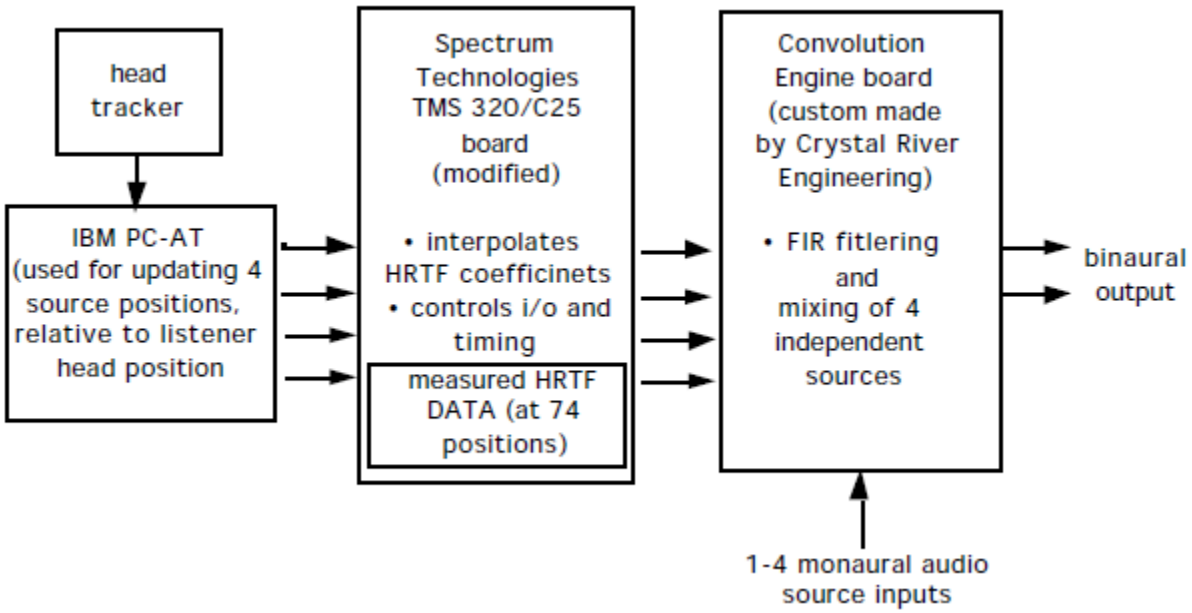


62. It was commonly known that the benefits of such systems was that the perceived location of the sound source can remain fixed in virtual space, regardless of the listener's head movement. Doing so would mimic spatial hearing in natural environments and enhance the immersivity of the environment:

> "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).

63. 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.").



64. Accordingly, in my opinion it was known well before the relevant time frame to (1) process and provide binaural sound so that it appears to originate from an SLP in an empty space, (2) move the SLP, and (3) maintain that SLP as the user's head changes position.

B. The'196 Patent

65. The '196 patent "is a computer system in which two electronic devices communicate with each other during a telephone call between two people." *Id.*, 2:42-46. "The computer system positions a voice of one person at a sound localization point that is proximate to the other person." *Id.* "In order to determine a location of sound, humans process and compare monaural cues from each other." Id., 2:53-54. "This comparison reveals difference cues or binaural cues that enable sound localization." *Id.*, 54-56. "These cues include interaural time differences

(ITDs), interaural level differences (ILDs), and head-related transfer functions (HRTFs)." *Id.*, 56-58. "Cues thus result from changes to the sound wave from an interaction with the human anatomy." *Id.*, 58-59. "Humans process these changes as impulse responses (HRIRs). The HRIRs and transformed HRTFs enable sound to be convolved such that a location of the sound or sound localization point can be changed." *Id.* 2:63-65. Ex. 1001, Abstract, Figure 1.

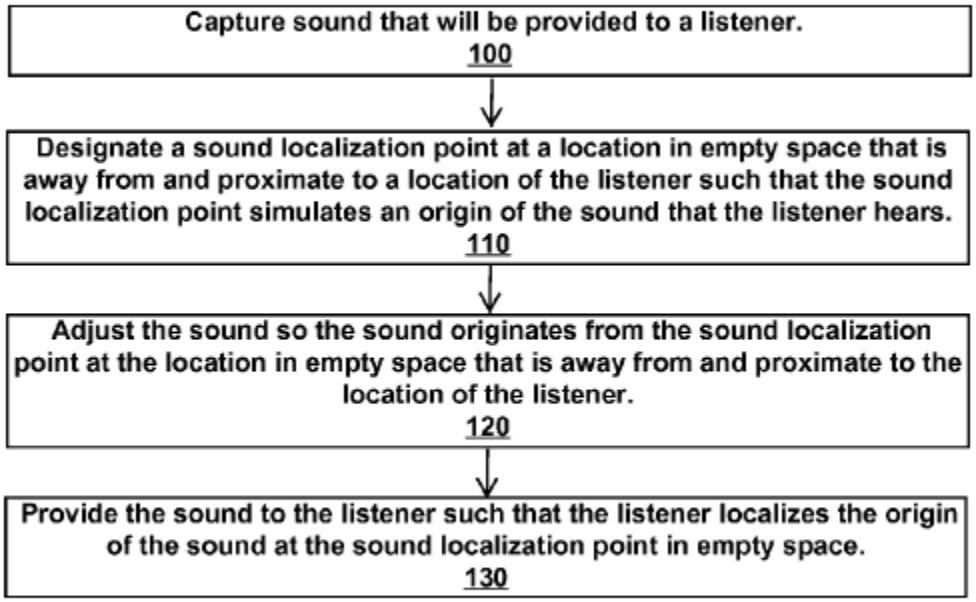


Figure 1

C. Prosecution History

66. I have reviewed the prosecution history of the '196 patent, and based on the prosecution history, the applicant obtained allowance of the '196 patent without required amendment by the Examiner. Ex. 1002, 141.

67. On January 27, 2016, the Examiner issued a Notice of Allowance and Fees Due, in which s/he allowed all twenty claims without amendment, stating, "These features, in combination with the other features of Claims 1, 8 and 15 (claims 21, 28, and 35 during prosecution) are not anticipated by, nor made obvious over, the prior art of record. *Id.*, 142. Specifically, the Examiner indicated that the novel claim element is:

> ". . . providing, through the electronic earphone being located at the head of the user and being in communication with the HPED, the binaural sound to the user during the voice exchange between the computer program and the user such that the origin of the voice of the computer program occurs at the SLP at the second location in empty space that is proximate to the first location and away from and proximate to the user."

68. In Comments on Statement of Reasons for Allowance date January 29, 2016, the Applicant noted that the novel element indicated by the Examiner is present independent claim 1, but not in independent claims 8 and 15.

D. Challenged Claims of the '196 Patent

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

70. The '196 patent has 20 claims, 3 independent claims and 17 dependent claims. Independent claim 1 is a method claim reciting a 4-step process:

(1) providing a binaural sound to the user such that an origin of the voice of the computer program occurs at a first location in empty space that is away from and proximate to the user;

(2) receiving a command to move the SLP r;

(3) moving the SLP from the first location in empty space to the second location in empty space; and

(4) providing the binaural sound to the user such that the origin of the Voice of the computer program occurs at the SLP at the second location in empty space;

71. Independent claim 8 is a process claim reciting a 4-step process:

(1) provid[ing], through a wearable electronic device on the person, binaural sound that occurs at the SLP at a first location in empty space that is away from and proximate to the person;

(2) receiv[ing], from the person, a command to move the SLP to a second location;

(3) mov[ing], in response to the command, the SLP from the first location in empty space to the second location in empty space; and

(4) provid[ing], through the wearable electronic device on the person, the binaural sound such that the origin occurs at the second location in empty space.

72. Independent claim 15 is a process claim reciting a 3-step process:

(1) a wearable electronic device that provides the voice of the computer program to the person at the SLP at a first location in empty space that is away from and proximate to the person;

(2) a memory that stores instructions; and

(3) a processor that receives a command to move the SLP from the first location in empty space to a second location in empty space moving the SLP in response to the command and providing binaural sound through the wearable electronic device such that the voice occurs at the SLP at the second location in empty space.

E. Effective Priority Date

73. I understand that the '196 patent was filed on November 20, 2015. I understand that cover of the '196 patent includes a reference to a non-provisional application filed on June 23, 2014. For the purpose of my analysis, I have been asked to assume that the challenged claims of the '196 patent are entitled to a priority date of June 23, 2014.

VII. CLAIM CONSTRUCTION

74. I understand that Patent Owner contends certain claim elements in the '196 patent should have their plain and ordinary meaning. Specifically, I understand

that 8KHZ's position is that if construed, the below terms in the '196 patent should have the following meanings:

- *away from and proximate to* in cls. 1, 4, 15, 19 should mean "a distance from"
- *wearable electronic device / WED* in cls. 15, 19 should mean "portable wireless electronic device that is worn on or attached to a person"
- *handheld portable electronic device / HPED* in cls. 1, 4 should mean "portable wireless electronic device that is held in the hand of 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

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

A. Li

76. U.S. Patent No. 9,131,305 to 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 in its conference call embodiment except *the voice of the computer program* which it discusses separately. Li was not cited or discussed during the '196 patent prosecution.

77. Li is generally directed to a 3D sound system that allows users to "experience[e] music, movies, [and] video games" and to provide user inputs to determine "movements of the sound sources." Ex. 1006, 7:21-22, 15:8, 16:4-7, 16:63-66, 19:39-20:52, Figs. 9, 16, 33.

78. Like the '196 patent, in its conferencing system embodiment, Li "generate[s] configurable 3D sounds," which involves processing the speaker's voice to position "different speaking users at different spots." *Id.*, 22:1-5, 33:14-18, 33:32-51, 33:62-66, 34:4-8, 34:20-51, 35:58-36:2. Li further describes "generat[ing] a configurable sound field that provides an interface to give the user the freedom of configuring the positions and movements of multiple sound tracks, in order to render a customized 3D binaural sound" (*id.*, 20:26-36). By allowing 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," Li discloses moving the SLP of a sound based on the user's command. *Id.*, 20:20-36, 21:57-22:5.

B. Bedingfield

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

all the limitations of the independent challenged claims in its conference call system except *the voice of the computer program*. Bedingfield was not cited or discussed during the '196 patent prosecution.

80. 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. In particular, Bedingfield describes an "improved telephone or network-based communication system [that] may more closely simulate an in-person conference setting." *Id.*, 3:25-27.

81. Like the '196 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's system allows the user to make selections that include new locations for the other participants in the sound field, meaning their SLPs move within the sound field delivered to the user. *Id.*, 5:51-61, 6:43-45, Fig. 3.

C. Ballas

82. U.S. Patent Pub. No. 2003/0059070 to Ballas, published March 27, 2003 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 required in certain dependent claims. Ballas was not cited or discussed during the '196 patent prosecution.

83. Like the '196 patent, Ballas "allow[s] the listener to change locations and head orientation without degradation of the audio signal." Ex. 1008, ¶¶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.

D. Sandgren

84. 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 an inverse-square relationship between sound intensity and distance, as well as changing reverberation and attenuation to simulate 3D audio required for certain dependent claims. Ex. 1017, ¶¶6, 12, 77. Sandgren was not cited or discussed during the '196 patent prosecution.

85. Like the '196 patent, Sandgren simulates sound-reflecting objects and contemplates what happens 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 and their interaction with the room boundaries. *Id.*, ¶¶31-33, 69.

E. Woo

86. U.S. Patent Pub. No. 2009/0060235 to Woo, published March 5, 2009 is prior art under at least 35 U.S.C. §§102(a)(1) and (a)(2). Woo discloses detecting and photographing actual objects in the user's room and adjusting a sound output based on the object's position in the room as required by certain dependent claims. Ex. 1016, ¶¶3, 53-54, Fig. 5. Woo was not cited or discussed during the '196 patent prosecution.

87. Like the '196 patent, Woo adjusts properties of sound to provide a sound effect, and does so based on the presence of objects in the user's room. *Id.*, ¶3, Abstract.

IX. GROUND 1: LI RENDERS OBVIOUS CLAIMS 1, 8, 15, AND 17

A. Motivation to Modify

88. In my opinion, it would have been obvious to a POSA to modify the conference call embodiment of Li to provide *a voice of a computer program* as taught in other parts of Li.

89. A POSA would have recognized that there are several advantages of a computer program synthesizing a voice to provide a *voice of a computer program* as

Li's sound source or binaural sound. For example, a computerized text-to-speech synthesis would have allowed a speaking-impaired person to participate in Li's conference call system. Therefore, the speaking-impaired may similarly participate in realistic conference calls.

90. In the relevant time frame, text-to-speech ("TTS") was an area of study involving a computer reading text and producing speech. Ex. 1042, 1; Ex. 1014. TTS is commonly used for telephone information systems, aiding those with speech impairments, and reading to the blind. *Id.*, 1, 2; Ex. 1014, [0051]. Researchers have been developing text-to-speech systems for decades and the concept of generating computerized voice from text dates back even further. Id., xix-xx, 1 ("The idea of artificially generated speech has of course been around for a long time . . . ."). Early TTS systems were limited due to by the robotic-sounding voice that was produced, but improvement to the computerized voice, due largely to the development of diphone libraries (i.e. recordings of speech sounds), has led to the wider incorporation of text-to-speech into various applications including those that aid remote distance conversation. Id., xix-xx, 2, 5.

91. By the mid-1990s, researchers and developers began to implement text-to-speech for teleconferences. Ex. 1040, 555. The HuMaNet system, for example, used a "general purpose text-to-speech synthesis system which can convert an arbitrary text message to an intelligible voice message . . . ." Id., 556. During the five

years preceding the introduction of the HuMaNet system, researchers developed "algorithmic as well as hardware advances [that] enabled TTS algorithm to migrate from general purpose workstations . . . to medium processing power PC" which allowed for easy implementation of the HuMaNet system in order to connect users via teleconference using their everyday computer devices. *Id*.

92. Later conferencing systems implementing text-to-speech broadened the application, integrating the text-to-speech function with a video system for video conferencing. See, Ex. 1041, 1. For example, the MASSY (Modular Audiovisual Speech Synthesizer) system read text and processed it with an audio synthesizer and a visual articulation module to develop both computer audio and video of an artificial person's head that appears to speak the computer audio. *Id*. 1, 2. The developers of MASSY recognized that text-to-speech audio had improved such that cues in natural human communication could be conveyed, but such audio-only systems failed to convey information associated with the visuals of communication. *Id*., 1. The MASSY system improved upon older systems by providing those missing visual cues. *Id*.

93. Another advantage is that computerized voice synthesis would enable Li's system to control for voice quality, speed, and volume to improve clarity and minimize potential miscommunication, while obviating a human speaker recording every possible utterance, thus saving time, cost, and storage space. Furthermore, a

computer-synthesized voice adds flexibility by removing the physical constraints of human speech and adapting to different languages, accents, and emotions, and to different vocal ranges. A user may therefore communicate with anyone in any manner the user chooses.

94. In my opinion, the modification would have been a natural extension of Li's disclosure of "3D sound…synthesis and reproduction" via sound generating devices or sound emitting devices. Ex. 1006, 3:11-16, 3:39-44, 6:67-7:4. Li contemplates using its system in "a simple videogame" and "sophisticated virtual reality simulators" (Ex. 1006, 1:32-35, 7:17-29, 10:3-11, 35:4-13). It is my opinion that using *a voice of a computer program* would enhance the user's experience.

B. Reasonable Expectation of Success

95. In my opinion, a POSA would have reasonably expected success modifying the conference call embodiment of Li to include *a voice of a computer program*. Li describes a portable device and wearable headphones to provide binaural sound that appears to the user to originate from an SLP. Li further describes "3D sound synthesis and reproduction to allow users to experience 3D sound…through a headset," and doing so in a conference call setting. Ex. 1006, 7:1-4. I further read Li to contemplate using its 3D sound system in conference calls, videogames, and virtual reality simulators. Ex. 1006, 7:17-29, 10:3-11, 21:57-22:5, 35:4-13. Li recognizes that sound generating devices or sound emitting devices

include "a person's voice [and] pre-recorded multiple sound tracks." *Id.*, 4:58-64, 14:40-46. Furthermore, Li discloses sound generating devices and different sound sources, including musical instruments, a person's voice and pre-recorded sounds. Ex. 1006, 4:58-64, 14:40-46.

96. Accordingly, I believe a POSA would have understood that although Li does not expressly refer to *a voice of a computer program* as part of the conference call embodiment, Li teaches using such a voice in relation in other applications and adding a *voice of a computer program* to a conference call would save time and cost, added flexibility, and improved clarity. Implementing Li's conference call embodiment with spatialized sound in Li would have been achieved by using a computer-generated voice as a sound source. Such an implementation would have involved merely combining known prior art elements (spatialization and computer-generated voice) according to known methods and techniques (applying HRTF filters) and the results of the combination would have been readily predictable (spatialized sound). In addition, the implementation would involve use of techniques that were widely known at the time (spatialization). In my opinion, a POSA would have performed the implementation to improve a similar device (wearable communication device) and would have done so in the same way (filtering sounds to introduce spatial cues).

C. Independent Claims 1 and 8:

1. 1[pre]/8[pre]

1[pre] A method executed by *one or more electronic devices in an electronic system that move a sound localization point (SLP) of a voice of a computer program during a voice exchange between the computer program and a user, the method comprising*:	8[pre] A non-transitory computer readable storage medium storing instructions that cause *one or more electronic devices in an electronic system* to execute a method *that moves a sound localization point (SLP) of a voice of a computer program during a voice exchange that is between the computer program and a person, the method comprising*:

97. In my opinion, to the extent the preamble is limiting, Li renders obvious a method executed by one or more electronic devices in an electronic system that move a sound localization point (SLP) of a voice of a computer program during a voice exchange between the computer program and a user.

98. Li discloses *one or more electronic devices in an electronic system* through its description of 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, 16, 33, reproduced below.

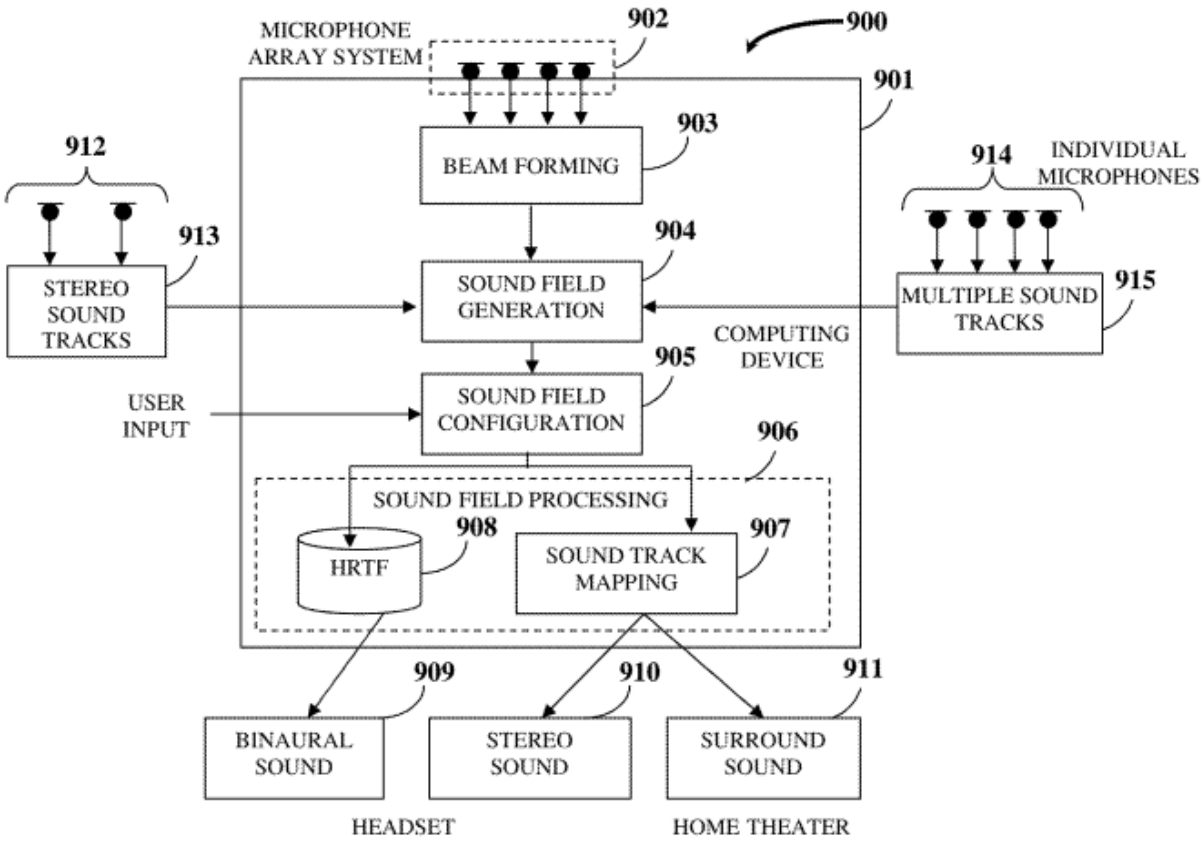


FIG. 9

As Li explains in relation to Fig. 9, the configurable 3D sound system 900 (*electronic system*) includes a computing device 901 (which the POSA would have known to include at least a processor and a memory, as it was well-known and commonplace for computing devices to include a memory storing instructions executed by a processor) and headphones or a headset (*one or more electronic devices*) that provide a "3D sound processing application" that "synthesizes and reproduces the user preferred sound field" and "performs sound track mapping 907 by convolving each

of the sound tracks with corresponding HRTFs… to synthesis [sic] 3D binaural sound 909 for a headset user." Ex. 1006, 17:1-4, 17:22-37.

99. Specific to element 8[pre], it is my opinion that Li discloses a *non-transitory computer readable storage medium storing instructions*. Li's configurable 3D sound system 900 includes computing device 901, which is an *electronic device*, and "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 (emphases added); *see also* 30:65-31:1, 35:24-50. Li's *electronic device* includes a memory unit 3402 (*non-transitory computer readable storage medium*) and a processor 3401 that executes the *instructions*.

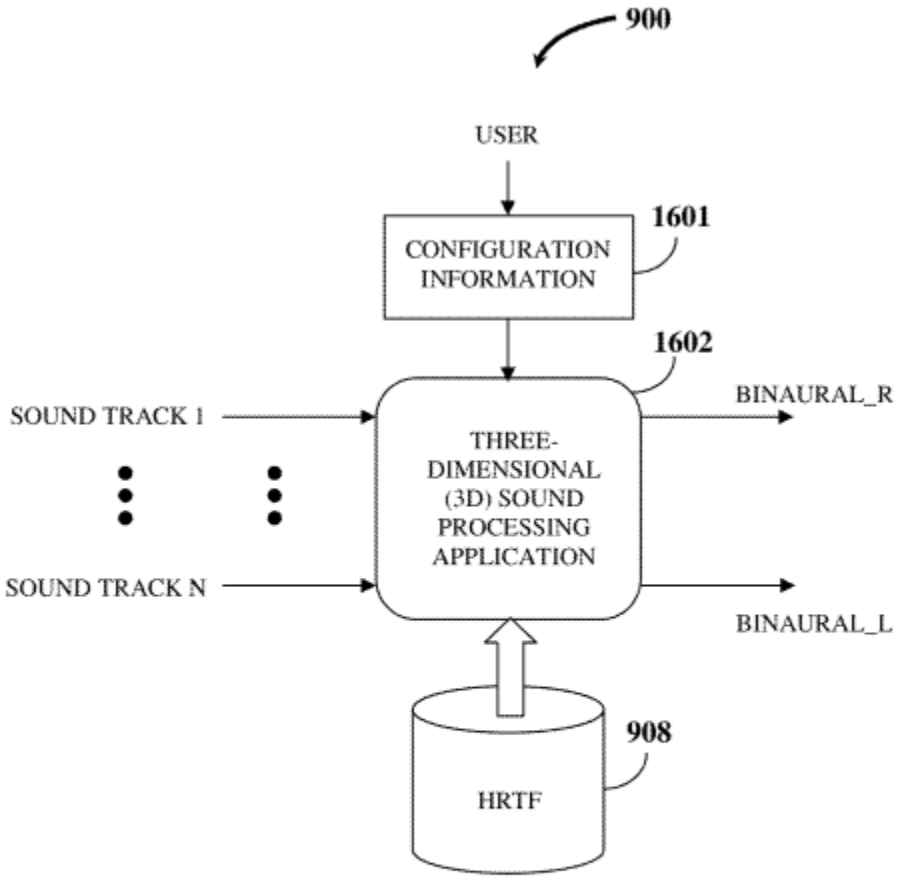


FIG. 16

100. Li further discusses the 3D sound processing application in relation to Fig. 16, which depicts the 3D sound processing application 1602 as "generat[ing] a configurable sound field that provides an interface to give the user the freedom of configuring the positions and movements of multiple sound tracks, in order to render a customized 3D binaural sound." *Id*., 20:26-36.

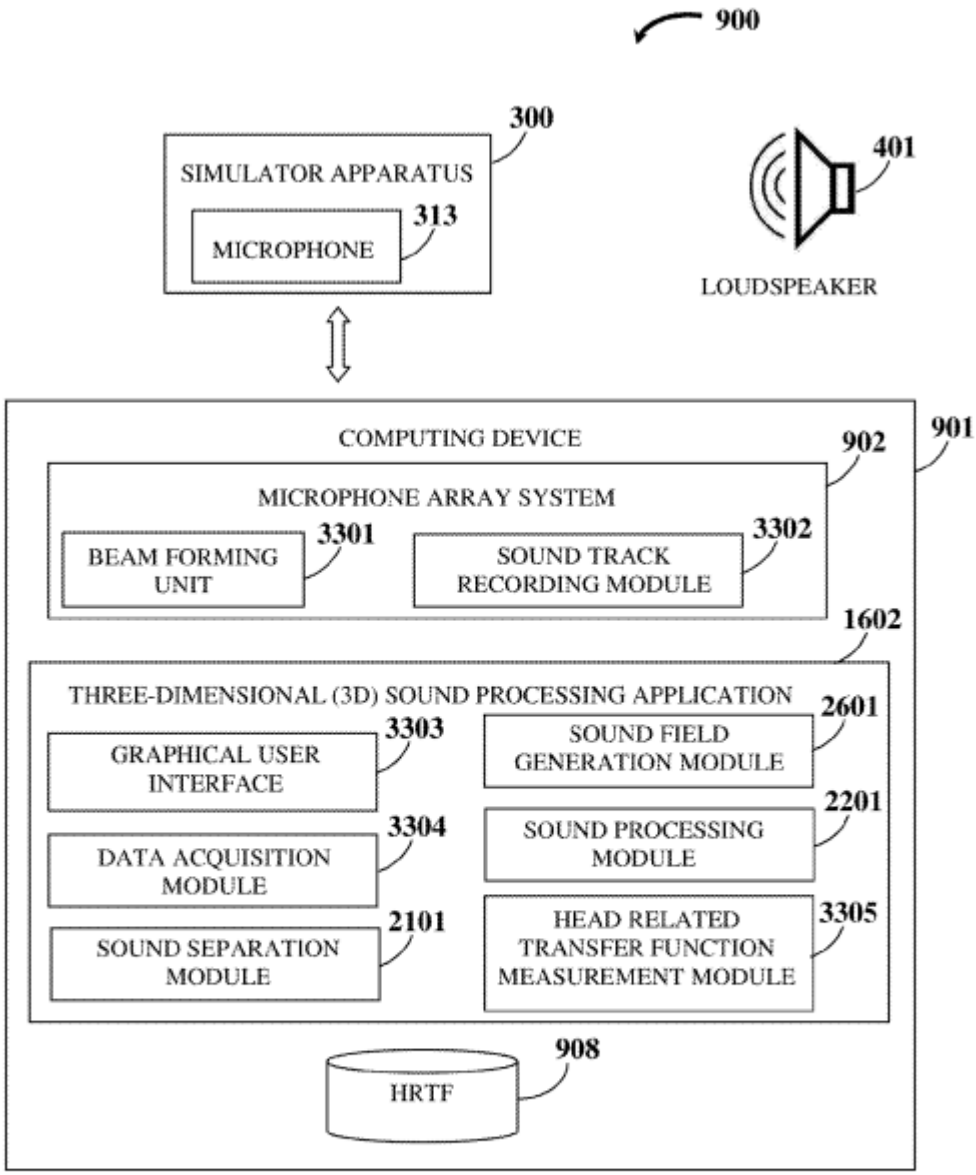


FIG. 33

101. Li further describes, in relation to Figs. 9 and 33, that "the user configures the placements and movements of any available sound sources as the inputs in order to obtain a virtual reality scene" *Id*., 22:53-67. Li's system sets the

SLP to that selected location and thus move[s] an SLP. *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").

102. Although Li does not expressly disclose *a voice of a computer program* as part of its conference call embodiment, I read Li to teach using such voice in relation to other embodiments. Li describes the sound source as "similar or different sound generating devices or sound emitting devices" including "a person's voice" and "pre-recorded multiple sound tracks." Ex. 1006, 4:58-67, 14:40-48. Li also discloses a *computer program* to "generate 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.

103. A POSA would have appreciated that implementing Li's sound source as *a voice of a computer program* would have been beneficial because it would have saved time, cost, and storage space, would have added flexibility to the vocal range of the voice, and would have improved clarity. A digital voice recording of a person requires a significant amount of storage space which increases the cost in storage needs. On the other hand, if voice recording of a person is synthesized as *a voice of a computer program*, storage demand is reduced and thus price decreased. Furthermore, *a voice of a computer* program may be digitally manipulated to

produce a sound output to a user's liking. Changes can be made to pitch, frequency, speed, or tone in order to enable a speaker to artificially reach a broader vocal range for singing or emphasizing important points. Lastly, implementing *a voice of a computer* improves clarity of the voice heard by a user. By using *a voice of a computer* rather than the original speaker's voice, noise can be filtered out to provide a clearing output and sounds from speech impediments (e.g., a lisp) may be removed so that the users can more clearly understand.

104. I read Li as disclosing that moving the sound source (*mov[ing] a SLP*) takes place *during a voice exchange between the computer program and a user/person*. When Li's user (*user/person*) "plac[es] different speaking users at different spots to mimic a real conference room environment" (*id.*, 21:66-22:5), this takes place *during* a "conference call," which involves *a voice exchange between the* processed voice of the speaker (*computer program*) *and a user/person*. Li further explains that rather than merely establishing a SLP location before a voice exchange begins, the user (*user/person*) is able to actively move the SLP which would occur after an initial placement has occurred the voice exchange has begun. *Id.*, 21:60-66.

2. 1[a] /8[a]

1[a] *providing, through* an electronic earphone being located at a head of the user and being in communication with a handheld portable electronic device (HPED),	8[a] *provide, through* a wearable electronic device on the person, *binaural sound to the person during the voice exchange that is between the computer program and the person such*

binaural sound to the user during the voice exchange between the computer program and the user such that an origin of the voice of the computer program occurs at the SLP at a first location in empty space that is away from and proximate to the user;	*that an origin of the voice of the computer program occurs at the SLP at a first location in empty space that is away from and proximate to the person*;

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

106. Li's *electronic system* includes headphones or a headset (*an electronic earphone on a person being located at a head of the user* (claim 1) and *a wearable electronic device on the person* (claim 8)) *being in communication with* a smart phone (*HPED*), as is additionally required by claim 1. Specifically, Li depicts the configurable 3D sound system 900 (*electronic system*) in Fig. 9 as providing 3D binaural sound 909 (*binaural sound*) over a "headset" (*electronic earphone being located at a head of the user*). Ex. 1006, 17:33-37; *see also* 20:44-48, 2:55-57, 3:6-10.

107. Specific to 1[a], Li's headset *communicat[es] with an HPED* because Li's configurable 3D sound system can be in a smart phone, which is an *HPED*, and the POSA would have readily understood in 2014 that it was commonplace for smart phones to connect to and communicate with headsets. Specifically, Li states the "configurable 3D sound system 900 [*electronic system*] can be used in portable devices, for example, smart phones [*HPED*], tablet computing devices, etc." for generating and playing a synthesized *binaural sound*, such as binaural sound 909.

Id., 34:52-56, 34:60-35:13. Therefore, because Li discloses the configurable 3D sound system 900 is included in a smart phone (*HPED*) and connects to a headset (*electronic earphone*), Li's headset (*electronic earphone*) is therefore *in communication with* a smart phone (*HPED*). *See also id.*, 3:6-8 ("consumers often use headphones for audio rendering on a mobile device."), 6:67-7:4, 7:33-37, 10:5-11, 14:14-15, 16:7, 17:35-37, 20:9-12, 20:44-48.

108. Specific to 8[a], I read Li's headset to render obvious a *WED* even under 8KHZ's position that a *WED* is a "portable wireless electronic device that is worn on or attached to a person." Not only were wireless headsets well-known and commonplace in 2014, but Li specifically describes various wireless connections, such as Bluetooth, infrared, Wi-Fi. *Id.*, 30:38-43.

109. I understand the remaining elements of 1[a] and 8[a] are common to both claims, and it is my opinion that Li discloses these elements. I read Li to disclose that the *SLP* is localized *at a first location* 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/person*, under 8KHZ's interpretation of *away from and proximate to*). Ex. 1006, 20:50-53 and Fig. 17. Based on the claim language describing moving an SLP form a first location to a second location, I understand that a first location refers to any location for an SLP that would occur prior to moving to the second location. 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 analysis is the same.

110. In my opinion, a POSA implementing Li would have found it obvious to position the *SLP* in *empty space*, or at a location where no physical object exists. For example, Li describes "placing different speaking users at difference spots to mimic a real conference environment." Ex. 1006, 22:1-5. 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.

111. 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:

Ex. 1037, Fig. 2

112. 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.

113. 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.

114. Furthermore, in relation to Fig. 1, Li describes measuring HRTFs "in a quiet sound treated room free of impulsive background noise." Ex. 1006, 10:39-54. These measured HRTFs "as disclosed in the detailed description of FIGS. 1-2" are "store[d]…in an HRTF database 908 exemplarily illustrated in FIG. 9." *Id.*, 19:62-67. The 3D sound processing application 1602 uses those same measured HRTFs from the database 908 to generate binaural sound 909. *Id.*, 17:27-37, 19:39-67. Because the HRTFs are measured in a "quiet sound treated room" to generate the binaural sound, I would have expected that the SLP would be placed in a similarly object-free space to replicate the environment in which it was record.

115. Because the sound sources are placed "at different spots to mimic a real conference room environment" (Ex. 1006, 21:66-22:5), the *SLP at a first location* is *away from and proximate to the user* such that the sound source would be placed at an empty chair at a distance away from the user.

116. As I explained for claims 1[pre] and 8[pre], Li renders obvious *the voice of the computer program*, whereby in Li, providing the *binaural sound* occurs *during* the conference call (*the voice exchange between the computer program and the user*) *such that* processed voice for different speakers are placed at different spots to mimic a real conference room environment for the user (*an origin of the voice of the computer program occurs at the SLP*). *See* Section IX.C.1.

3. 1[b]/8[b]

1[b] *receiving, from the user* and at the HPED, *a command to move the SLP of the voice of the computer program from the first location in empty space to a second location in empty space that is proximate to the first location and away from and proximate to the user*;	8[b] *receive, from the person, a command to move the SLP of the voice of the computer program from the first location in empty space to a second location in empty space that is proximate to the first location and away from and proximate to the person*;

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

118. Li's configurable 3D sound system 900 "acquires [*receiving*] configuration information 1601 [*command*] from a user" that 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. Those 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; *see also* 15:29-32, 15:44-47. By allowing the user to

move the sound source, Li discloses *receiving, from the user,* configuration information (*a command*) *to move the* sound source for a speaker on the conference call around the user, or up and down (*SLP of the voice*) *from the first location 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/person*.

119. Moreover, specific to element [1[b], Li's *command* is received *at the HPED* because the "user may configure the positions and movements of the sound sources on the generated configurable sound field 1700… by entering his/her preferences on the generated configurable sound field 1700 via the GUI," where the GUI is "on the user's computing device 901," which is implemented a mobile phone (*HPED*). *Id.*, 20:62-21:9, 21:47-60; *see also* 7:15-17 ("Users can control the configurations, for example, using a touch screen on their handheld devices."), 15:27-52, 15:61-16:8, 17:38-40, 18:1-21.

120. As I explained for claims 1[pre] and 8[pre], Li renders obvious the *voice of the computer program* because a POSA would be motivated to use a processed voice for the speaker. *See* Section IX.C.1.

4. 1[c]/8[c]

1[c] *moving, based on the command* and during the voice exchange between the computer program and the user, *the SLP*	8[c] *move, in response to the command, the SLP of the voice of the computer program from the first location in empty*

of the voice of the computer program from the first location in empty space to the second location in empty space that is proximate to the first location and away from and proximate to the user; and	*space to the second location in empty space that is proximate to the first location and away from and proximate to the person*; and

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

122. Li's configurable 3D sound system (*electronic system*) *moves, based on* the configuration information (*the command*) *and,* specific to 1[c], *during the* conference call (*voice exchange between the computer program and the user*)*, the SLP of the* processed voice (*voice of the computer program*) around the user, or up and down, to different locations in an object-free space (*from the first location in empty space to the second location in empty space*) *that are* at a distance from the user to mimic a real conference room environment (*proximate to the first location and away from and proximate to the user*, under 8KHZ's position). Ex. 1006, 21:57-22:5, 15:29-48; *see also* Sections IX.C.1-3 (Ground 1, 1[pre]/8[pre]-1[b]/8[b]).

5. 1[d] /8[d]

1[d] *providing, through the* electronic earphone being located at the head of the user and being in communication with the HPED, *the binaural sound to the user during the voice exchange between the computer program and the user such that the origin of the voice of the computer program occurs at the SLP at the second*	8[d] provide, through the wearable electronic device on the person, *the binaural sound to the person during the voice exchange that is between the computer program and the person such that the origin of the voice of the computer program occurs at the SLP at the second location in empty space that is proximate to the first*

location in empty space that is proximate to the first location and away from and proximate to the user.	*location and away from and proximate to the person.*

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

124. Li's configurable 3D sound system (*electronic system*) *provid[es], through* a headset (*electronic earphone being located at a head of the user* (claim 1) or *wearable electronic device on the person* (claim 8)), specific to 1[d], *being in communication with* a smart phone (*HPED*), the 3D binaural sound (*binaural sound*) *to the user during* the conference call (*the voice exchange between the computer program and the user*). *See* Section IX.C.2 (Ground 1, 1[a]/8[a]). Moreover, in Li, after the user moves the sound source around the user, or moves the sound source up and down as described in 1[b]/8[b]-1[c]/8[c], the spots for the speakers have changed to new locations in the object-free space (*such that the origin of the voice of the computer program occurs at the SLP at the second location in empty space that is away from and proximate to the user*). *See* Sections IX.C.1.-4 (Ground 1, 1[pre]/8[pre]-1[c]/8[c]).

D. Independent Claims 15:

1. 15[Pre]: *An electronic system with one or more electronic devices that move a sound localization point (SLP) of a voice of a computer program during a voice exchange between the computer program and a person, the electronic system comprising.*

125. To the extent limiting, it is my opinion that Li discloses the preamble as Li's user is a *person*. *See* Section IX.C.1. (Ground 1, 1[pre]/8[pre]).

2. 15[a]: *a wearable electronic device that the person wears and that provides the voice of the computer program to the person at the SLP at a first location in empty space that is away from and proximate to the person;*

126. In my opinion, Li renders obvious this limitation, as Li's headphones or headset is *a wearable electronic device that the* user (*person*) *wears*. *See* Sections IX.C.1-2 (Ground 1, 1[pre]/8[pre]-1[a]/8[a]).

3. 15[b]: *a memory that stores instructions; and*

127. In my opinion, Li renders obvious this limitation. Li's configurable 3D sound system 900 is an *electronic system* (*see* 1[pre]/8[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 (emphases added); *see also* 30:65-31:1, 35:24-50. Because Li discloses that the functionality sought is achieved by "employ[ing] a computer system . . . including a *memory* unit," I understand Li to disclose that the memory unit is part of *an electronic system with one or more electronic devices*, within the meaning of claim 15. Furthermore, a POSA would understand that in a computerized system, such as Li's, instructions, or software, for performing Li's functions would need to be stored. A memory (no matter the physical form) is a term

that describes the required software storage device. Li's *electronic system* therefore includes a memory unit 3402 (*memory*) *that stores instructions*.

4. 15[c]: *a processor that executes the instructions to . . .*

128. In my opinion, Li renders obvious this limitation. Li's configurable 3D sound system 900 is an *electronic system* (*see* 1[pre]/8[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: 7-19, 30:26-29 ("The memory unit 3402… stores… *instructions* for *execution* by the *processor* 3401.") (emphases added); *see also* 31:18-35, 32:32-33:31, 34:36-51, 35:17-50, 36:37-41. Consistent with my discussion of claim 15[b] above, because Li discloses that the functionality sought is achieved by "employ[ing] a computer system" including a *processor* 3401, I understand Li to disclose that *processor* 3401 is part of *an electronic system with one or more electronic devices*, within the meaning of claim 15. Furthermore, a POSA would recognize that to perform the functionality disclosed in Li would require a processor to read an execute instructions. Therefore, Li's *electronic system* therefore includes a processor 3401 (*processor*) *that executes the instructions*.

5. 15[c1]: *receive a command from the person to move the SLP of the voice of the computer program from the first location in empty space to a second location in empty space that is proximate to the first location and away from and proximate to the person,*

129. In my opinion, Li renders obvious this limitation for at least the reasons I discussed in Section IX.C.3 (Ground 1, 1[b]/8[b]).

6. 15[c2]: *move the SLP of the voice of the computer program in response to the command from the first location in empty space to the second location in empty space, and*

130. In my opinion, Li renders obvious this limitation for at least the reasons I discussed in Section IX.C.4 (Ground 1, 1[c]/8[c]).

7. 15[c3]: *provide binaural sound to the person through the wearable electronic device during the voice exchange between the computer program and the person such that an origin of the voice of the computer program occurs at the SLP at the second location in empty space that is proximate to the first location and away from and proximate to the person.*

131. In my opinion, Li renders obvious this limitation for at least the reasons I discussed in Section IX.C.5 (Ground 1, 1[d]/8[d]).

E. Dependent Claim 17: *processor further executes the instructions to change an interaural time difference (ITD) and an interaural level difference (ILD) of the binaural sound in order to move the SLP of the voice of the computer program from the first location in empty space to the second location in empty space*

132. In my opinion, Li renders obvious this limitation. As I explained for claim 15[c2] and 1[pre]/8[pre]-1[c]/8[c], a POSA would have found it obvious for Li's *processor* 3401 to *execute[] the instructions to change* HRTFs *of the binaural sound in order to move the SLP of the voice of the computer program from the first location in empty space to the second location in empty space*. *See* Sections IX.C.1-4, IX.D.6 (Ground 1, 1[pre]/8[pre]-1[c]/8[c], 15[c2]). Li discloses sound

localization cues including *an ITD and an ILD* that form the "combination of [] factors [that are] modeled by" HRTFs, such that both cues are included in the HRTFs that are used to *move the SLP*. Ex. 1006, 1:65-2:21.

133. Sound originating from a location reaches the two listener's ears at different times because the ears are at different locations on either side of the head. That time difference depends on the angle the sound is originating from as it travels from the source to the listener. The listener's brain uses this interaural time difference (ITD to help determine the direction of the sound.

134. Similarly, the intensity of the sound will differ from one ear to the other. The listener's brain uses this intensity difference (interaural level difference, ILD) cue to help determinate the angle of the sound source.

135. The sound localization ability of the ear using ITD and ILD cues was first published by Lord Rayleigh in 1907 and is known as the "duplex theory" [Strutt, J. W. (1907). ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232.]

136. Sounds that originate directly in front or directly behind the listener, however, will have identical ITD and ILD characteristics. The same is true for sounds in symmetric front hemisphere and rear hemisphere locations (e.g., at 30° to the front right or 150° to the rear right of the listener). The listener is able to distinguish these directions because there are additional cues that the ear uses. These cues have to do with differences in the spectrum of the sound as it is altered (uniquely

for every angle) by the outer ear (pinna) and upper torso of the listener. The ITD, ILD, and spectral cues are all contained in the frequency domain HRTF (or the time domain HRIR). Ex 1006, 1:65-2:21.

137. Accordingly, in order to produce an SLP that will provide a user with the requisite cues to perceive the sound at a selected location, Li's system uses an HRTF (or HRIR) to reproduce ILD and ITD cues. When the user changes the SLP location, the ILD and ITD must change, as well, to provide the new cues to the listener. Therefore, in moving the SLP, Li's system must change an interaural time difference (ITD) and an interaural level difference (ILD) of the binaural sound.

X. GROUND 2: LI IN VIEW OF BALLAS RENDERS OBVIOUS CLAIMS 4, 6, 9-10, 14, 18

A. Motivation to Combine Li with Ballas

138. In my opinion, it would have been obvious to a POSA to modify Li 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 recognizes that to determine a sound source, humans rely on various cues including head motion. Ex.1006, 1:38-42. In my opinion, a POSA would have been motivated to combine the teachings of Li and Ballas because a POSA would have recognized that both references have a shared goal of providing a realistic experience for the user and implementing Ballas's head tracking would relieve the user of having to input manual commands.

Users do not remain perfectly still while using a headset. Therefore, if the HRTF remains unchanged, the location of an SLP will move with the movement of the user's head. An SLP that moves with the user's head does not establish a realistic experience for the user. To prevent movement of the SLP, a POSA would have tracked the user's head orientation and movements and provide the orientation and movement to Li's sound field so that Li's system could account for the user's orientation and movements, using conventional head tracking and spatialization techniques taught by Ballas.

139. In my opinion, 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. Ballas shares this 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 contemplates use of its method for potentially lengthy stretches of time, such as for conferences, concerts, or radio dramas. Ex. 1006, 21:57-22:-5. For such applications, head tracking is key to make the experience as realistic as possible for the user. Specifically, 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 with Ballas's teachings of head tracking and adjusting the sound field 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.

140. The modification would have been a natural extension of Li's use of HRTFs to spatialize sound and generate a sound field. Like Li, Ballas discloses an "[HRTF] filter to produce spatialized audio." Ex. 1008, Abstract; *see also id.*, ¶¶16-17, 25. Specifically, Ballas 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. Thus, the use of 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 advanced the common goals of both Li and Ballas to provide a realistic environment in a teleconferencing application. In order for Ballas's HRTF filters to "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," Ballas's system must continually perform and adjust HRTFs because the relative location of the SLP compared to the user's position changes.

B. Reasonable Expectation of Success

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

142. Consistent with Li's system, Ballas's speakers 110 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 processed using HRTF filters 106 that are used to produce the spatialized audio. *Id.* Accordingly, in my opinion, a POSA would have recognized that Li and Ballas, because each describes an apparatus designed to be fixed to a user's head and produce binaural sound, are structurally and functionally similar.

143. In my opinion, a POSA implementing the Li system would have recognized that the headset described in Li could be attached to Ballas's frame. A headset, such as that disclosed in Li, consists of small speakers that could easily have been mounted to Ballas's frame using small screws or an epoxy or integrated using a molding technique. 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. 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). The computer systems of Li and Ballas could easily be integrated in a single processor and memory system or through two separate systems that communicate electronically.

C. Dependent Claim 4:

1. 4[a]: *tracking head orientation of user as user moves with respect to SLP at the second location*

144. In my opinion, Li in view of Ballas discloses this limitation. 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 discloses "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

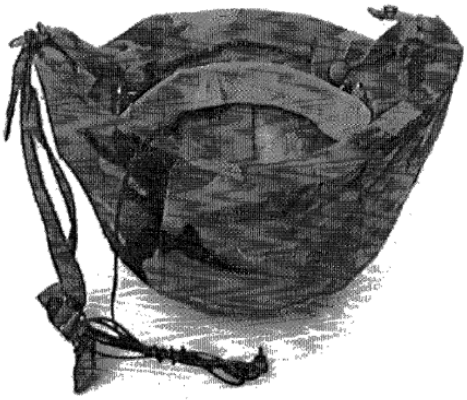


Figure 4. Helmet mount prototype

Mounted to the 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" (*tracking a head orientation of the user*). Ex. 1008, ¶25.

145. Ballas discloses "a pair of head related transfer functions (HRTFs)" that are "for a particular location" (*the SLP at the second location*). *Id.*, ¶16. Specifically, Ballas's 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. 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 *as the user moves with respect to the SLP at the second location*. The SLP remains in a fixed position while the user moves. Because the SLP and user's head do not move together, the Ballas's system detects movement of the user *with respect* to the SLP.

146. In my opinion, a POSA would have been motivated to implement Li's headphones as part of Ballas's wearable frame having a head tracking system, as 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 X.A (Motivation to Combine Li with Ballas). When a listening participant in an in-person conversation moves, there is no effect on the position of a speaking participant. As the listening participant moves, the listening participant receives ILD and ITD cues that that varying with the distance relative to the speaker and are interpreted by listening participant's brain to determine the fixed position of the speaker. To provide a realistic virtual experience that mimics this in-person interaction, the varying of ILD and ITD cues must be replicated. In order to replicate the sound cues, Li's system would need to receive information regarding the user's movements. Ballas's wearable frame with head tracking system provides that functionality and thus enables a realistic virtual experience.

2. 4[b]: *adjusting voice of computer program that user hears through electronic earphone so voice continues to appear to originate from SLP at the second location even as the user moves with respect to the second location.*

147. In my opinion, Li in view of Ballas discloses this limitation.

148. As I explained for claim 4[a], Ballas discloses applying HRTF filters (*adjusting*) to the *sound* in order to maintain the sound's virtual location (*so the*

sound… continues to appear to originate from the SLP at the second location), which allows the user to move without degradation of the virtual location (*even as the user moves with respect to the second location*). *See* Section X.C.1.

D. Dependent Claim 6: *changing an amplitude, an interaural time difference, and an interaural level difference of the voice of the computer program as the user moves with respect to the SLP at the second location so a sound of the voice of the computer program continues to appear to originate from the SLP at the second location as the user moves*

149. In my opinion, Li discloses this limitation. As I already explained above for 4[a]-4[b], Li-Ballas discloses adjusting (*changing*)… *the voice of the computer program as the user moves with respect to the SLP at the second location so a sound of the voice of the computer program continues to appear to originate from the SLP at the second location as the user moves. See* Section X.C.

150. Li 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 (emphases added). Li defines the ITD as "the difference of arrival times of transmitted sound between the two ears" and the ILD as "the difference in level and/or intensity of the transmitted sound received between the two ears." *Id.,* 1:65-2:4; *see also id.*, 12:54--57. Li also discloses *an amplitude* by describing "volume" as a "configurable parameter associated with the sound sources." *Id.*, 5:18-22.

151. Li also discloses *changing* each of the *ILD* and *ITD for the voice* as part of its bank of HRTF filters. Li explains that 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 (HRTFs). 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. Li applies one HRTF (from the bank of HRTFs) to produce a spatialized 3D sound for a specific position (*first location*), and applies a different HRTF to position the 3D sound at a new position (*SLP* at the *second location*). *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. Because 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 for the voice*.

152. Li also discloses *changing* the *amplitude for the voice* as part of its bank of HRTF filters. Specifically, Li explains that "distance" and "volume" (*amplitude*) are each a "configurable parameter[] associated with the sound sources" and are therefore adjustable as *the user moves with respect to the SLP at the second location*, as contemplated in Ballas with its head tracking system. Ex. 1006, 5:18-22; *see also id.*, 15:36-40, 28:29-35, claims 8, 16, 19, 25, 35. In addition, in a 3D sound field, a person's brain perceives sound signals having a low volume to be further away and louder sounds to be closer. An amplitude of a sound decays over distance. Thus,

closer sounds are louder and farther sounds softer. By having a bank of HRTFs for different "positions" (not just directions) Li applies different HRTFs for different locations which would include distances. Applying different HRTFs at different distances which would involve *changing the amplitude* (i.e. volume) *for the voice* as the positions shift to change distance.

E. Dependent Claim 9: *maintain the origin of the voice of the computer program at the SLP at the second location in empty space but alter an amplitude of the voice of the computer program that the person hears in response to the person moving closer to or away from the SLP at the second location*

153. In my opinion, Li in view of Ballas discloses this limitation. See Section X.D (Ground 2, Claim 6).

F. Dependent Claim 10

1. **10[a]**: *monitor movements of head orientation of person in response to person moving with respect to SLP at second location in empty space; and*

154. In my opinion, Li in view of Ballas discloses this limitation. *See* Section X.C.1. (Ground 2, 4[a]).

2. **10[b]**: *adjust an interaural time difference (ITD) and an amplitude of the voice of the computer program in response to the movements of the head orientation of the person as the person moves*

155. In my opinion, Li in view of Ballas discloses this limitation. *See* Section IX.E (Ground 1, cl. 17) and Sections X.C.2 and X.D (Ground 2, 4[b] and 6).

G. Dependent Claim 14: *adjust interaural level difference, interaural time difference, and amplitude for voice person hears so SLP of voice remains fixed at second location in empty space while head orientation of person moves*

156. In my opinion, Li in view of Ballas discloses this limitation.

157. As I explained for claims 4[a] and 4[b], Li in view of Ballas discloses *adjust[ing]* a *voice that the person hears so the SLP of the voice of the computer program remains fixed at the second location in empty space while a head orientation of the person moves with respect to the second location. See* Section XI.C. As I explained for claim 6, Li discloses *adjusting an ILD, and ITD, and an amplitude for the voice*. *See* Section X.D.

H. Dependent Claim 18: *processor further executes instructions to change interaural time difference (ITD) and amplitude of binaural sound to maintain SLP fixed at second location in empty space during voice exchange while person physically moves away from second location in empty space and changes head orientation with respect to second location in empty space*

158. In my opinion, Li in view of Ballas discloses this limitation.

159. As I explained for claim element 6[a], Li in view of Ballas discloses *chang[ing] an ITD and an amplitude of the binaural sound in order to maintain the SLP fixed at the second location in empty space… while the person… changes a head orientation with respect to the second location in empty space.* As I further explained for claim element 15[c], Li's *processor* 3401 *executes instructions* to

perform this processing, and as explained in 1[pre]/8[pre], this occurs *during the voice exchange*.

160. Ballas explains "[t]he HRTF filter maintains virtual location of the audio signals, thus allowing the listener to change locations and head orientation without degradation of the audio signal." Ex. 1008, ¶17; *see also id.*, Abstract, ¶18, claim 1. Ballas's head tracking system 104 "measures the location and orientation of a user's head." *Id.*, ¶25; *see also id.*, Abstract, ¶¶17-20, 26, claims 1, 7, 13. As the POSA would have understood, Ballas's head tracking system would have tracked *the person physically mov[ing] away from the second location in empty space.* Ballas's head tracking system would track any movement of the head including movements away from a second location. To provide as realistic and as immersive of an experience as possible for the user, the combined system of Li and Ballas would have processed the sound *in order to maintain the SLP fixed at the second location* regardless of the user's movements or head orientation. When a listening participant in an in-person conversation moves, there is no effect on the position of a speaking participant. As the listening participant moves, the listening participant receives ILD and ITD cues that that varying with the distance relative to the speaker and are interpreted by listening participant's brain to determine the fixed position of the speaker. To provide a realistic virtual experience that mimics this in-person interaction, the varying of ILD and ITD cues must be replicated. In order to replicate

the sound cues, Li's system would need to receive information regarding the user's movements. Ballas's wearable frame with head tracking system provides that functionality and thus enables a realistic virtual experience.

XI. GROUND 3: LI IN VIEW OF BALLAS AND SANDGREN RENDERS OBVIOUS CLAIM 5

A. Motivation to Combine

161. In my opinion, and as I explained above for Ground 2, a POSA would have been motivated to combine the teachings of Li and Ballas. *See* Section X.A (Ground 2, Motivation to Combine Li with Ballas). As I previously explained, I understand Li's primary goal to be providing 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. In addition, Li recognizes that the "volume" is a "configurable parameter[] associated with the sound sources." Ex. 1006, 5:18-22. As was well-known in the field, a sound signal having a low volume is perceived to be further away than a louder sound, and the POSA would have readily known to apply this principle to Li's 3D sound field, a sound signal.

162. It is my opinion that a POSA would have been motivated to look to Sandgren to precisely determine how to vary Li's sound volume with distance. Doing this would have helped to achieve Li's stated goal of providing 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.

163. Sandgren specifically shares this goal and additionally states that the sound intensity follows an inverse-square law with distance. Ex. 1017, ¶¶1, 77, Abstract. Thus, in my opinion, applying Sandgren's teachings to Li to adjust a sound amplitude according to an inverse-square law with distance would have provided a more realistic and immersive experience for the user, and would have been a natural extension of Li's disclosure of a configurable volume.

B. Reasonable Expectation of Success

164. In my opinion, the POSA would have had a reasonable expectation of success in modifying Li-Ballas in accordance with the teachings of Sandgren. Both Li and Sandgren use HRTFs to generate spatialized binaural sound and provide such sound over headphones. Moreover, Sandgren's system is consistent with Li's system in that Sandgren also 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, a POSA would have recognized that Li and Sandgren are structurally and functionally similar.

165. Moreover, the POSA implementing the Li-Ballas system would have appreciated that the HRTFs used to process and spatialize the sound would have been accompanied with Sandgren's attenuation filters and reverberation generators. Doing so, as Sandgren teaches, would have had the benefit of artificially simulating the presence of virtual objects. This 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 5:** *adjust amplitude of voice of computer program to be inversely proportional to square of distance between user and SLP at second location in empty space as user moves with respect to SLP at second location in empty space*

166. In my opinion, Li in view of Ballas and Sandgren discloses this limitation.

167. As I explained for claim 6, Li in view of Ballas discloses *adjust[ing] an amplitude of the voice of the computer program… as the user moves with respect to the SLP at the second location in empty space. See* Section X.D.

168. Neither Li nor Ballas expressly recites *adjust[ing]* the *amplitude… to be inversely proportional to a square of a distance between the user and the SLP*, but this would have been obvious in view of Sandgren. Sandgren discloses "[s]ound intensity generally follows an inverse-square law, meaning that for each doubling of distance, the intensity drops by 6 dB." Ex. 1017, ¶77. Indeed, the reduction of sound level (amplitude) as a function of distance from the source is a well-established and would be common knowledge for a POSA. Therefore, in order to provide as realistic of a sound field as possible, the POSA would have found it obvious to *adjust an amplitude… to be inversely proportional to a square of a distance between the user and the SLP*. If the sound intensity were not adjusted to be inversely proportional to a square of a distance between the user and the SLP, a user's brain would not be able to properly determine the intended location of the SLP.

XII. GROUND 4: LI IN VIEW OF SANDGREN RENDERS OBVIOUS CLAIM 19

A. Motivation to Combine

169. In my opinion, it would have been obvious to a POSA to modify Li 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. Li recognizes that to determine a source of a sound, humans rely on various types of cues including reverberation. Ex. 1006, 1:38-42. It is my opinion that a POSA

would have been motivated to combine the teachings of Li and Sandgren because a POSA would have recognized that both references have a shared goal of providing a realistic and immersive experience for the user. To make this goal a reality, 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.

170. Moreover, it is my opinion that the POSA would have understood to simulate sound-reflecting objects that do not actually exist but are perceived to exist in the virtual environment, using conventional sound enhancement techniques taught by Sandgren. Thus, although Li does not describe changing attenuation and reverberation of the binaural sound to add artificial audial cues of virtual objects, it is my opinion that a POSA would have been motivated to perform such an implementation, as doing so has the benefits described below.

171. 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. As Sandgren appreciates, "[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 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 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.

172. Moreover, in my opinion, the modification would have been a natural extension of Li's use of HRTFs to spatialize sound and generate a sound field. Like Li, Sandgren 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. 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. In my opinion, a POSA would have recognized that the teachings of Sandgren would have enhanced Li's use of HRTFs for 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 Li and Sandgren to provide a realistic and immersive experience for the user.

B. Reasonable Expectation of Success

173. In my opinion, a POSA would have had a reasonable expectation of success in modifying Li in accordance with the teachings of Sandgren. Li and Sandgren involve the use of HRTFs. HRTFs can be easily modified to incorporate additional elements.

C. Dependent Claim 19:

1. 19[a]: *The electronic 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 person,*

174. In my opinion, Li in view of Sandgren discloses this limitation.

175. While Li recognizes that *reverberation* is a "source[] of localization cues" (Ex. 1006, 1:39-42), Li does not explicitly disclose that its *processor executes*

the instructions to perform this limitation, but, in my opinion, this would have been obvious in view of Sandgren.

176. Sandgren relates to simulating sound-reflecting objects (*virtual objects*) to enhance the "[e]xternalization of the simulated sound source." Ex. 1017, Abstract. According to Sandgren, "[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.

177. In my opinion, Sandgren discloses *chang[ing] reverberation*. "Reflected 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 (emphases added); *see also* ¶¶23-25, 48-66, 68-74, 100. As Sandgren describes, "[w]hen 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."[1] *Id.*, ¶69, Figs. 5A-C; *see also id.*, 74 (describing updating early reflections when simulating a dynamic scene with moving objects).

[1] Sandgren refers to element 502 as an "early-reflection generator" as well as an "early-reverberation generator" and uses these names interchangeably. Therefore, I refer to Sandgren's element 502 as an "early-reverberation generator."

By updating parameters for a reverberation generator to simulate moving objects, Sandgren discloses *chang[ing]… reverberation of the binaural sound in order to add artificial audial cues of virtual objects near the person*."

178. Sandgren also discloses *chang[ing] attenuation*. Sandgren's Figs. 5A-C depict audio simulators "having HRTF processors and an early-reflection generator," and each of these figures additionally depict attenuation filters 504, 506, 520, and 522. Ex. 1017, ¶¶31-33. These attenuation filters represent a gain or amplification applied to the direct sound (*e.g.*, 504, 506) or early reflections (*e.g.*, 520, 522). *Id.*, ¶¶49, 62, Figs. 5A-C. As the POSA would have appreciated, the attenuation applied to both direct sound and early reflections (including $a_k(n)$ representing the attenuation filter for the k-th reflection) will *change* as Sandgren contemplates its *virtual objects* moving and changing location. Ex. 1017, ¶¶74-75, 78-80, 89, Fig. 1. Therefore, Sandgren discloses *change… attenuation of the binaural sound in order to add artificial audial cues of virtual objects near the person*."

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

179. In my opinion, Li in view of Sandgren discloses this limitation.

180. As I explained for claim 19[a], Sandgren describes simulating sound-reflecting objects (*virtual objects*) to enhance the "[e]xternalization of the simulated

sound source." Ex. 1017, Abstract. These objects 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*. *Id*., Abstract, ¶¶6, 23-25, 46-48, claims 1, 10, 19.

XIII. GROUND 5: LI AND WOO RENDERS OBVIOUS CLAIM 11

A. Motivation to Combine

181. In my opinion, it would have been obvious to a POSA to modify Li to adjust the voice to include artificial cues based on actual physical objects in the user's room, as taught in Woo. As the POSA would have recognized, an implementation that delivers a sound perceived by the user to actually originate from the SLP in the user's room would have enhanced Li's conferencing system and achieved Li's goal of providing as realistic and immersive an experience as possible. Woo discloses photographing an actual object in the user's room and adjusts the sound according to the object's position in the room to provide optimized audio to the listener. By photographing objects in the room, Woo is able to establish adjusting sound based on the object's location and a "sweet spot" which is the listener location relative to the sound source that the designer of the sound processing apparatus considers ideal for the best listening experience. EX.1016, [0006]-[0008]. In my opinion, the POSA would have known to combine this conventional teaching in Woo

with Li's system so that the sound is perceived by the user to actually originate from the SLP in the user's room, and accounts for physical objects in the room.

182. Furthermore, in my opinion, the combination would have been a natural extension of Li's remote conferencing system. For example, using Woo's detection of actual physical objects in the room and corresponding adjustment of the sound would have helped to achieve Li's primary goal of providing 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. Moreover, in my opinion, the POSA would have understood that adjusting a sound based on photographs of actual objects in the room would have enhanced Li's delivery of sound and provided a more realistic and immersive experience for the user.

B. Reasonable Expectation of Success

183. It is my opinion that a POSA would have had a reasonable expectation of success in modifying Li in accordance with the teachings of Woo. Consistent with Li's system, Woo tracks a user's position, adjusts certain properties of a sound and optimizes the delivery of the sound for the user. Accordingly, a POSA would have recognized that Li and Woo are structurally and functionally similar.

184. In my opinion, a POSA implementing the Li system would have recognized that the processing of the sound would have been accompanied with

Woo's adjustment of the sound based on actual physical objects in the user's room, in order to artificially simulate their presence. It is further my opinion that such an implementation would have merely involved a combination of known prior art elements (camera and sound processor) according to known methods and techniques (simulate spatialized sounds) to yield predictable results (producing sound so that the user perceives the sound as if it originates in the user's actual room), and involves use of a known technique (generating realistic spatializations for sounds) 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 11:

1. 11[a]: *determine physical objects located in a room with the person*

185. In my opinion, Li in view of Woo discloses this limitation.

186. While Li discloses providing the user with an immersive experience that "mimic[s] a real conference room environment" (Ex. 1006, 22:1-5), Li does not expressly disclose this limitation, but this would have been obvious in view of Woo. Woo relates to "adjust[ing] properties of sound [] provid[ing] a sound effect." Ex. 1016, ¶3, Abstract. Specifically, Woo discloses "photograph[ing] the object 200 and generat[ing] the plurality of images," and then "recogniz[ing] the position of the object 200" from the images. *Id.*, ¶¶53-54, Fig. 5; *see also id.*, ¶¶40, 42-44, 47-49,

52, 55, Abstract. As the POSA would have appreciated, recognizing an object's position would have meant that Woo's system *determine[s] physical objects.* By photographing and then detecting positions of object 200, the POSA would have understood that Woo *determine[s] physical objects located in a room with the person*.

2. 11[b]: *adjust, based on physical objects located in room, the voice of computer program that person hears to emulate sound that person would hear if voice of computer program actually originated from SLP at second location.*

187. In my opinion, Li in view of Woo discloses this limitation.

188. As Woo explains, "[t]he sound processing apparatus 100 adjusts the property of the sound corresponding to the position of the object 200" that is determined from the photographic images. Ex. 1016, ¶¶54-55, Fig. 5; *see also id.*, ¶¶3, 40, 45, 48-49, claims 1, 7, 13, 19. By describing that the sound is adjusted based on the object's position in the images, Woo discloses *adjust* the sound field *based on the* object 200 (*physical objects located in the room*). Woo modifies a source of the sound such that it is located at a distance that is the "sweet spot," a location for ideal sound perception. Ex. 1016, [0008]-[0010]. Examples of properties that Woo adjusts are "frequency, waveform, delay time, volume according to a frequency band and left/right balance" (Ex. 1016, ¶45), which the POSA would have understood to be properties that would need to be accounted for in the sound field if there are

physical objects that reflect and/or absorb sound. In my opinion, in order to meet Li's goal of providing as realistic and immersive an experience as possible, it would have been obvious to, upon determination of physical objects in the same room as the person, to provide a realistic sound field that accounts for the presence and movement of those physical objects, and provide *the voice of the computer program that the person hears to emulate a sound that the person would hear if the voice of the computer program actually originated from the SLP at the second location*. *See* Section IX.C.5. (Ground 1, 1[d]/8[d]).

XIV. GROUND 6: BEDINGFIELD RENDERS OBVIOUS CLAIMS 1, 8, 15, AND 17

A. Motivation to Modify

189. In my opinion, it would have been obvious to a POSA to modify Bedingfield to (1) provide as binaural sound *a voice of a computer program during a voice exchange between the computer program and a user (or person)*, and (2) implement the multi-channel processing device as an *HPED*. A POSA would have been motivated to implement this modification for the following reasons.

1. *Voice of a computer Program.*

190. For the same reasons I explained for Li in Ground 1, the POSA would have recognized certain advantages to provide a *voice of a computer program* as Bedingfield's audio signal, as a computerized text-to-speech synthesis would have

(1) allowed a speaking-impaired person to participate in Bedingfield's conference call system, (2) enabled Bedingfield's system to improve clarity while saving time, cost, and storage space, and (3) adding flexibility. *See* Section IX.A (Ground 1, Motivation to Modify).

191. Moreover, the modification would have been a natural extension of Bedingfield's disclosure of "synthetic audio placement" and rendering of a "sound field" of different audio sources. Ex. 1005, Abstract, 1:15-17, 1:50-2:16, 4:18-19. Bedingfield already contemplates "generating a synthetic audio signal" (Ex. 1005, 1:63-64). In my opinion, a POSA would have found it obvious that such a synthesized audio signal would have been *a voice of a computer program*.

2. *HPED.*

192. In my opinion, a POSA would have recognized the advantages of Bedingfield's listener 102 (*user* or *person*) having an *HPED*. For example, an *HPED* would have been easily transported without requiring any fixed location or physical wired connection. Being *handheld* and *portable* would have means that the listener 102 would have been able to remain connected to Bedingfield's conference call system while moving around and changing locations.

193. Further in my opinion, the modification would have been a natural extension of Bedingfield's disclosure of various *HPEDs* such as "telephones, wireless phones, [or] mobile phones" as connecting participants to the conference

call system. *See* Ex. 1005, 4:23-30. Bedingfield further contemplates using "mobile, handheld, or laptop computer; a mobile telephone; a [PDA]" and similar devices as part of a system to provide synthetic audio placement (*see id.*, 11:41-51), which the POSA would have understood are *HPEDs*, and any of which would have been a device for the listener 102.

B. Reasonable Expectation of Success

194. In my opinion, a POSA would have had a reasonable expectation of success in modifying Bedingfield with (1) *a voice of a computer program* and (2) implement the multi-channel processing device as an *HPED*.

195. The proposed modification involves (1) a simple addition of another sound source and (2) using a handheld portable device, both of which have many advantages. Therefore, in my opinion, the combination is nothing more than applying known prior art elements according to known methods and known techniques to yield predictable results, and involves use of known techniques to improve a similar device in the same way.

1. *Voice of a computer Program.*

196. Bedingfield provides audio signals to both ears that the listener perceives to originate from an SLP. Bedingfield discloses generating a "sound field providing synthetic audio placement of each of the plurality of remote audio signals" and "project[ing] the plurality of remote audio within the sound field to a listener."

Ex. 1005, Abstract. Bedingfield contemplates using its sound field in a conference call setting. *Id.*, Abstract, 1:21-46, 3:25-34, 5:4-15, Figs. 1-3). While Bedingfield does not expressly disclose *a voice of a computer program*, a POSA would have had a reasonable expectation of success using such a voice in Bedingfield, to save time and cost, add flexibility, and improve clarity. Moreover, modifying Bedingfield would have been as simple as using a computer-synthesized voice (instead of a recorded voice) as an audio signal, and would have been obvious especially in view of Bedingfield's teaching of "generating a synthetic audio signal." Ex. 1005, 1:63-64.

2. *HPED.*

197. Bedingfield already describes providing sounds to the listener over headphones, which are an *HPED*. In my opinion, the POSA would have had a reasonable expectation of success implementing the multi-channel processing device 118 to be *handheld* and *portable* because such processors were well-known and widely used well before the relevant time period. Bedingfield already contemplates various *HPEDs*, which the POSA would have understood could have easily incorporated the multi-channel processing device 118.

C. Independent Claims 1 and 8:

1. 1[Pre]/8[pre]

1[pre] A method executed by *one or more electronic devices in an electronic system that move a sound localization point (SLP) of a voice of a computer program during a voice exchange between the computer program and a user, the method comprising*:	8[pre] A non-transitory computer readable storage medium storing instructions that cause *one or more electronic devices in an electronic system* to execute a method *that moves a sound localization point (SLP) of a voice of a computer program during a voice exchange that is between the computer program and a person, the method comprising*:

198. To the extent limiting, it is my opinion that Bedingfield renders obvious the preambles.

199. Bedingfield discloses *one or more electronic devices in an electronic system* by describing a communication system 116 and a multi-channel processing device 118 that *execute a method*. Ex. 1005, 4:31-51, Fig. 1, reproduced below).

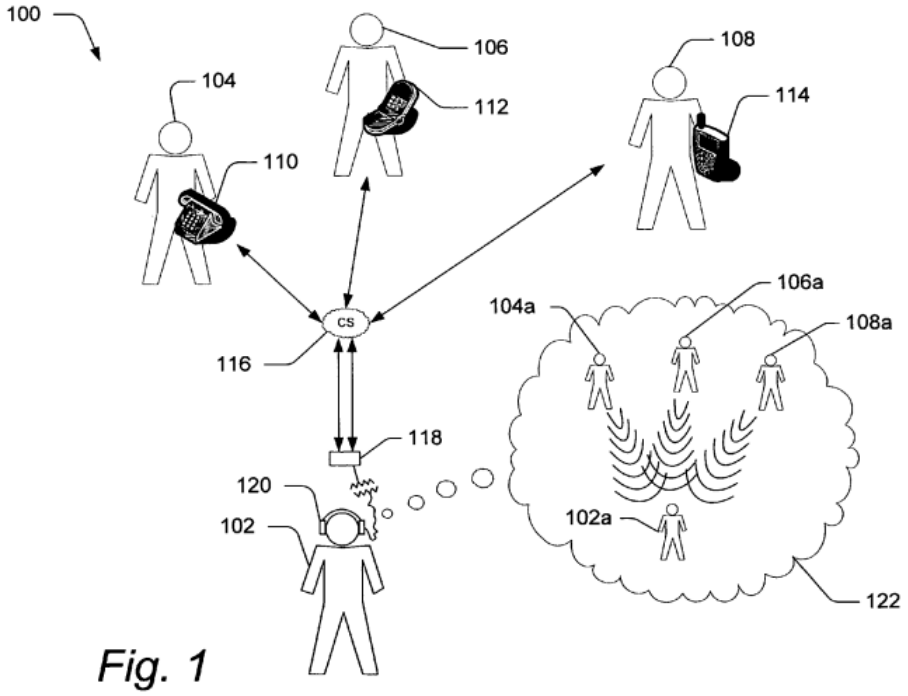


Fig. 1

As Bedingfield explains in relation to Fig. 1, the combination of the communication system 116, a multi-channel processing device 118, and multi-channel speaker system 120 (*electronic system*) has *one or more electronic devices* (which the POSA would have known to include a processor and memory), including the individual components 116, 118, and 120. *Id.*, 4:31-51.

200. Specific to element 8[pre], it is my opinion that Bedingfield discloses a *non-transitory computer readable storage medium storing instructions*. Bedingfield states "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 (emphases added); *see also* 11:23-36 (describing "computer executable *instructions*" as being stored in "*memory* storage devices" (emphases added)), 12:5-44 (describing a processor that reads "executable *instructions* from… the storage medium" (emphasis added)), Fig. 12. Bedingfield therefore contemplates that the communication system 116, the multi-channel processing device 118, and/or the multi-channel speaker system 120 (together, *the electronic system*) implement functions using software and would include *a memory that stores instructions*.

201. In Bedingfield, the communication system 116 and the multi-channel processing device 118 "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." *Id.*, 4:66-5:3. In other words, Bedingfield's *electronic system* simulates the sound field 122 so that the listener's perception is as if the listener and participants "were in the same location and speaking to each other directly (without the communication system 116)." *Id.*, 4:52-61.

202. Bedingfield further describes *moving an SLP of a voice… during a voice exchange*. In Fig. 3 (reproduced below), Bedingfield depicts "a block diagram of a method 300 of how synthetic audio placement may be organized, including selections a listener may make," including "on the multi-channel processing device 118" of Fig. 1. *Id.*, 5:51-55.

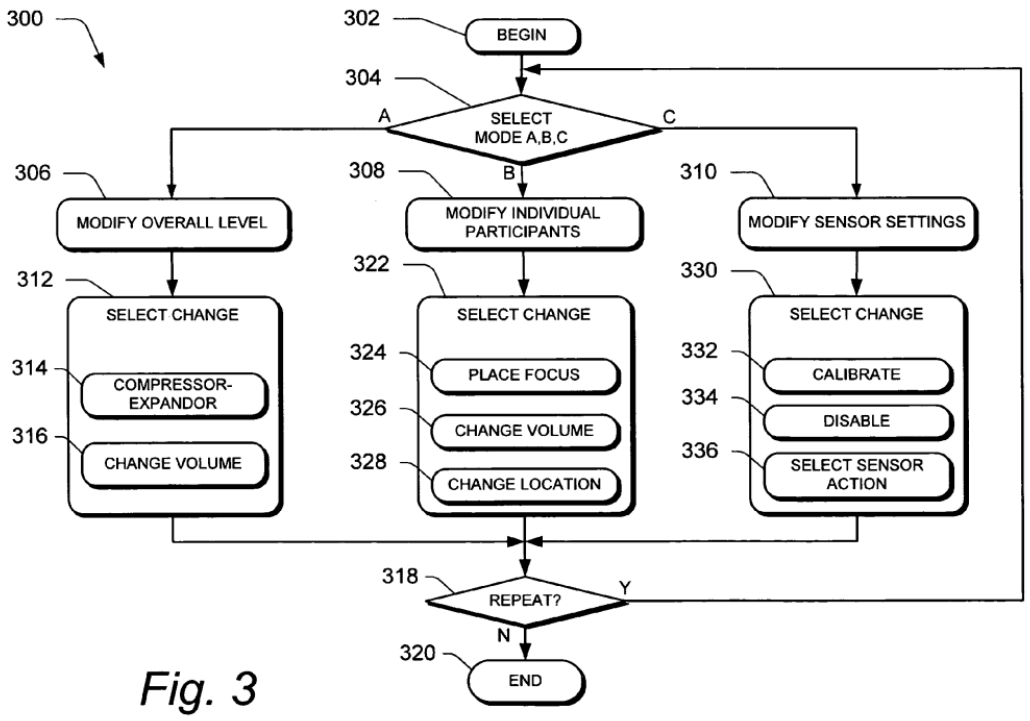


Fig. 3

"[A]t the block 328, the listener selects a participant and then selects a new location for the participant within the sound field" such as sound field 122. *Id.*, 6:43-45; *see*

also 6:29-32, 9:60-64, 14:7-10, 16:33-35. By using the multi-channel processing device 118 (part of the *electronic system*) to move a participant's location within the sound field, Bedingfield's *one or more electronic devices in an electronic system move an SLP of a voice*.

203. In Bedingfield, this takes place during a *voice exchange between the computer program and a user/person* because Bedingfield contemplates the environment of Fig. 1 is used to connect participants to one another in a conference call (*see, e.g.,*, *id.*, 5:4-10, Fig. 2) and that the block diagram of Fig. 3 is used in the context "when multiple participants are speaking during a conference" (*id.*, 6:36-37), when a *voice exchange* amongst participants occurs.

204. Bedingfield renders obvious the sound signal is *a voice of a computer program*. Bedingfield contemplates the participants communicate over its system using their *voices.* Ex. 1005, 4:25-30 (describing the communication devices as "voice" devices), 4:36-61 (describing the participants as "speaking"); *see also* 6:9-10, 9:60-64. Bedingfield also discloses "*computer program* products," including "software modules" that "provide synthetic audio placement." *Id.*, 2:17-24, 12:9-15 (emphasis added). For the same reasons as I explained above for Li, a POSA would have appreciated that implementing Bedingfield's sound source as *a voice of a computer program* would have been advantageous because such an implementation

would have saved time, cost, and storage space, added flexibility to the vocal range of the voice, and improved clarity.

2. 1[a]/8[a]

1[a] *providing, through* an electronic earphone being located at a head of the user and being in communication with a handheld portable electronic device (HPED), *binaural sound to the user during the voice exchange between the computer program and the user such that an origin of the voice of the computer program occurs at the SLP at a first location in empty space that is away from and proximate to the user*;	8[a] *provide, through* a wearable electronic device on the person, *binaural sound to the person during the voice exchange that is between the computer program and the person such that an origin of the voice of the computer program occurs at the SLP at a first location in empty space that is away from and proximate to the person*;

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

206. Bedingfield's *electronic system* includes a multi-channel speaker system 120 depicted in Fig. 1 as headphones, which includes *an electronic earphone on a person being located at a head of the user and being in communication with a* multi-channel processing device 118 (which the POSA would have known to implement as an *HPED*) (1[a]), or *a WED on the person* (8[a]). Ex. 1005, 4:31-51, Fig. 1; *see also* 7:43-46, 13:34-36. It was well-known and common-place in 2014 for *electronic earphones* such as earbuds or headphones to connect to such processing devices. Specifically, Bedingfield states the audio signals provided to the multi-channel speaker system 120 "enable the listener 102 to hear the speaking

participant in the sound field 122," which represents the perception for the listener 102 of the participants 104, 106, and 108 as if they "were in the same location and speaking to each other directly" such that the listener "would hear sounds from the participants [] through both of the listener's [] ears." *Id.*, 4:47-5:3. By rendering a sound field (which includes multiple SLPs) making use of both ears of a listener (*user*), Bedingfield discloses *providing binaural sound to the user*.

207. Specific to 1[a], while Bedingfield does not expressly describe the multi-channel processing device 118 as an *HPED*, in describing Fig. 1, Bedingfield describes participants of the conference call as having "communication devices 110, 112, 114 [that] may be telephones, wireless phones, [or] mobile phones" (Ex. 1005, 4:23-30; *see also* 11:41-51 (describing computing devices as "mobile, handheld, or laptop computer; a mobile telephone; a [PDA]," and so on), any of which are *HPEDs*. Based on these teachings, it is my opinion that the POSA would have understood one option for the multi-channel processing device 118 would have been an *HPED*, as such an implementation would have made it possible for the listener 102 to move around with the multi-channel processing device 118 and remain connected to the conference call. Moreover, specific to 8[a], Bedingfield's headphones were wireless and therefore in my opinion meet *WED* under 8KHZ's position. Ex. 1005, 13:39-43 (describing wireless connection).

208. As I explained for claims 1[pre] and 8[pre], Bedingfield renders obvious *the voice of the computer program*, whereby in Bedingfield, providing the *binaural sound* occurs *during* a conference call (*the voice exchange between the computer program and the user*) *such that* a participant has a specific location within the sound field (*an origin of the voice of the computer program occurs at the SLP*). *See* Section XIV.C.1 (Ground 6, 1[pre]/8[pre]).

209. In Bedingfield, the *SLP* is *at a first location in empty space that is away from and proximate to the user* because as Bedingfield depicts in Fig. 1, the sound field 122 "represents the perception resulting from the audio output" for the listener 102 and is similar to "an actual sound field where the listener 102 is present between all of participants 104, 106, 108." Ex. 1005, 4:40-5:3; *see also* 9:60-64 (describing a voice "occupy[ing] a 'vacant' location"). Because each participant has its own location in the sound field that is separate and at a distance from the listener 102's (*see* analysis for 1[pre]/8[pre]), each participant's *SLP* is *at a first location in empty space that is away from and proximate to the* listener 102 (*user*). Moreover, as I already explained above in Ground 1 for Li, it would have been obvious to position the *SLP* in *empty space*, at a location where no physical object exists. For example, Bedingfield's system "closely simulate[s] an in-person conference setting" (Ex. 1005, 3:25-27, 4:31-5:3), such that the POSA would have found it obvious to seat

the *user/person* at a physical conference table, where an empty seat at the table (*empty space*) is allocated for a virtual participant, as was known in the field.

3. 1[b]/8[b]

1[b] *receiving, from the user* and at the HPED, *a command to move the SLP of the voice of the computer program from the first location in empty space to a second location in empty space that is proximate to the first location and away from and proximate to the user*;	8[b] *receive, from the person, a command to move the SLP of the voice of the computer program from the first location in empty space to a second location in empty space that is proximate to the first location and away from and proximate to the person*;

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

211. Bedingfield's *electronic system* includes the multi-channel processing device 118, which as I explained for claims 1[a] and 8[a], a POSA would have appreciated would have been implemented as an *HPED* (specific to 1[b]). Moreover, as I explained for claims 1[pre] and 8[pre], Bedingfield describes the "listener may make selections on the multi-channel processing device 118 [*HPED*]." Ex. 1005, 5:51-61, Fig. 3 (reproduced below). Therefore, Bedingfield's multi-channel processing device 118 (*HPED*) *receives* selections *from the* listener (*user/person*).

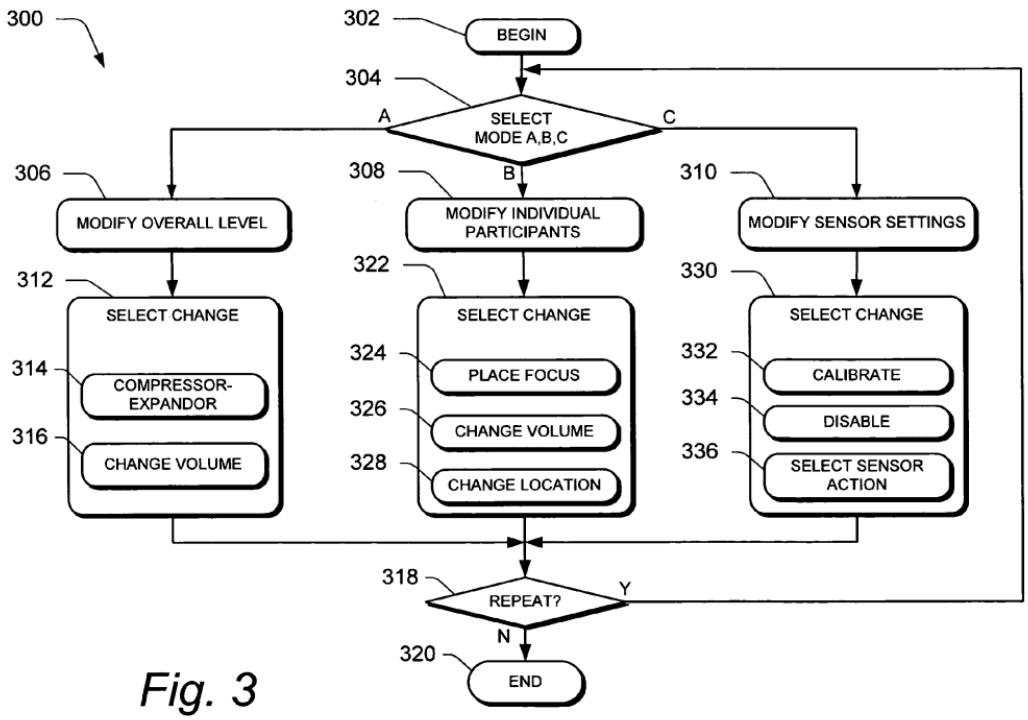


Fig. 3

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. By allowing the listener (*user/person*) to select a participant of the conference call and select a new location for that participant, Bedingfield discloses *receiving, from the user/person, a command to move the SLP of the voice… from the first location in empty space to a second location in empty space that is proximate to the first location and away from and proximate to the user/person*.

212. Regarding the *voice* being the *voice of the computer program*, as I explained for claims 1[pre] and 8[pre], Bedingfield renders obvious that the *voice* is the *voice of the computer program*.

4. 1[c]/8[c]

1[c] *moving, based on the command* and during the voice exchange between the computer program and the user, *the SLP of the voice of the computer program from the first location in empty space to the second location in empty space that is proximate to the first location and away from and proximate to the user*; and	8[c] *move, in response to the command, the SLP of the voice of the computer program from the first location in empty space to the second location in empty space that is proximate to the first location and away from and proximate to the person*; and

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

214. Bedingfield's *electronic system moves, based on the command and during the* conference call (*voice exchange between the computer program and the user*), *the* location (*SLP*) *of* a speaker's voice (*voice of the computer program*) *from the* previous position (*first location in empty space*) *to the* "new location" (*second location in empty space*), where the new location (*second location*) *is* within the sound field (*proximate to the first location*) *and away from and proximate to the* listener (*user*). Ex. 1005, 4:18-5:3, 5:51-61, 6:29-32, 6:43-45, Figs. 1, 3; *see also* analysis of 1[pre]/8[pre], 1[a]/8[a], 1[b]/8[b].

5. 1[d]/8[d]

1[d] *providing, through the* electronic earphone being located at the head of the user and being in communication with the HPED, *the binaural sound to the user during the voice exchange between the computer program and the user such that the origin of the voice of the computer*	8[d] provide, through the wearable electronic device on the person, *the binaural sound to the person during the voice exchange that is between the computer program and the person such that the origin of the voice of the computer program occurs at the SLP at the second location in empty*

program occurs at the SLP at the second location in empty space that is proximate to the first location and away from and proximate to the user.	*space that is proximate to the first location and away from and proximate to the person.*

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

216. Bedingfield's *electronic system provides, through* headphones (*electronic earphone being located at a head of the user being in communication with the HPED* (1[d]) or *WED on the person* (8[d])) connected to a multi-channel processing device 118 (*HPED*), the audio rendering of the sound field making use of both ears of the listener (*binaural sound*) *to the user/person during* the conference call (*the voice exchange between the computer program and the user/person*). *See* Section XIV.C.2 (Ground 6, 1[a]/8[a]). Moreover, in Bedingfield, after the listener (*user/person*) selects at block 328 a new location for the participant within the sound field in 1[c]/8[c], the location for that participant is updated to the new location (*such that the origin of the voice of the computer program occurs at the SLP at the second location in empty space that is away from and proximate to the user/person*). *See* Section XIV.C.1-4 (Ground 6, 1[pre]/8[pre]-1[c]/8[c]).

D. Independent Claims 15:

1. 15[Pre]: *An electronic system with one or more electronic devices that move a sound localization point (SLP) of a voice of a computer program during a voice exchange between the computer program and a person, the electronic system*

comprising.

217. To the extent limiting, it is my opinion that Bedingfield renders obvious the preamble, and Bedingfield's user is a *person*. *See* Section XIV.C.1 (Ground 6, 1[pre]/8[pre]).

2. 15[a]: *a wearable electronic device that the person wears and that provides the voice of the computer program to the person at the SLP at a first location in empty space that is away from and proximate to the person;*

218. In my opinion, Bedingfield renders obvious this limitation, and Bedingfield's headphones is *a wearable electronic device that the* listener (*person*) *wears*. *See* Section XIV.C.1, XIV.C.2 (Ground 6, 1[pre]/8[pre], 1[a]/8[a]).

3. 15[b]: *a memory that stores instructions; and*

219. In my opinion, Bedingfield discloses this limitation. Bedingfield states "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 (emphases added); *see also* 11:23-36 (describing "computer executable *instructions*" as being stored in "*memory* storage devices" (emphases added)), 12:5-44 (describing a processor that reads "executable *instructions* from… the storage medium" (emphasis added)), Fig. 12. Bedingfield therefore contemplates that the communication system 116, the multi-channel processing device 118, and/or the multi-channel speaker system 120 (together, *the*

electronic system) implement functions using software and would include *a memory that stores instructions*.

4. 15[c]: *a processor that executes the instructions to . . .*

220. In my opinion, Bedingfield discloses this limitation. Bedingfield's communication system 116, multi-channel processing device 118, and the multi-channel speaker system 120 (together, *the electronic system*) include *a processor that executes the instructions*. Ex. 1005, 11:13-16 (describing processor(s) that execute program code stored in memory); *see also* 11:60-12:16 (describing processors coupled to user interfaces and reading executable instructions from memory), 12:25-43 (describing processors executing software modules). Bedingfield's *electronic system* therefore includes a *processor that executes the instructions*. *See* Section XIV.C.1 (Ground 6, 1[pre]/8[pre]).

5. 15[c1]: *receive a command from the person to move the SLP of the voice of the computer program from the first location in empty space to a second location in empty space that is proximate to the first location and away from and proximate to the person,*

221. In my opinion, Bedingfield discloses this limitation. *See* Section XIV.C.3 (Ground 6, 1[b]/8[b]).

6. 15[c2]: *move the SLP of the voice of the computer program in response to the command from the first location in empty space to the second location in empty space, and*

222. In my opinion, Bedingfield discloses this limitation. *See* Section XIV.C.4 (Ground 6, 1[c]/8[c]).

7. 15[c3]: *provide binaural sound to the person through the wearable electronic device during the voice exchange between the computer program and the person such that an origin of the voice of the computer program occurs at the SLP at the second location in empty space that is proximate to the first location and away from and proximate to the person*

223. In my opinion, Bedingfield discloses this limitation. *See* Section XIV.C.5 (Ground 6, 1[d]/8[d]).

E. Dependent Claim 17: *processor further executes the instructions to change an interaural time difference (ITD) and an interaural level difference (ILD) of the binaural sound in order to move the SLP of the voice of the computer program from the first location in empty space to the second location in empty space*

224. In my opinion, Bedingfield discloses this limitation. Bedingfield discloses that *ITD* and *ILD* are cues that provide directionality and place a sound source in the sound field. Regarding *ILD*, Bedingfield explains in relation to Fig. 6 that 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. Regarding *ITD*, Bedingfield refers to applying a "delay" as an audio technique to place a sound source in the sound field.

Id., 5:34-43. As the POSA would have readily appreciated, Bedingfield's delay would have included the well-known concept of an *ITD*.

225. As Bedingfield describes each of these cues as being used to set the direction or position of a sound source in the sound field, the POSA would have known to use Bedingfield's *processor* to *execute instructions to change an ITD and an ILD of the binaural sound in order to move the SLP*.

XV. GROUND 7: BEDINGFIELD IN VIEW OF BALLAS RENDERS OBVIOUS CLAIMS 4, 6, 9-10, 14, AND 18

A. Motivation to Combine Bedingfield with Ballas

226. In my opinion, it would have been obvious to a POSA to modify Bedingfield 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. 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. Bedingfield further discloses the listener providing 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. A POSA would have been motivated to combine the teachings of Bedingfield and Ballas because a POSA would have recognized that both references have a shared goal of providing a realistic experience for the user and implementing Ballas's head tracking would relieve the user of

having to input manual commands. To make this goal a reality, the POSA would have understood to track the user's head orientation and movements and provide Bedingfield's sound field to account for the user's orientation and movements, using conventional head tracking and spatialization techniques taught by Ballas. Thus, although Bedingfield does not describe tracking a head orientation and adjusting the sound field based on the user's movement, it is my opinion that a POSA would have been motivated to perform such an implementation, as doing so has the benefits described below.

227. In my opinion, the primary goal of Bedingfield is to provide a sound field that "closely simulate[s] an in-person conference setting." Ex. 1005, 3:25-27. As I explained in Ground 2, Ballas shares this goal. Ex. 1008, ¶¶2, 9, 10. Bedingfield contemplates detecting the listener's orientation and changing the sound field. Ex. 1005, 1:56-59, 13:38-39, 13:58-64, 14:32-38, 18:49-52. To simulate the in-person conference setting, tracking the listener's head and changing the sound field so that the sound sources appear fixed to the listener are key. Modifying Bedingfield in accordance with Ballas's teachings of head tracking and adjusting the sound field to appear fixed to the user would have allowed the user to move while staying connected to the event, and achieved Bedingfield's express purpose to simulate an in-person conference setting.

228. In my opinion, the modification would have been a natural extension of Bedingfield's use of motion detectors, sensors, and devices to detect the listener's orientation. Ex. 1005, 13:54-56. As I explained in Ground 2, 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 would have allowed the user to change head location and orientation while maintaining fixed virtual sound locations, simulating an in-person conference setting. This implementation would have advanced the common goals of both Bedingfield and Ballas to provide a realistic environment in a teleconferencing application.

B. Reasonable Expectation of Success

229. In my opinion, a POSA would have had a reasonable expectation of success in modifying Bedingfield in accordance with the teachings of Ballas. Bedingfield and Ballas both track the listener's movement and orientation and change the sound field accordingly. In Bedingfield, the user wears "wearable speakers" in the form of stereo headphones. Ex. 1005, 13:35-36. As I explained in Ground 2, Ballas also discloses speakers that are attached to a wearable head frame. Ex. 1008, ¶25. Accordingly, a POSA would have recognized that Bedingfield and Ballas are structurally and functionally similar.

230. In my opinion, a POSA implementing the Bedingfield system would have recognized that Bedingfield's headphones could be 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 4

1. 4[a]: *tracking head orientation of user as user moves with respect to SLP at the second location*

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

232. Bedingfield discloses "a device to detect the orientation of a listener 104" (Ex.1005, 13:38-39; *see also id.*, 6:54-58, 13:52-57), which is *tracking a[n] orientation of the user*. The listener's orientation is relative to *the SLP*. *See* Ex. 1005, 1:56-59, 18:49-52. However, Bedingfield does not explicitly disclose that what is *track[ed]* is the *orientation* of the user's *head, as the user moves with respect to the SLP at the second location*. However, this would have been obvious in view

of Ballas, which discloses this limitation as I explained in Ground 2. *See* Section X.C.1 (Ground 2, 4[a]).

233. POSA would have been motivated to implement Bedingfield's wearable speakers as part of Ballas's wearable frame having a head tracking system, as 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 XV.A (Motivation to Combine Bedingfield with Ballas).

2. 4[b]: *adjusting voice of computer program that user hears through electronic earphone so voice continues to appear to originate from SLP at the second location even as the user moves with respect to the second location.*

234. In my opinion, Bedingfield in view of Ballas discloses this limitation.

235. 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. As Bedingfield describes for Figures 14 and 15 (reproduced below with my 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.

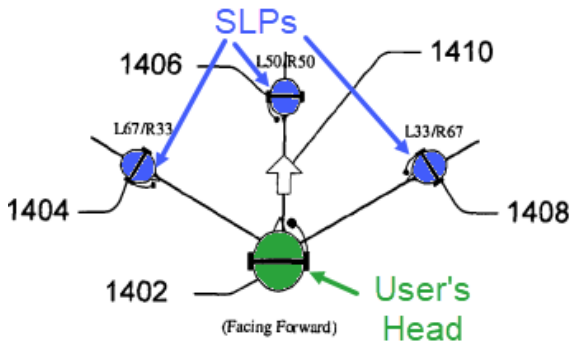


Fig. 14

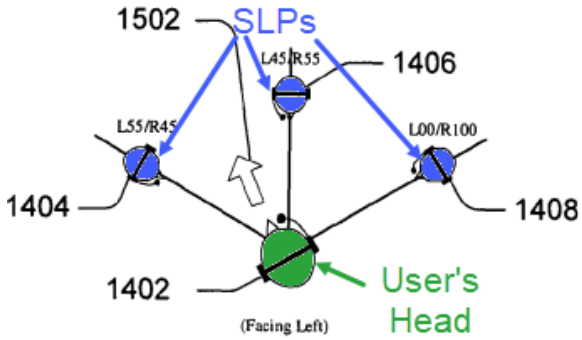


Fig. 15

In this manner, Bedingfield discloses *adjusting a sound… that the user hears through the electronic earphone so the sound of the voice of the computer program continues to appear to originate from the SLP at the second location.*

236. While Bedingfield's sound field is adjusted based on manual input, not according to the user's movement, this would have been obvious in view of Ballas, which discloses this limitation as discussed for Ground 2. *See* Section X.C.2 (Ground 2, 4[b]).

D. Dependent Claim 6: *changing an amplitude, an interaural time difference, and an interaural level difference of the voice of the computer program as the user moves with respect to the SLP at the second location so a sound of the voice of the computer program continues to appear to originate from the SLP at the second location as the user moves*

237. As I explained in Ground 6 for claim 17, Bedingfield renders obvious *changing an ILD* and an *ITD*. *See* Section XIV.E (Ground 6, 17). Bedingfield also refers to overall "volume" (*amplitude*) as providing relative distance information.

Ex. 1005, 3:45-47 ("A signal of lower volume sounds farther away than a signal of higher volume."). As Bedingfield describes each of these cues as being used to set the direction or position of a sound source in the sound field, the POSA would have found it obvious to *chang[e] an amplitude, an ITD, and an ILD of the voice* to modify the sound field *so a sound of the voice… continues to appear to originate from the SLP… as the user moves*.

E. Dependent Claim 9: *maintain the origin of the voice of the computer program at the SLP at the second location in empty space but alter an amplitude of the voice of the computer program that the person hears in response to the person moving closer to or away from the SLP at the second location*

238. In my opinion, Bedingfield in view of Ballas discloses this limitation. *See* Section XV.D (Ground 7, 6).

F. Dependent Claim 10

1. 10[a]: *monitor movements of head orientation of person in response to person moving with respect to SLP at second location in empty space; and*

239. In my opinion, Bedingfield in view of Ballas discloses this limitation. *See* Section XV.C.1 (Ground 7, 4[a]).

2. 10[b]: *adjust an interaural time difference (ITD) and an amplitude of the voice of the computer program in response to the movements of the head orientation of the person as the person moves*

240. In my opinion, Bedingfield in view of Ballas discloses this limitation. *See* Section XIV.E (Ground 6, 17) and Section XV.C.2 and XV.D (Ground 7, 4[b] and 6).

G. **Dependent Claim 14:** *adjust interaural level difference, interaural time difference, and amplitude for voice person hears so SLP of voice remains fixed at second location in empty space while head orientation of person moves*

241. In my opinion, Bedingfield in view of Ballas discloses this limitation.

242. As I explained for claims 4[a] and 4[b], Bedingfield in view of Ballas discloses *adjust[ing]* a *voice that the person hears so the SLP of the voice of the computer program remains fixed at the second location in empty space while a head orientation of the person moves with respect to the second location. See* Section XV.C (Ground 7, 4). As I explained for claim 6, Bedingfield renders obvious *adjusting an ILD, and ITD, and an amplitude for the voice*. *See* Section XV.D (Ground 7, 6).

H. **Dependent Claim 18:** *processor further executes instructions to change interaural time difference (ITD) and amplitude of binaural sound to maintain SLP fixed at second location in empty space during voice exchange while person physically moves away from second location in empty space and changes head orientation with respect to second location in empty space*

243. As I explained for claim 6, Bedingfield in view of Ballas discloses *chang[ing] an ITD and an amplitude of the binaural sound in order to maintain the SLP fixed at the second location in empty space… while the person… changes a head*

orientation with respect to the second location in empty space. Section XV.D (Ground 7, 6). As I explained for claim 15[c], Bedingfield's *processor executes instructions* to perform this processing, and as I explained for claims 1[pre] and 8[pre], this occurs *during the voice exchange*. *See* Section XIV.D.4 (Ground 6, 15[c]) and Section XIV.C.1 (Ground 6, 1[pre]/8[pre]).

244. As I explained in Ground 2, Ballas further discloses *the person physically mov[ing] away from the second location in empty space. See* Section X.H (Ground 2, 18). To provide as realistic and as immersive of an experience as possible for the user, the combined system of Bedingfield and Ballas would have processed the sound *in order to maintain the SLP fixed at the second location* regardless of the user's movements or head orientation.

XVI. GROUND 8: BEDINGFIELD IN VIEW OF BALLAS AND SANDGREN RENDERS OBVIOUS CLAIM 5

A. Motivation to Combine

245. In my opinion, A POSA would be motivated to combine the teachings of Bedingfield and Ballas. *See* Section XV.1 (Ground 7, Motivation to Combine Bedingfield with Ballas). Further, Bedingfield recognizes that a sound's "volume" (*amplitude*) provides distance information. Ex. 1005, 3:45-47 ("A signal of lower volume sounds farther away than a signal of higher volume."). A POSA would have been motivated to look to Sandgren to determine how to vary Bedingfield's sound

volume with distance, to achieve Bedingfield's stated goal of providing a realistic conferencing experience. Sandgren shares this goal and additionally states that the sound intensity follows an inverse-square law with distance. Ex. 1017, ¶¶1, 77, Abstract. Modifying Bedingfield and Ballas in accordance with Sandgren's teachings to adjust a sound amplitude according to an inverse-square law with distance would have provided a more realistic and immersive experience for the user, and would have been a natural extension of Bedingfield's disclosure of a volume setting. There is nothing in Ballas that would prevent a POSA from modifying Bedingfield with Sandgren's teachings.

B. Reasonable Expectation of Success

246. In my opinion, a POSA would have had a reasonable expectation of success in modifying Bedingfield with Ballas. *See* Section XV.B (Ground 7, Reasonable Expectation of Success). A POSA would have had a reasonable expectation of success in further modifying Bedingfield in accordance with the teachings of Sandgren. Bedingfield and Sandgren both simulate realistic auditory scenes having spatialized binaural sound and provide such sounds over headphones. As I explained in Ground 3, Sandgren creates virtual 3D audio scenes to externalize virtual sound sources. *See* Section XI.B (Ground 3, Reasonable Expectation of Success). Accordingly, Bedingfield and Sandgren are structurally and functionally similar.

247. In my opinion, a POSA implementing the Bedingfield system would have recognized that the spatialization of sound would have been accompanied 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 (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 5: *adjust amplitude of voice of computer program to be inversely proportional to square of distance between user and SLP at second location in empty space as user moves with respect to SLP at second location in empty space*

248. In my opinion, Li in view of Ballas and Sandgren discloses this limitation.

249. As I explained for claim 6, Bedingfield in view of Ballas discloses *adjust[ing] an amplitude of the voice of the computer program… as the user moves with respect to the SLP at the second location in empty space. See* Section XV.D (Ground 7, 6).

250. As I explained in Ground 3, Sandgren discloses an inverse-square law, and to provide as realistic of a sound field as possible, the POSA would have found it obvious to *adjust an amplitude… to be inversely proportional to a square of a distance between the user and the SLP*. *See* Section XI.C (Ground 3, claim 5).

XVII. GROUND 9: BEDINGFIELD IN VIEW OF SANDGREN RENDERS OBVIOUS CLAIM 19

A. Motivation to Combine

251. In my opinion, it would have been obvious to a POSA to modify Bedingfield 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. A POSA would have been motivated to combine the teachings of Bedingfield and Sandgren because a POSA would have recognized that both references have a shared goal of providing a realistic and immersive experience for the user. To make this goal a reality, the POSA would have understood to simulate sound-reflecting objects that do not actually exist but are perceived to exist in the virtual environment, using conventional sound enhancement techniques taught by Sandgren. Thus, although Bedingfield does not describe changing attenuation and reverberation of the binaural sound to add artificial audial cues of virtual objects, a POSA would have been motivated to perform such an implementation, as doing so has the benefits described for Ground 4. Modifying Bedingfield 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.

252. Moreover, it is my opinion that the modification would have been a natural extension of Bedingfield's 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. For the same reasons as discussed for Ground 4, a POSA would have recognized that the teachings of Sandgren 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 and Sandgren to provide a realistic and immersive experience for the user.

B. Reasonable Expectation of Success

253. In my opinion, a POSA would have had a reasonable expectation of success in modifying Bedingfield in accordance with the teachings of Sandgren. *See* Section XVI.B (Ground 8, Reasonable Expectation of Success).

C. Dependent Claim 19

1. 19[a]: *The electronic 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 person,

254. In my opinion, Bedingfield in view of Sandgren discloses this limitation. Sandgren discloses reverberation generators and attenuation filters *to change attenuation and reverberation of binaural sound to add artificial audial cues of virtual objects near a person*. *See* Section XII.C.1(Ground 4, 19[a]).

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

255. In my opinion, Bedingfield in view of Sandgren discloses this limitation. Sandgren discloses objects that *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*. *See* Section XII.C.2 (Ground 4, 19[b]).

XVIII. GROUND 10: BEDINGFIELD AND WOO RENDERS OBVIOUS CLAIM 11

A. Motivation to Combine

256. In my opinion, it would have been obvious to a POSA to modify Bedingfield to adjust the voice to include artificial cues based on actual physical objects in the user's room, as taught in Woo. As the POSA would have recognized, an implementation that deliver a sound perceived by the user to actually originate from the SLP in the user's room, would have enhanced Bedingfield's conferencing system and achieved Bedingfield's goal of simulating a face-to-face conference as

closely as possible. Woo photographs an actual object in the user's room and adjusts the sound according to the object's position in the room. The POSA would have known to combine this conventional teaching in Woo with Bedingfield's system so that the sound is perceived by the user to actually originate from the SLP in the user's room, and accounts for physical objects in the room.

257. The combination would have been a natural extension of Bedingfield's remote conferencing system. For example, using Woo's detection of actual physical objects in the room and corresponding adjustment of the sound would have helped to achieve Bedingfield's primary goal of providing an enhanced sound field that "closely simulate[s] an in-person conference setting." Ex. 1005, 3:25-27. Moreover, in my opinion, the POSA would have understood that adjusting a sound based on photographs of actual objects in the room would have enhanced Bedingfield's sound delivery and provided a more realistic conferencing experience.

B. Reasonable Expectation of Success

258. In my opinion, a POSA would have had a reasonable expectation of success in modifying Bedingfield in accordance with the teachings of Woo. Consistent with Bedingfield's system, Woo tracks a user's position, adjusts certain properties of a sound and optimizes the delivery of the sound for the user. Accordingly, a POSA would have recognized that Bedingfield and Woo are structurally and functionally similar.

259. In my opinion, a POSA implementing the Bedingfield system would have recognized that the processing of the sound would have been accompanied with Woo's adjustment of the sound based on actual physical objects in the user's room, in order to artificially simulate their presence and simulate an in-person conference. Such an implementation would have merely involved a combination of known prior art elements (camera and sound processor) according to known methods and techniques (simulate spatialized sounds) to yield predictable results (producing sound so that the user perceives the sound as if it originates in the user's actual room), and involves use of a known technique (generating realistic spatializations for sounds) 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 11

1. 11[a]: *determine physical objects located in a room with the person*

260. In my opinion, Bedingfield in view of Woo discloses this limitation. Woo discloses *to determine physical objects located in room with person* by photographing objects to recognize their location in a room. *See* Section *(*Ground 5, 11[a]).

2. 11[b]: *adjust, based on physical objects located in room, the voice of computer program that person hears to emulate sound that*

> *person would hear if voice of computer program actually originated from SLP at second location.*

261. In my opinion, Bedingfield in view of Woo discloses this limitation. Woo *adjust* the sound field *based on the* object 200 (*physical objects located in the room*) by describing that the sound is adjusted based on the object's position in the images *to emulate sound that a person would hear if a voice of the computer program actually originated from SLP at the second location*. *See* Section XIII.C.2 (Ground 5, 11[b]).

XVI. SECONDARY CONSIDERATIONS

262. I am not aware of any secondary considerations of non-obviousness for the claims of the '196 patent. If Patent Owner raises any secondary considerations in its briefing or expert's declaration, I reserve the right to respond to those considerations.

XVII. CONCLUSION

263. For these reasons, it is my opinion that claims 1, 4-6, 8-11, 14-15, and 17-19 of the '196 patent are obvious.

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

Dr. Chris Kyriakakis