

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re *Inter Partes* Review of:)
U.S. Patent No. 9,674,628)
Issued: Jun. 6, 2017)
Application No.: 15/441,199)
Filing Date: Feb. 23, 2017)

For: Providing Binaural Sound to Localize at an Image During a Telephone Call

FILED VIA P-TACTS

**DECLARATION OF DR. CHRIS KYRIAKAKIS IN SUPPORT OF
PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 9,674,628**

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

TABLE OF CONTENTS

I.	INTRODUCTION.....	8
II.	EXPERIENCE, QUALIFICATIONS, AND STATUS AS AN INDEPENDENT EXPERT WITNESS	9
III.	UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES	14
	A. Claim Construction	14
	B. Obviousness	15
IV.	SUMMARY OF MATERIALS REVIEWED AND CONSIDERED	17
V.	LEVEL OF SKILL IN THE ART AND PERSPECTIVE APPLIED IN THIS DECLARATION	20
VI.	BACKGROUND.....	22
	A. Overview of the Technology	22
	1. Wearable virtual reality	22
	2. Binaural sound.....	25
	3. Sound source localization.....	30
	4. Creating binaural sound.....	34
	5. Compensating for head movement.....	37
	B. The '628 Patent.....	42
	C. Prosecution History	45
	D. Challenged Claims of the '628 Patent	46
	E. Effective Priority Date	47
VII.	CLAIM CONSTRUCTION	47
VIII.	THE PRIOR ART	48
	A. Billinghamst	48
	B. Schröder	51
	C. Zhou	52
	D. Haddick.....	53
	E. Ballas	54

IX.	GROUND 1: BILLINGHURST IN VIEW OF SCHRÖDER RENDERS OBVIOUS CLAIMS 9-14 AND 16-20	55
A.	Motivation to Combine Billinghamurst with Schröder	55
B.	Reasonable Expectation of Success.....	59
C.	Independent Claim 9	61
	1. 9[pre]: “ <i>A method executed by one or more electronic devices in an electronic system that improves an experience of a calling party and a called party talking to each other during a telephone call, the method comprising: ”</i>	61
	2. 9[a]: “ <i>processing, during the telephone call and with a processor, a voice of the calling party with head related transfer functions (HRTFs) so the voice of the calling party localizes to the called party as binaural sound at a sound localization point (SLP)that is one meter to two meters away from the called party; ”</i>	67
	3. 9[b]: “ <i>displaying, during the telephone call and to the called party with a wearable electronic device worn on a head of the called party, an image that represents the calling party that is seen by the called party as being at the SLP that is one meter to two meters away from the called party; ”</i>	80
	4. 9[c]: “ <i>tracking, during the telephone call and with the wearable electronic device, head movements of the called party with respect to the image that represents the calling party and that is seen by the called party as being at the SLP; and”</i>	83
	5. 9[d]: “ <i>changing, during the telephone call and in response to the head movements of the called party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the SLP that is seen by the called party as being the image that represents the calling party that is one meter to two meters away from the called party while the head of the called party moves with respect to the image that represents the calling party. ”</i>	87
D.	Dependent Claims 10-14	94

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

1. Claim 10: “*The method of claim 9 further comprising: 94 tracking, during the telephone call and with a wearable electronic device worn on a head of the calling party, head movements of the calling party; 95 moving, during the telephone call, the image that represents the calling party to copy the head movements of the calling party so the called party sees, with the display of the wearable electronic device worn on the head of the called party, the head movements of the calling party; and 95 changing, during the telephone call and in response to the head movements of the calling party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the image that represents the calling party that is seen by the called party as being at the SLP that is one meter to two meters away from the called party while the image that represents the calling party moves.*” 95
2. Claim 11: “*The method of claim 9 further comprising: 99 tracking, during the telephone call and with a wearable electronic device worn on a head of the calling party, head movements of the calling party; and 99 changing, during the telephone call and in response to the head movements of the calling party, the voice of the calling party so the voice of the calling party emulates a sound that the called party would hear if the calling party were physically present at the SLP and talking to the called party.*” 99
3. Claim 12: “*The method of claim 9 further comprising: displaying, during the telephone call and to the called party with the wearable electronic device worn on the head of the called party, the image that represents the calling party as a virtual image of the calling party.*” 100

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

4.	Claim 13: “ <i>The method of claim 9 further comprising: displaying, during the telephone call and to the called party with the wearable electronic device worn on the head of the called party, the image that represents the calling party as an augmented reality image of the calling party.</i> ”	101
5.	Claim 14: “ <i>The method of claim 9 further comprising:</i> ”	102
	<i>moving, during the telephone call, the image that represents the calling party around a room where the called party is located; and</i>	102
	<i>changing, during the telephone call and in response to movements in the room of the image that represents the calling party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the image that represents the calling party even as the image that represents the calling party moves around the room.”</i>	102
E.	Independent Claim 16:.....	105
1.	16[pre]: “ <i>A non-transitory computer readable storage medium storing instructions that one or more electronic devices in an electronic system execute as a method during a telephone call between a calling party and a called party, the method comprising:</i> ”	105
2.	16[a]: “ <i>displaying, during the telephone call and to the called party with a wearable electronic device worn on a head of the called party, an image of the calling party that appears to the called party as being at least one meter away from a head of the called party;</i> ”	106
3.	16[b]: “ <i>processing, during the telephone call, a voice of the calling party with head related transfer functions (HRTFs) so the voice of the calling party localizes to the called party as binaural sound at the image of the calling party that appears at least one meter away from the head of the called party;</i> ”	106
4.	16[c]: “ <i>tracking, during the telephone call, head movements of the called party; and.</i> ”	107

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

5.	16[d]: “ <i>changing, during the telephone call and in response to the head movements of the called party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the image of the calling party that appears at least one meter away from the head of the called party as the head of the called party moves and changes head orientations.</i> ”	107
F.	Dependent Claims 17-20:	108
1.	Claim 17: <i>The non-transitory computer readable storage medium storing instructions of claim 16 further to cause the one or more electronic devices in the electronic system to execute the method comprising: tracking, during the telephone call, head movements of the calling party; and</i>	108
	<i>changing, during the telephone call and in response to the head movements of the calling party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the image of the calling party that appears at least one meter away from the head of the called party as a head of the calling party moves.</i>	108
2.	Claim 18: “ <i>The non-transitory computer readable storage medium storing instructions of claim 16, wherein a location of the image of the calling party also functions as a virtual microphone point (VMP) for sound that is captured by the wearable electronic device worn on the head of the called party, and the sound captured by the wearable electronic device worn on the head of the called party is provided to the calling party.</i> ”	109

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

3.	Claim 19: “ <i>The non-transitory computer readable storage medium storing instructions of claim 16 further to cause the one or more electronic devices in the electronic system to execute the method comprising: changing, during the telephone call, the HRTFs processing the voice of the calling party so the voice of the calling party moves and follows the image of the calling party as the image of the calling party moves with respect to the called party.</i> ”	112
4.	Claim 20: “ <i>The non-transitory computer readable storage medium storing instructions of claim 16 further to cause the one or more electronic devices in the electronic system to execute the method comprising: providing the image of the calling party as being a virtual microphone point; and</i> ”	113
	<i>reducing, during the telephone call, an amplitude of a voice of the called party that is provided to the calling party when the called party moves away from the virtual microphone point.</i> ”	113
X.	GROUND 2: BILLINGHURST AND SCHRÖDER IN VIEW OF ZHOU AND HADDICK RENDERS OBVIOUS CLAIMS 9-14 AND 16-20.....	114
A.	Motivation to Combine Billinghamhurst and Schröder with Zhou	114
B.	Motivation to Combine Billinghamhurst and Schröder with Haddick.....	116
C.	Reasonable Expectation of Success.....	118
D.	Claims 9-14 and 16-20.....	120
XI.	GROUND 3: BILLINGHURST IN VIEW OF BALLAS RENDERS OBVIOUS CLAIMS 9-14 and 16-20	123
A.	Motivation to Combine Billinghamhurst with Ballas	123
B.	Reasonable Expectation of Success.....	124
C.	Claims 9-14 and 16-20.....	125

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

XII. GROUND 4: BILLINGHURST AND BALLAS IN VIEW OF ZHOU AND HADDICK RENDERS OBVIOUS CLAIMS 9-14 AND 16-20.....	127
XIII. SECONDARY CONSIDERATIONS	128
XIV. CONCLUSION	128

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

I. INTRODUCTION

1. I have been retained by Meta Platforms, Inc. (“Petitioner”) to provide my opinions concerning the validity of claims 9-14 and 16-20 of U.S. Patent No. 9,674,628 (the “’628 patent”).

2. I am simultaneously engaged by Petitioner to provide declarations for IPR petitions challenging U.S. Patent No. 9,226,090 and U.S. Patent No. 9,282,196.

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 9-14 and 16-20 of the ’628 patent are invalid based on the following grounds:

Ground 1	Claims 9-14 and 16-20 are obvious under 35 U.S.C. § 103 over Billingham in view of Schröder.
Ground 2	Claims 9-14 and 16-20 are obvious under 35 U.S.C. § 103 over Billingham in view of Schröder in further view of Zhou and Haddick.
Ground 3	Claims 9-14 and 16-20 are obvious under 35 U.S.C. § 103 over Billingham in view of Ballas.
Ground 4	Claims 9-14 and 16-20 are obvious under 35 U.S.C. § 103 over Billingham in view of Ballas in further view of Zhou and Haddick.

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

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

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

III. UNDERSTANDING OF RELEVANT LEGAL PRINCIPLES

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

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

A. Claim Construction

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

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

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

B. Obviousness

24. I understand that a prior art reference can render a patent claim obvious to one of ordinary skill in the art if the differences between the subject matter set forth in the patent claim and the prior art are such that the subject matter of the claim would have been obvious at the time the claimed invention was made. In analyzing obviousness, I understand that it is important to consider the scope of the claims, the level of skill in the relevant art, the scope and content of the prior art, the differences between the prior art and the claims, and any secondary considerations.

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

26. I understand that obviousness cannot be based on the hindsight combination of components selectively culled from the prior art. I understand that in an obviousness analysis, neither the motivation nor the avowed purpose of the inventors control the inquiry. Any need or problem known in the field at the time of the invention and addressed by the patent can provide a reason for combining elements. For example, I understand that it is important to consider whether there existed at the time of the invention a known problem for which there was an obvious solution encompassed by the patent's claims. I understand that known techniques can have obvious uses beyond their primary purposes, and that in many cases a person of ordinary skill can fit the teachings of multiple pieces of prior art together like pieces of a puzzle.

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

28. I understand that secondary considerations can include, for example, evidence of commercial success of the invention, evidence of a long-felt need that was solved by an invention, evidence that others copied an invention, or evidence that an invention achieved a surprising result. I further understand that such evidence must have a nexus, or causal relationship to the elements of a claim, in an order to be relevant. I am unaware of any such secondary considerations for the '628 patent.

IV. SUMMARY OF MATERIALS REVIEWED AND CONSIDERED

29. All of the opinions contained in this Declaration are based on the documents I reviewed and my knowledge and professional judgment. In forming 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 '628 patent, its file history,

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

the prior art, and all exhibits cited in this declaration, including the documents in the table below.

Exhibit	Description
1001	U.S. Patent No. 9,674,628 (“’628 patent”)
1002	Prosecution History of U.S. Patent No. 9,674,628 (“’628 FH”)
1004	Curriculum Vitae of Dr. Chris Kyriakakis
1008	U.S. Patent App. Pub. No. 2003/0059070 to Ballas (“Ballas”)
1009	U.S. Patent No. 7,606,375 to Bailey, et al. (“Bailey”)
1010	Billinghurst, Mark, et al. "A wearable spatial conferencing space." Digest of Papers. Second International Symposium on Wearable Computers. IEEE, 1998 (“Billinghurst”)
1011	International Pub. No. WO 2013/049248 to Haddick, et al. (“Haddick”)
1012	International Pub. No. WO 2005/032209 to Schröder (“Schröder”)
1013	U.S. Patent App. Pub. 2014/0064526 to Otto, et al. (“Otto”)
1015	Rumsey, Francis “Whose head is it anyway? Optimizing binaural audio.” J. Audio Eng. Soc., Vol. 59, No. 9, 2011 (“Rumsey”)
1018	Zhou, Zhiying, et al. “An experimental study on the role of software synthesized 3D sound in augmented reality environments.” <i>Interacting with Computers</i> 16.5 (2004): 989-1016 (“Zhou”)
1020	Scheinerman, Matt “Exploring Augmented Reality.” Haverford College Computer Science (2009) (“Scheinerman”)
1021	Comeau, Charles et al. “Headsight TV System.” <i>Electronics</i> , 1961, pp. 86-90 (“Comeau”)
1022	Sutherland, Ivan “The Ultimate Display.” <i>Proceedings of IFIP Congress</i> , pp. 506-508 (1965) (“Sutherland1965”)
1023	Sutherland, Ivan “A head-mounted three dimensional display.” <i>Fall Joint Computer Conference</i> , pp. 757-764 (1968) (“Sutherland1968”)

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

Exhibit	Description
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”)
1028	Harma, Aki. “Ambient Human-to-Human Communication.” Handbook of Ambient Intelligence and Smart Environments, 2010, pp. 795-823 (“Harma”)
1029	Okada, Ken-ichi, et al. “Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design.” Association of Computing Machinery, 1994, pp. 385-393 (“Okada”)
1030	Nova, Nicolas. “A Review of How Space Affords Socio-Cognitive Processes during Collaboration.” PsychNology Journal, 2005 (“Nova”)
1032	May 22, 2023 Deposition Transcript of Durand Begault, Ph.D. from <i>Eight KHZ, LLC v. Meta Platforms, Inc. et al.</i> , 6:22-cv-575-ADA-DTG (W.D. Tex.)
1033	U.S. Patent No. 257,453 to Ader (“Ader”)
1034	Olson, Harry. “Sound Reproducing Systems - Monaural, Binaural, Monophonic, and Stereophonic.” Audio, 1958 (“Olson”)
1035	Kyriakakis, Chris. “Fundamental and Technological Limitations of Immersive Audio Systems.” Proceedings of the IEEE, Vol. 86, No. 5, 1998 (“Kyriakakis 1998”)
1036	Begault, Durand. “3-D Sound for Virtual Reality and Multimedia.” 2000 (“Begault”)
1039	Grämer, T “Efficient modeling of head movements and dynamic scenes in virtual acoustics.” University of Zurich, Faculty of Medicien, 2010 (“Grämer”)
1043	Laing, Gordon “Sound effects.” Personal Computer World, June 1997 (“Laing”)
1044	U.S. Patent No. 2,955,156 to Heilig (“Heilig”)

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

31. My opinions are additionally guided by my appreciation of how a POSA would have understood the claims of the '628 patent at the time of the alleged invention, which I have been asked to assume is June 23, 2014.

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

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

33. The invention of the '628 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, 1:15-16.

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

claimed priority date of the '628 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.

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

36. I believe that I would have qualified as a person of at least ordinary skill in the art as of the earliest claimed June 23, 2014 priority date on the face of the '628 patent. I believe that I have a sufficient level of knowledge, experience, and education to provide an expert opinion in the field of the '628 patent.

37. My opinions in this Declaration are based on the perspective of a person of ordinary skill in the art as of June 23, 2014 or somewhat before. This is true even if the testimony is stated in the present tense. Each of the statements below reflects my opinion based on my review of the disclosures of the '628 patent, the file history of the '628 patent, and the challenged claims.

VI. BACKGROUND

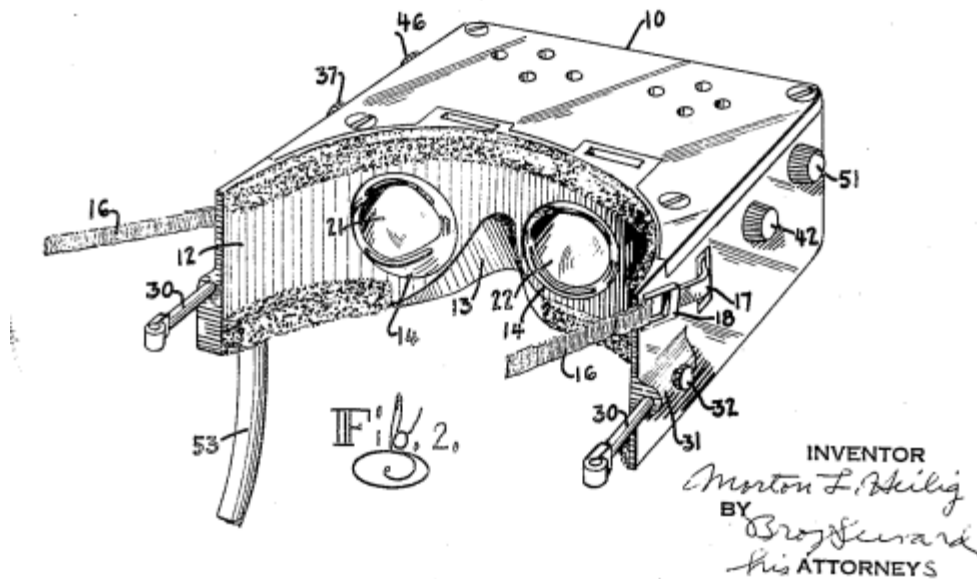
A. Overview of the Technology

1. Wearable virtual reality

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

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

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



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

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

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

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

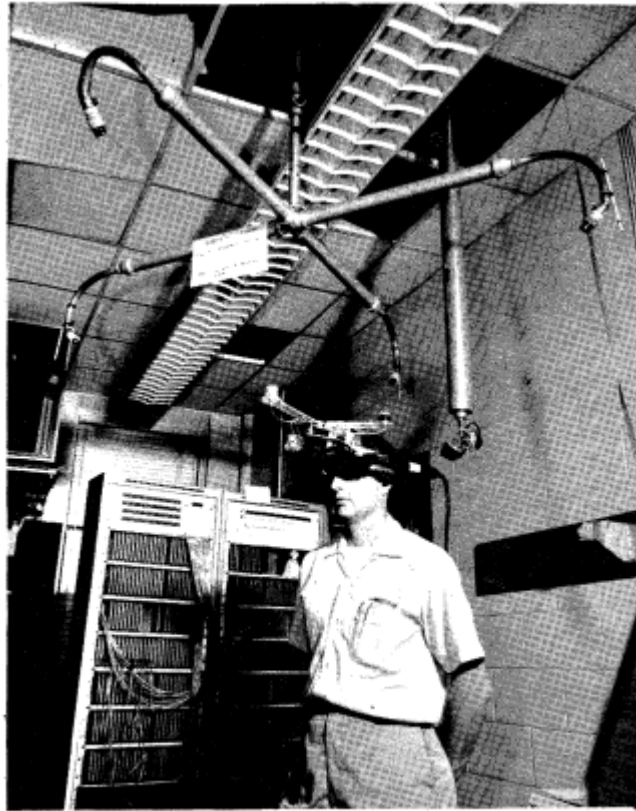


FIGURE 3—The mechanical head position sensor in use

2. Binaural sound

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

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

44. Ader's patent teaches:

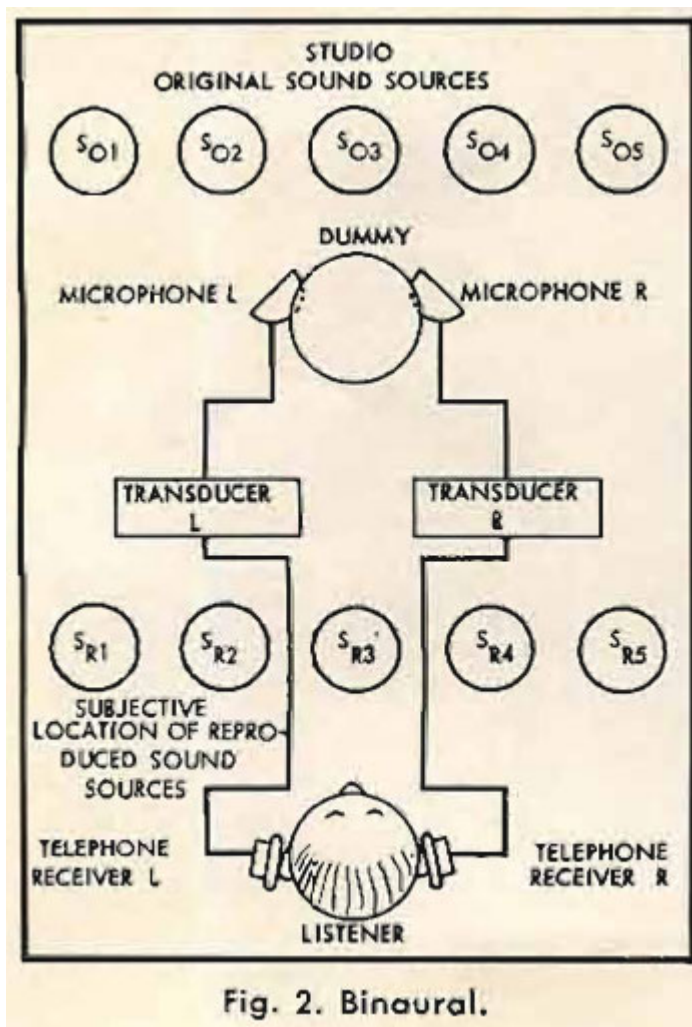
"I place a series of transmitters on the stage... each of these transmitters being arranged, as usual in a local circuit, with a battery and induction-coil and a line-circuit extending from each such coil to the houses of the respective subscribers. Two line-circuits extend to the house of each subscriber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

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

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

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

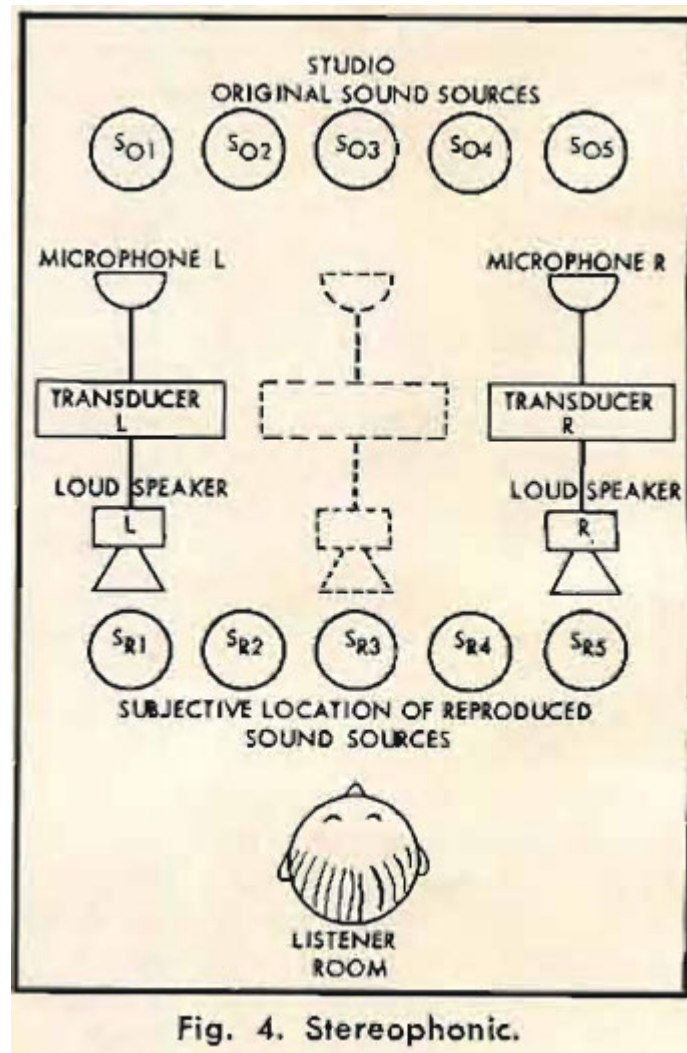
channels. Olson was a prominent acoustical engineering pioneer whose work in microphone and loudspeaker development would have been well-known to a POSA. In September 1958, Olson wrote in an article published in “Audio,” which was the most popular audio magazine when it was in print from 1947 to 2000.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



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

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

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



3. Sound source localization

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

Ex. 1035, p. 944.

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

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

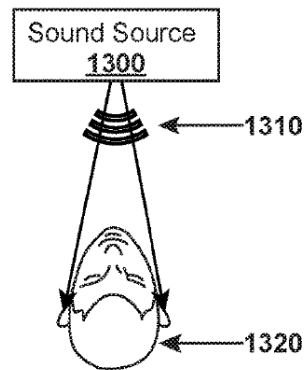


Figure 13A

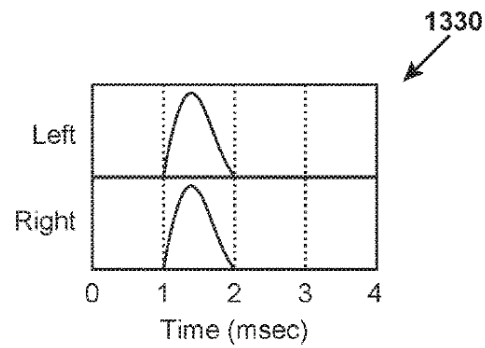


Figure 13B

Ex. 1001, Figs. 13A-13B

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

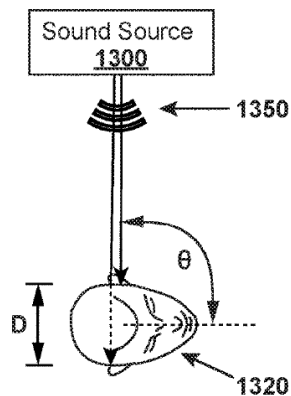


Figure 13C

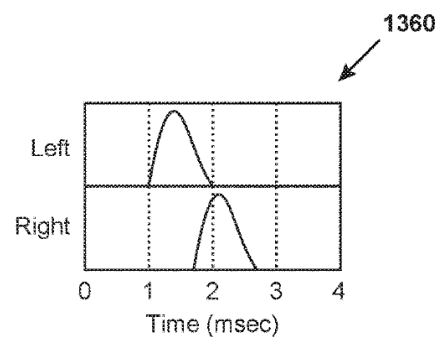


Figure 13D

Ex. 1001, Figs. 13C-13D

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

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

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

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

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

Ex. 1035, 944.

4. Creating binaural sound

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

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

content with the left and right HRTFs of the desired placement angle and position. For example, a monaural sound source, such as a person's voice, is processed (via convolution) with a left ear HRTF and a right ear HRTF, where each HRTF is designated at the desired placement angle and position, such as 30° to the front left of the listener and at a specified distance from the listener. Ex. 1025, 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, 2, 5. At the relevant time period, HRTFs had been known for decades and were commonly used to characterize the manner in which an ear receives a sound from a point in space. Ex. 1024, 1; Ex. 1025, 2, 5.

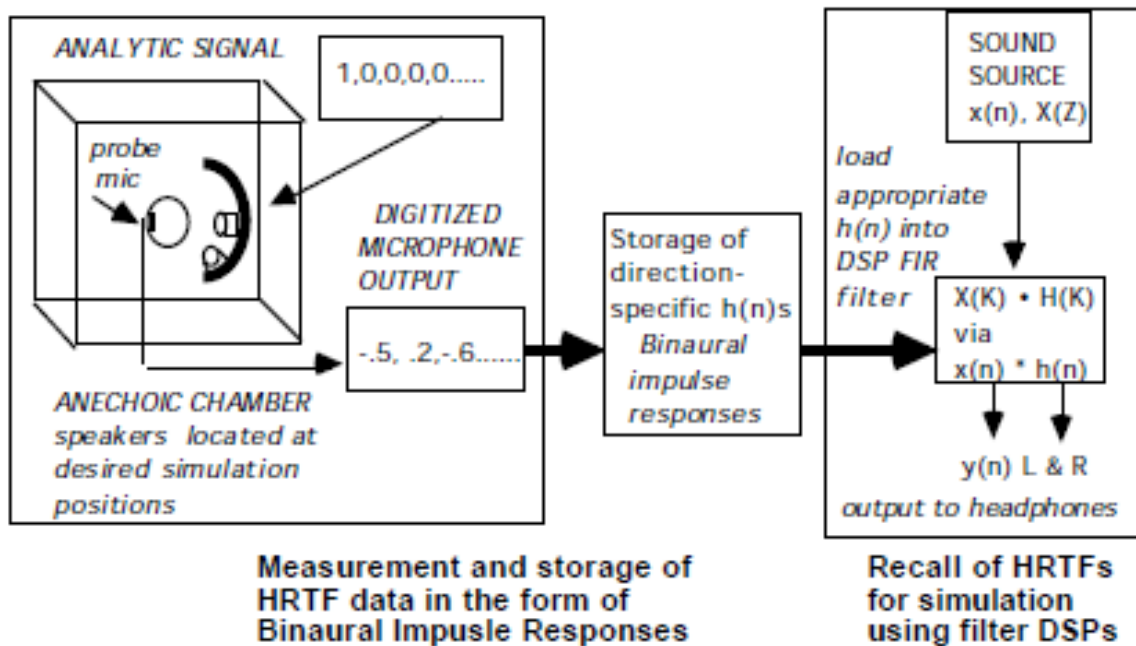
56. HRTFs are typically measured in advance (e.g., in a laboratory) and then made available to anyone wishing to produce synthetic binaural sound. A commonly used set of HRTFs for several hundred angles around the head was made publicly available by Bill Gardner and Keith Martin of the MIT Media Lab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

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

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

Ex. 1036, 109.

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



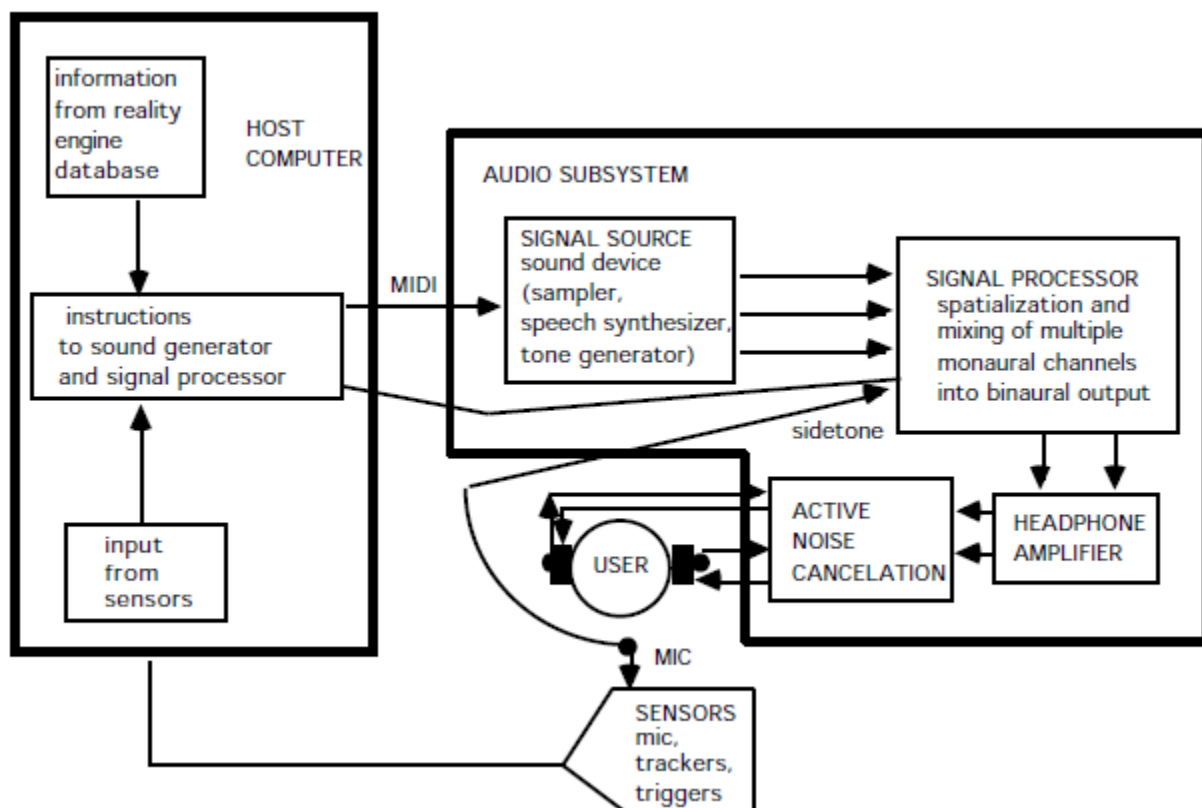
59. Such 3-D sound was useful in various virtual reality environments, such as videoconferencing. *Id.*, 174 (describing arranging spatialized audio signals in virtual space for multiconversation teleconferencing with broadcast video images of the participants).

5. Compensating for head movement

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

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

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

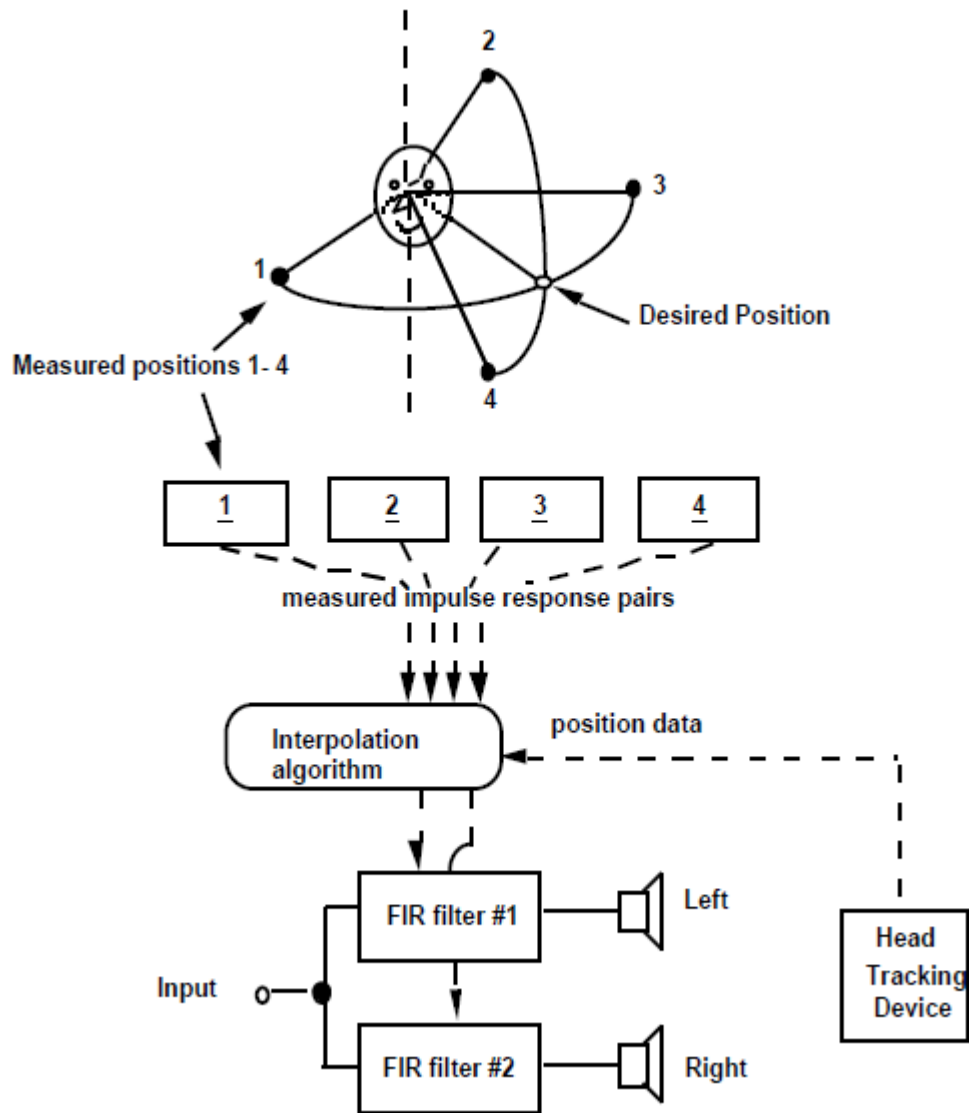


62. These “trackers” were for tracking the head movement of the listener, to adjust the binaural sound delivered to the listener. This was done so that the perceived location of the sound source remains fixed in virtual space, regardless of the listener’s head movement:

“If one of the perceptual requirements is to have virtual sound images remain fixed in virtual space independent of where one’s head moves, a six degree-of-freedom tracker system is necessary, consisting of a source, sensor, and a hardware unit housing power supply, electronics, and RS-232 interface to send positional data to a host computer.”

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

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



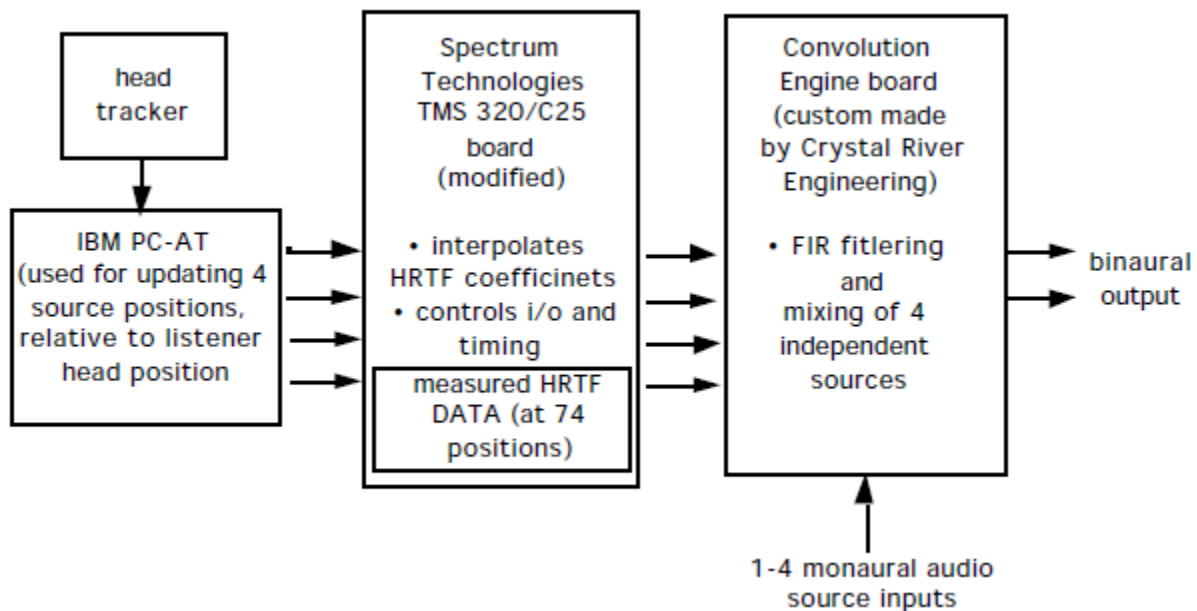
64. It was commonly known that the benefits of such systems was that the perceived location of the sound source can remain fixed in virtual space, regardless of the listener's head movement. Doing so would mimic spatial hearing in natural environments and enhance the immersivity of the environment:

“In virtual reality, the key to immersivity is directly tied to the success of the interface in providing realistic information at the effectors and the degree to

which sensors are able to track human interaction. For this reason, head-tracked virtual acoustic systems represent the best approach to implementing sound, so that **sound sources can remain fixed in virtual space independent of head movement**, as they are in natural spatial hearing.”

Ex. 1036, 167 (emphasis added).

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



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

B. The '628 Patent

67. The '628 patent describes providing sound to listeners as "if they were communicating face-to-face with each other." Ex. 1001, 1:15-16. In my opinion, these concepts were well-known in 2014, particularly in view of the development of wearable technology, spatialization cues, binaural sound synthesis, HRTFs, and virtual reality head-tracking systems discussed above. The components of the patent were conventional and operate in a conventional manner.

68. The '628 patent claims embodiments in which “one or more electronic devices in an electronic system provides binaural sound of a voice of a person to localize at an image during a telephone call.” *Id.*, Abstract, Fig. 9 (reproduced below).

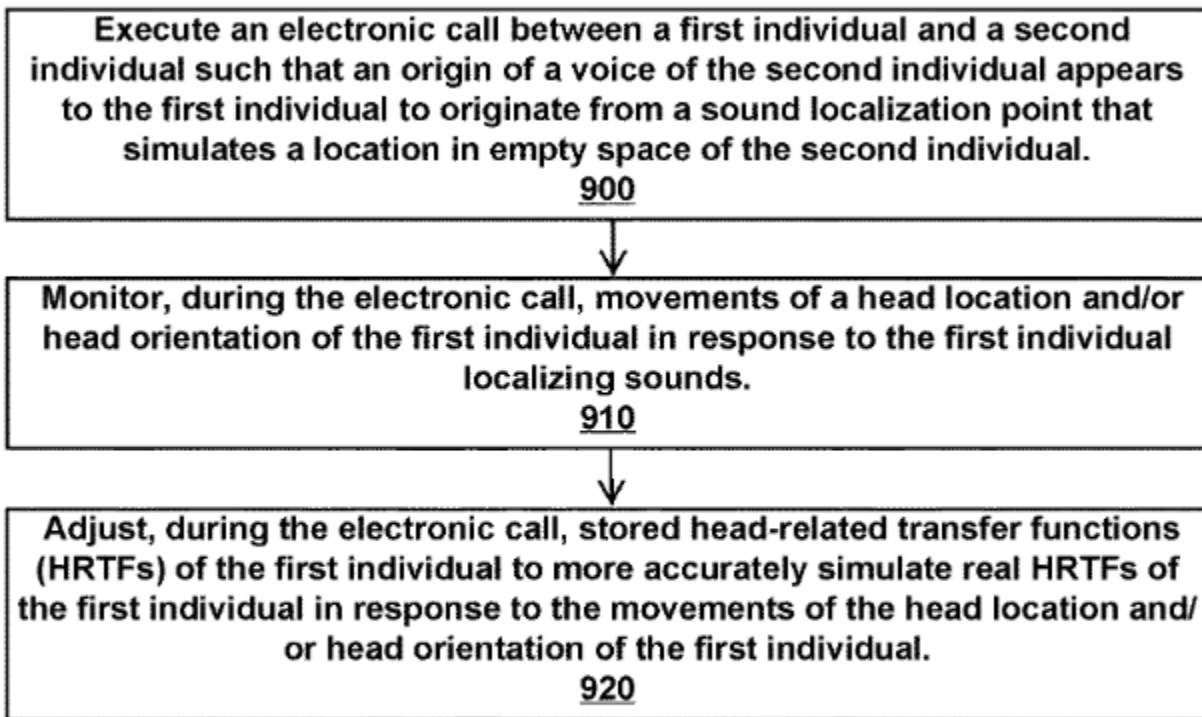


Figure 9

69. In my opinion, there is nothing unconventional about the claimed components, their configuration, or how they operate. As the patent describes, it is during “an electronic call between a first individual [e.g., called party] and a second individual [e.g., calling party]” that the “voice of the [calling party] appears to the [called party] to originate from a sound localization point that simulates a location

in empty space of the [calling party].” *Id.*, 21:36-40. According to an example in the patent, “a listener sits at his kitchen table and has an electronic telephone call with his mother.” *Id.*, 5:37-38. “The listener designates an empty kitchen chair next to him as a sound localization point for the voice of his mother.” *Id.*, 5:43-45. Then, “the voice of the mother appears to originate from the empty kitchen chair such that the listener localizes her voice to this location.” *Id.*, 5:45-47.

70. As the patent further describes, the listener is wearing a wearable device that tracks the listener’s head movements relative to the SLP. *Id.*, 21:51-54. “For instance, a determination is made as to whether and/or when the head turns left or right, turns up or down, or rotates back and forth” compared to the SLP. *Id.*, 21:57-66. The voice of the calling party is processed in response to these head movements so that the called party continues to localize the calling party’s voice at the SLP.

71. The listener’s wearable device also includes “a display that provides an image or video of the speaker 2010 in the field of view 2080 of the listener” at the sound localization point. *Id.*, 33:22-24. For example, an image of the mother will appear at the kitchen chair.

72. Despite these descriptions in the ’628 patent, neither the specification nor the file history describe how these teachings improve upon the prior art. Notably, as I discuss below, my review of the prosecution history indicates that the Examiner allowed the wrong claims, and it appears that the Examiner did not

consider the issued claims against any prior art. *See* Section VI.C (Prosecution History).

C. Prosecution History

73. I have reviewed the prosecution history of the '628 patent, and based on the prosecution history, the Examiner appears to have mistakenly allowed the wrong claims. I am unaware of any definitive evidence that the Office ever examined the issued claims of the '628 patent.

74. According to the prosecution history, the application leading to the '628 patent was filed on February 23, 2017, with twenty original claims. Ex. 1002, 1-116. During prosecution, the examiner issued no rejection. *See generally id.* On April 17, 2017, the Applicant cancelled original claims 1-20, and added new claims 21-40. *Id.*, 127-39.

75. On April 26, 2017, the Examiner issued a Notice of Allowance, allowing the cancelled claims 1-20 but did not mention the pending new claims 21-40. *Id.*, 149.

76. The Examiner's statement of reasons for allowance referred solely to the cancelled claims 1, 8 and 15 (and not the pending claims 21-40) as requiring: "... providing, with an electronic earphone located at a head of the listening person the binaural sound to the listening person during the telephone call such that the

origin the binaural sound for the listening person occurs at the sound localization point in e [sic] space that is away from and proximate to the listening person.” *Id.*

77. The examiner also stated in the reasons for allowance that it was “These features, in combination with the other features of Claims 1, 8 and 15 [that] are not anticipated by, nor made obvious over, the prior art of record.” *Id.* However, the exact limitation referenced in the examiner’s statement for reasons for allowance is not recited in the new claims 21-40 that were actually pending at the time of the allowance.

78. The ’628 patent issued on June 6, 2017, with pending claims 21-40 renumbered as issued claims 1-20.

D. Challenged Claims of the ’628 Patent

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

80. The ’628 patent has 20 claims, 3 independent claims and 17 dependent claims. Independent claim 9 is a method claim that recites the following four steps during a call:

- (1) processing the calling party’s voice with HRTFs so it localizes to the called party as binaural sound at an SLP that is one to two meters away;
- (2) displaying with a wearable electronic device an image representing the calling party at the SLP;

- (3) tracking the called party's head movements; and
- (4) changing the HRTFs so the voice of the calling party continues to localize to the SLP while the called party's head moves with respect to the image.

81. Independent claim 16 recites a non-transitory computer readable storage medium storing instructions that one or more electronic devices in an electronic system execute as a method, that is materially similar to claim 9. Dependent claims 10, 11, and 17 recite additional limitations relating to tracking the calling party's head movements. Dependent claims 12-14 and 19 recite additional limitations relating to the image representing the calling party. Dependent claims 18 and 20 recite additional limitations relating to a virtual microphone point.

E. Effective Priority Date

82. I understand that the '628 patent was filed on February 23, 2017. I understand that cover of the '628 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 '628 patent are entitled to a priority date of June 23, 2014.

VII. CLAIM CONSTRUCTION

83. I understand that Patent Owner contends certain claim elements in the '628 patent should have their plain and ordinary meaning. Specifically, I understand

that 8KHZ's position is that if construed, the below terms in the '628 patent should have the following meanings:

- *telephone call* in cls. 9, 12, 16 should mean “voice communication over an electronic network between at least two parties”
- *HRTFs* in cls. 9, 12, 16 should mean “functions that describe transformation of sound that modify the spectrum and time of arrival at a listener's ears”
- *wearable electronic device / WED* in cls. 9, 12, 16 should mean “portable wireless electronic device that is worn on or attached to a person”

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

VIII. THE PRIOR ART

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

A. Billinghamurst

85. The paper titled “A Wearable Spatial Conference Space” was authored by Professor Mark Billinghurst et al. and published by IEEE in October 1998 (“Billinghurst”) and is prior art under at least 35 U.S.C. §§102(a)(1) and (a)(2). In my opinion, Billinghurst discloses each limitation of the independent challenged claims of the ’628 patent, with the exception that Billinghurst does not expressly disclose the use of HRTFs. In my review of the prosecution history, Billinghurst was not cited or discussed during the ’628 patent prosecution.

86. Similar to the systems and devices that I discussed above, Billinghurst describes an immersive virtual experience that provides sounds that users located remotely from one another would hear if they were communicating face-to-face in person with one another. Ex. 1010, 2 (“In collaborative virtual environments (CVEs) spatial visual and audio cues can combine in natural ways to aid communication.”).

87. As Billinghurst explains, “[t]he result [of wearables that provide spatialized 3D graphics and audio cues to aid communication] is a wearable augmented reality communication space with audio enabled avatars of the remote collaborators surrounding the user. The user can use natural head motions to attend to the remote collaborators, can communicate freely while being aware of other side conversations and can move through the communication space.” *Id.*, 1.

88. In this manner, Billinghamurst is similar to the wearable technology discussed in the above Background section, as Billinghamurst describes, “wearables can be used to provide spatialized 3D graphics and audio cues to aid communication” by creating “a wearable augmented reality communication space with audio enabled avatars of the remote collaborators surrounding the users.” *Id.*, 1.

89. Billinghamurst is also similar to the '628 patent in that during a call between remote users, the voice of the calling party appears to the called party to originate from a sound localization point that simulates a location in empty space of the calling party. *Id.*, 5. In Billinghamurst, “each user determines the position and orientation of their own avatar in space, which changes as they move or look about [the] environment.” *Id.* Billinghamurst explains that “[t]he well known ‘cocktail party’ effect shows that people can easily monitor several spatialized audio streams at once, selectively focusing on those of interest [22], [23].” *Id.*, 2-3.

90. Billinghamurst’s wearable device also includes a head tracking device that tracks the head movements of the listener. “The wearable user has their head tracked so they can simply turn to face the speakers they are interested in.” *Id.*, 5. The audio of the user “is spatialized according to the distance and direction between speaker and listener.” *Id.* This means that “[a]s users face or move closer to different speakers the speaker volume changes due to the sound spatialisation.” *Id.*

91. In addition to providing spatialized audio based on the wearer's head orientation and movement, the Billingham wearable device also includes a display that provides an image or video of the speaker in the field of view of the listener. *Id.*, 4. According to Billingham, "[w]ith this display configuration a wearable conference space could be created that allows remote collaborators [to] appear as virtual avatars distributed about the user (figure 2.0)." *Id.* "Since the displays are see-through or see-around the user could also see the real world at the same time, enabling the remote collaborators to help them with real world tasks." *Id.* "These remote users may also be using wearable computers and head mounted displays." *Id.*

B. Schröder

92. I understand that International Publication No. WO 2005/032209 to Schröder ("Schröder") was filed on September 15, 2004 and published on April 7, 2005 and is prior art under at least 35 U.S.C. §§102(a)(1) and (a)(2). In my opinion, Schröder discloses the use of HRTFs to localize binaural sound at an SLP, which is missing from Billingham. In my review of the prosecution history, Schröder was not cited or discussed during the '628 patent prosecution.

93. Schröder is similar to the '628 patent, Billingham, and the wearable devices discussed above in the Background section that produce binaural sound. Schröder describes reproducing binaural sound using headphones. Ex. 1012, 1:5-6

(“the binaural reproduction of aural events using headphones.”). Specifically, Schröder’s headphones include head-tracking systems that “detect the rotational movements of the head and are used to keep the position of a virtual sound source outside the head constant.” *Id.*, 1:7-10. In Schröder, the sound appears to remain spatially fixed even when “the listener turns his head.” *Id.*, 5:27-30. To implement the sound’s “constant spatial direction,” Schröder’s system changes the HRTF filter that is applied. *Id.*, 5:30-35 (“[t]he signals for the left and right ears are adjusted, on the basis of the position of the listener or of his head, using variable filters for the HRT function such that the direction from which the aural event appears to come remains stable in space.”), *see also id.*, 13:8-11.

C. Zhou

94. The paper titled “An experimental study on the role of software synthesized 3D sound in augmented reality environments” was authored by Professor Zhiying Zhou et al. and published in *Interacting with Computers* in 2004 (“Zhou”) and is prior art under at least 35 U.S.C. §§102(a)(1) and (a)(2). In my opinion, Zhou discloses using HRTFs to provide binaural sound that localizes the sound *at an SLP one or two meters away*. In my review of the prosecution history, Zhou was not cited or discussed during the ’628 patent prosecution.

95. Zhou is similar to the ’628 patent, Billinghamurst, and the wearable devices discussed above in the Background section that produce binaural sound in

virtual reality environments. Zhou describes 3D sound in augmented reality environments, which it concludes “contributes significantly to the feeling of ‘human presence and collaboration’ and helps the subjects to ‘identify spatial objects.’” Ex. 1018, 1-2. Zhou explains a series of experiments, where “[i]n the first experiment the sound sources are always placed around 1.5 m away in the front of the listener.” *Id.*, 5. Zhou teaches that at this distance, “the addition of 3D sound can be a very effective way to complement the user experiences in AR environments.” *Id.*, 26. Zhou further found that not only does 3D sound help with object identification and human-human interactions, but it also “helps depth perception” and “enables a more realistic environment and more immersive feeling of being inside a both visual and auditory augmented reality world.” *Id.*, 26.

D. Haddick

96. I understand that International Publication No. WO 2013/49249 to Haddick (“Haddick”) was filed on September 26, 2012 and published on April 4, 2013 and is prior art under at least 35 U.S.C. §§102(a)(1) and (a)(2). In my opinion, Haddick discloses a wearable head mounted display device that provides both visual and audio input for live communication with one or more other parties over a *telephone call*. In my review of the prosecution history, Haddick was not cited or discussed during the ’628 patent prosecution.

97. Haddick is similar to the '628 patent, Billingham, and the HMD devices discussed in the Background section in that it describes a wearable device that provides both visual and audio sensory input for live communication with one or more other parties in a videoconference and that can “reproduce [a] 3D sound profile.” Ex. 1011, ¶653. Haddick’s “wearer with headphones, hearing loss, in a loud environment, and the like, may not otherwise be able to tell what’s being said and/or the direction of the person speaking, but is provided with a 3D sound enhancement from the eyepiece.” *Id.*

E. Ballas

98. I understand that U.S. Patent Pub. No. 2003/0059070 to Ballas (“Ballas”) was published on March 27, 2003 and is prior art under at least 35 U.S.C. §§102(a)(1) and (a)(2). In my opinion, Ballas discloses the use of HRTFs to localize binaural sound at an SLP missing from Billingham. In my review of the prosecution history, Ballas was not cited or discussed during the '628 patent prosecution.

99. Ballas is similar to the '628 patent, Billingham and the wearable head tracking systems described in the Background section. Ballas describes “[s]patialized audio [that] is sound that is processed to give the listener an impression of a sound source within a three-dimensional environment” and provides “a more realistic experience.” Ex. 1008, ¶2. Just like in Billingham, Ballas’s wearable devices (e.g., an eyeglass frame or helmet with mounted speakers) are “mounted on

the user's head" and produce spatialized audio to the user and further "allow[s] the listener to change locations and head orientation without degradation of the audio signal." *Id.*, ¶¶17-18, 25, cl. 1, 7, 13. Ballas provides additional detail regarding how the spatialized audio is generated, using HRTFs. *Id.*, ¶¶17-18, 25-26 (describing how the wearable device includes a head tracking system that 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.") According to Ballas, by using HRTF filters, the system is able to "maintain[] the virtual location of the audio signals/sound" and "[t]he source of the sound is kept on axis with [the] user's ear regardless of the head movement." *Id.*, Abstract, ¶17.

IX. GROUND 1: BILLINGHURST IN VIEW OF SCHRÖDER RENDERS OBVIOUS CLAIMS 9-14 AND 16-20

A. Motivation to Combine Billinghamurst with Schröder

100. In my opinion, it would have been obvious to a POSA to modify the CPU in Billinghamurst to process audio streams using the HRTFs taught by Schröder. For the following reasons, it is my opinion that a POSA would have been motivated to combine the teachings of Billinghamurst and Schröder. A POSA would have recognized that Billinghamurst's audio streams would have been spatialized using the conventional HRTF techniques described in Schröder.

101. As I explained above in the Background section, it was commonly known to synthesize binaural sound with left and right HRTFs for the desired placement angle and position. Processing sounds with HRTFs to create a virtual auditory space was common knowledge, and at the relevant time period, HRTFs had been known for decades and were readily available to anyone wishing to produce synthetic binaural sound. Accordingly, at the relevant time period, a POSA would have readily understood that HRTFs was the most common technique to spatialize 3D sound. *See, e.g.*, Ex. 1018, 3 (“[e]xact ***spatial 3D sound*** in virtual and augmented environments requires perceptual externalization of the sound source, which can only be brought about using ***head related transfer functions (HRTFs)***.”). Thus, in my opinion, although Billinghamurst does not expressly describe in detail the exact spatialization techniques used to localize the audio stream of the remote collaborators to their respective avatars, a POSA would have understood that these details were described in HRTF techniques, including those described in Schröder.

102. Billinghamurst’s primary goal is to support collaboration between remote users, by providing an immersive virtual experience with sounds that the remote users would hear if they were communicating face-to-face in person with one another. Ex. 1010, 1 (“The problem we are interested in is how a wearable computer can be used to support collaboration between multiple people.”). To provide such an immersive virtual experience for a teleconference, Billinghamurst describes (1)

visual and audio enhancements to aid communication between remote users, (2) how to create a collaborative communications space between remote users, and (3) how to represent remote users in such a space, all with a focus on spatialization. *Id.*

103. According to Billinghamurst, and as would have been appreciated by a POSA, spatialization is key to creating an immersive virtual environment for communication and collaboration because it emulates face-to-face interaction. Ex. 1010, 2 (“In collaborative virtual environments (CVEs) spatial visual and audio cues can combine in natural ways to aid communication.”). In Billinghamurst, creating such an environment need not be complicated, as “[e]ven a simple virtual avatar representation and spatial audio model enables users to discriminate between multiple speakers.” *Id.*, 3. Yet Billinghamurst does not describe with much detail precisely how such spatialization would have been implemented, most likely because such techniques including HRTFs were well-known in the field. Therefore, in my opinion, a POSA implementing Billinghamurst’s wearable teleconferencing system would have looked to other references, such as Schröder, to spatialize Billinghamurst’s audio streams provided through headphones, so as to spatialize the audio streams in real time. Combined with Billinghamurst’s head-tracking system and “body-stabilized” or “world-stabilized” configurations, this implementation would have ensured that the spatialized audio streams appear as if they emit from the corresponding avatar even when the users move.

104. Similar to Billingham, Schröder shares this goal of spatialization so that a sound source provided over head-tracking headphones is perceived to remain fixed in space even when the listener moves. Ex. 1012, 5:10-13 (explaining its objective is “to specify an arrangement for locating aural events such that they have a constant spatial direction using headphones.”).

105. Indeed, Schröder’s goal is to provide an immersive experience, so that the user “perceive[s] aural events as having a constant spatial direction on average, i.e. as coming from one preferred direction on average in virtually stationary fashion, even when the listener is moving, without restricting the ability to locate individual sound sources in the aural event.” *Id.*, 5:5-10.

106. Similar to Billingham, Schröder’s wearable device includes a head tracking system and alters the sound output over the headphones to provide a stable virtual position. *Id.*, 4:13-17 (“rotational movements of the head are therefore evaluated and the signal for the left and right ears are altered such that a stable virtual position for sound sources in space becomes established.”). Schröder provides additional detail regarding how this is implemented, by adjusting HRTFs. Specifically, Schröder describes “[t]he signals for the left and right ears are adjusted, on the basis of the position of the listener or of his head, using variable filters for the HRT function such that the direction from which the aural event appears to come remains stable in space, that is to say has a constant spatial direction.” *Id.*, 5:30-36.

Thus, as the POSA would have recognized upon reading Billinghamurst and Schröder, using HRTFs would have advanced the common spatialization goal of both references, and further advanced Billinghamurst's goal for immersive and realistic collaboration.

B. Reasonable Expectation of Success

107. In my opinion, a POSA would have had a reasonable expectation of success in combining the teachings of Billinghamurst and Schröder.

108. Both systems included wearable devices having processors that adjust and process audio streams that are output through headphones. In Billinghamurst, that wearable computer "is a custom built 586 PC 104 based computer with 20mb of RAM running Windows 95." Ex. 1010, 4. Using the computer's processor, Billinghamurst describes that the audio stream "is spatialized according to the distance and direction between speaker and listener." *Id.*, 5. This means that "the speaker volume changes due to the sound spatialisation." *Id.*

109. Schröder's system is very similar to the Billinghamurst system. Just as in Billinghamurst, in Schröder's system, "[a] listener 1 is wearing headphones 11 for binaural reproduction of aural events." Ex. 1012, 11:22-24. Schröder provides additional detail regarding how those aural events are reproduced binaurally, using a "control circuit containing variable HRTF filters," "the audio signal for the left and right ears of the listener 1" are adjusted "such that the aural event is perceived as

coming unchanged from direction 6” when “the listener’s head is turned.” *Id.*, 12:3-4, 12:20-24. Specifically, Schröder’s system includes a “control circuit HRTF [that] outputs signals for the left and right channels which have been altered... such that it is possible to perceive an aural event to have a stable spatial direction.” *Id.*, 15:17-21.

110. Accordingly, it is my opinion that a POSA would have recognized that Billinghamurst and Schröder are structurally and functionally similar. Both systems use a processor to adjust an audio stream to provide spatialized binaural sound over headphones. Ex. 1010, 5; Ex. 1012, 15:17-21. Each system further includes a head tracking system that tracks the head orientation, position, and movement of the listener, and adjusts the audio stream to maintain the sound localization point so that it appears fixed in space, despite the listener’s head movement. Ex. 1010, 4-5; Ex. 1012, 12:3-24. While Billinghamurst does not provide explicit details regarding how the audio stream is spatialized, it would have been well-known to implement spatialized audio streams with HRTFs, and Schröder expressly provides this detail. In my opinion, the POSA implementing the Billinghamurst system would have therefore recognized that Billinghamurst’s spatialized sound would be achieved by programming the Billinghamurst processor with Schröder’s HRTF functionality, to process the audio stream to localize it to a specific perceived location.

111. In my opinion, this programming involves a simple and obvious implementation to Billinghurst's system. Such an implementation would have merely involved a combination of known prior art elements (sound processor as included in Billinghurst's wearable device) according to known methods and techniques (HRTF filters as taught by Schröder) to yield predictable results (spatialized sound described in both references), and involves use of a known technique (spatialization) to improve a similar device (a wearable communication device described in both references) in the same way (filtering sounds to introduce spatial cues).

C. Independent Claim 9

1. **9[pre]:** *“A method executed by one or more electronic devices in an electronic system that improves an experience of a calling party and a called party talking to each other during a telephone call, the method comprising:”*

112. To the extent the preamble of claim 9 is limiting, it is my opinion that Billinghurst discloses the preamble.

113. In Billinghurst, a set of wearable hardware *execute[s]* a *method* for implementing a “wearable spatial conferencing space.” Ex. 1010, Title. Billinghurst's wearable hardware (*one or more electronic devices*) is included in a videoconferencing system (*an electronic system*) to execute the *method*. Billinghurst's Fig. 3.0 (which I reproduced below) is an image of a user wearing and

using that “wearable hardware,” which includes a “wearable computer,” a display having a head tracking system, and a sound source. Specifically, the display of the wearable hardware takes the form of “a pair of Virtual i-O iglasses! converted into a monoscopic display by the removal of the left eyepiece,” and its head tracking system takes the form of an “inclinometer and a magnetometer used as a three degree of freedom orientation tracker.” *Id.*, 4-5. The sound source of the wearable hardware takes the form of “a soundBlaster compatible soundboard with headmounted microphone.” *Id.*



Figure 3.0 The Wearable Hardware.

114. As Billinghamurst explains, the wearable hardware provides a “wearable conferencing space... [that] allows remote collaborators [to] appear as virtual avatars distributed about the user (figure 2.0). The avatars could be live video streams and as they speak their audio streams [are] spatialized in real time so that they appear to emit from the corresponding avatar.” *Id.*, 4, Fig. 2.0 (reproduced

below). In this manner, Billinghurst's wearable hardware provides an immersive virtual videoconferencing environment in which users talk to each other and audio streams and live video stream avatars of the conference participants appear co-located to the user.

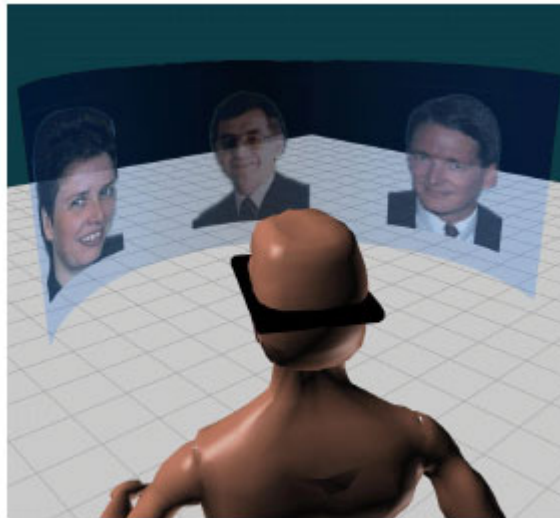


Figure 2.0 A Spatial Conferencing Space.

115. In my opinion, in Billinghurst's the wearable hardware (*one or more electronic devices*) works together as part of a videoconferencing system (*electronic system*) to execute a *method* that *improves an experience* for videoconference users, where those users include *a calling party and a called party talking to each other*. As Billinghurst explains, the wearable hardware and videoconferencing system provide an environment that is “[j]ust as in face-to-face collaboration, [where] users could turn to face the collaborators they wanted to talk to while still being aware of the other conversations taking place.” *Id.*

116. The Billingham system also *improves an experience* for videoconference participants, when compared to conventional videoconferencing environments. For example, compared to conventional videoconferences where the users are static and have undefined or fixed spatial relationships with one another, the POSA would have readily appreciated that one such improvement in Billingham includes allowing “[t]he user [to] move about the space enabling them to choose their own viewpoint and the spatial relationships between the collaborators.” *Id.* As Billingham explains, another important improvement is that, compared to conventional systems where the sizes of the remote avatars is small and dependent on the size of the screen, Billingham’s system “allow[s] the faces of remote users to appear life-size, a crucial factor for establishing equal relationships in remote collaboration.” *Id.*

117. The POSA would have readily appreciated that Billingham teaches other *improvements to the experience* for videoconference participants in Billingham’s *electronic system*. Such *improvements* include easy discrimination when speakers talk at the same time, participant preferences to rely on non-verbal cues, allowing users to multi-task while participating in the conference, and ease-of-use. *Id.*, 6 ((1) “[u]sers are able to easily discriminate between three simultaneous speakers when their audio streams are spatialized,” (2) “[p]articipants preferred seeing a visual representation of their collaborators as opposed to just hearing their

speech,” which “allowed them to use some of the non-verbal cues commonly used in face-to-face communication such as gaze modulation and body motion,” (3) users “could continue doing real world tasks while talking to collaborators in the conferencing space,” and (4) “[t]he interface is easy and intuitive to use,” and “the head tracking on the wearable was easier than the mouse only desktop interface.”)

118. Billinghamhurst further specifically contemplates a *telephone call*. Ex. 1010, 2 (describing telephones incorporating computing power and conference phone calls), *see also id.*, 5 (describing how the wearable interface supports “3D spatialized Internet telephony”). I understand that Patent Owner’s position is that a *telephone call* should mean “voice communication over an electronic network between at least two parties.” Under that interpretation, Billinghamhurst’s videoconferencing system provides a *telephone call* because it provides voice communication over the internet amongst multiple parties. In addition, to the extent that it is Patent Owner’s position that two parties participating in a conference call would meet the limitations *calling party* and *called party*, it is my opinion that the user of Billinghamhurst’s wearable hardware or any of the remote collaborators would have been the *calling party* or *called party*. Specifically, the ’628 patent does not refer to “calling party” or “called party” anywhere other than in its abstract and in the claim language. Nor does the patent refer to either party as having to perform

any step other than those that are claimed, which do not require any party to call another party, and also do not require any party to accept a call from another party.

119. In addition, it is my opinion that Billinghurst even contemplates a *telephone call* under a narrower interpretation requiring a “call between at least two parties using at least one phone.” As Billinghurst recognized in 1998, the technology in telephones and portable computers were merging such that “telephones incorporate more computing power and portable computers become more like telephones... [and] a conference phone call may be just as effective as [computer-mediated communication].” Ex. 1010, 2. Therefore, it is my opinion that even if *telephone call* were to be interpreted to require a call using a phone, the POSA would have recognized that Billinghurst’s teachings would have applied to a *telephone call*. Moreover, the POSA would have known that conventional phone calls take place between a *calling party* (which the POSA would have recognized as any of the users in Billinghurst’s videoconferencing system, but for the purposes of this declaration, is mapped to Billinghurst’s remote collaborator) and a *called party* (which the POSA similarly would have recognized as any of the users in Billinghurst’s videoconferencing system, but for the purposes of this declaration, is mapped to Billinghurst’s user or wearer of the wearable hardware). In my opinion, implementing Billinghurst’s wearable system during a call using a phone between a calling party and a called party would have been a simple, natural, and

straightforward extension of Billingham's teachings. This implementation would have improved a user's experience of a conventional telephone call, as it would have provided spatial auditory cues to the listener, making it easier for the listener to distinguish amongst the remote collaborators, and improved the listener's experience by making it feel more realistic and natural.

120. For these reasons, the POSA would have appreciated that Billingham discloses *a calling party and a called party talking to each other during a telephone call*.

121. Therefore, it is my opinion that Billingham discloses a *method executed by wearable hardware (one or more electronic devices) in a videoconferencing system (an electronic system) that provides an easy and intuitive interface that improves remote collaboration (that improves the experience) between users of the Billingham system (of a calling party and a called party talking to each other during a telephone call)*.

2. **9[a]:** *"processing, during the telephone call and with a processor, a voice of the calling party with head related transfer functions (HRTFs) so the voice of the calling party localizes to the called party as binaural sound at a sound localization point (SLP) that is one meter to two meters away from the called party;"*

122. In my opinion, Billingham in view of Schröder discloses this limitation.

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

- a) *Billinghurst discloses this limitation but is not explicit it uses HRTFs*

123. Billinghurst describes “a wearable conferencing space... that allows remote collaborators [to] appear as virtual avatars distributed about the user (figure 2.0).” Ex. 1010, 4. In that wearable conferencing space, “the avatars could be live video streams and as they speak their audio streams [are] spatialized in real time so that they appear to emit from the corresponding avatar.” *Id.* In this manner, Billinghurst explains the auditory and positional relationships between the called and calling parties. As I explained above for 9[pre], the POSA would have understood that the users of Billinghurst’s conferencing space (which is produced during a *telephone call* under Patent Owner’s interpretation) include a *calling party* and a *called party*, where the wearer or user of the head mounted display in Fig. 2.0 would be the *called party* and each remote collaborator would be a *calling party*. See Section IX.C.1 (Ground 1, 9[pre]).

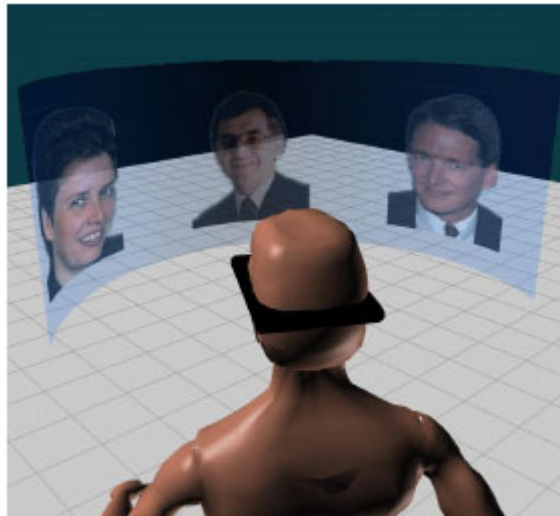


Figure 2.0 A Spatial Conferencing Space.

124. Further, a POSA would have understood that each audio stream in Billinghamhurst's conferencing space is a *voice* from a remote collaborator. That remote collaborator is represented in the virtual conferencing space by the virtual avatar. As the POSA would have readily known, conferencing spaces commonly presented voices of remote conference participants. In Billinghamhurst, that *voice* of the remote collaborator is provided as *binaural sound* to the user, and *localizes* to the user *at a sound localization point (SLP)* that corresponds to the perceived position of that remote collaborator's avatar in the virtual conferencing space. *See, e.g., id.*, 4, Fig. 2.0. I note that 8KHZ's expert in the parallel litigation admitted it was known to provide binaural sound to listeners through earphones before 2014. Ex. 1032, 176:3-5.

125. As Billinghamhurst explicitly states, "current wearable computers use see-through or see-around monoscopic head mounted displays with stereo headphones."

Id., 3. “Each user determines the position and orientation of their own avatar in space.” *Id.*, 5.

126. In my opinion, the POSA would have understood that when users (including the user or wearer of the HMD and the remote collaborators) connect to the Billinghurst system, “their audio is broadcast to all the other users in the space” using the headphones, such that the audio (corresponding to their *voice*) “is spatialized according to the distance and direction between speaker [*calling party*] and listener [*called party*]” as *binaural sound* that localizes the *calling party* at the avatar’s position (*sound localization point*). *Id.*

127. I note that Billinghurst’s system has remote users that connect to the conferencing space, and the POSA would have understood that any of the users are speakers or listeners. For clarity, I rely on Billinghurst’s “listener,” the wearer of the wearable hardware, as the *called party* and the “speaker,” the remote collaborator, as the *calling party*, though the POSA would have understood that any of the users or remote collaborators in Billinghurst’s system can be the listener or speaker at any time during the videoconference.

128. The POSA would have readily understood from reading Billinghurst that Billinghurst’s system has remote users that connect to the conferencing space, where any of the users are speakers or listeners. For clarity, and for the purposes of this declaration, I rely on Billinghurst’s “listener,” the wearer of the wearable

hardware, as the *called party* and the “speaker,” the remote collaborator, as the *calling party*, though it is my opinion that the POSA would have understood that any of the users or remote collaborators in Billinghamhurst’s system can be the listener or speaker at any time during the videoconference.

129. In my opinion, the POSA would have also readily understood that Billinghamhurst discloses processing the audio stream of the remote collaborator (*voice of the calling party*) with a CPU (*with a processor*). In Billinghamhurst, the wearable hardware includes a wearable computer, which is “a custom built 586 PC 104 computer with 20mb of RAM running Windows 95.” *Id.*, 4. When the audio is broadcast in Billinghamhurst, “culling is also used so that only the audio streams from the speakers closest to the listener are broadcast and spatialized,” and this culling “significantly reduces the CPU load.” *Id.*, 5. In view of these teachings, the POSA would have understood that Billinghamhurst’s CPU in the computer (*processor*) processes the audio stream of the speaker (*voice of the calling party*), for example, to perform such culling.

130. Thus, it is my opinion that Billinghamhurst discloses *processing, during telephone call and with a CPU (processor)*, an audio stream of a remote collaborator (*a voice of the calling party*) so the audio stream of the remote collaborator (*voice of the calling party*) *localizes to the* listener (*called party*) as spatialized sound (*binaural sound*) at the avatar’s position (*a sound localization point*).

131. It is my opinion that this limitation would have been obvious over Billinghamurst in view of Schröder because Schröder described in detail converting propagation time, amplitude, phase and frequency into HRTFs (Ex. 1012, 2:3-12). I note that PO's expert in the parallel litigation even admitted that processing with HRTFs was well-known. Ex. 1032, 59:4-16 (there were "many references" from the 1990s that "discuss[ed] using convolution processes to process head related transfer functions with digital signal processors for binaural audio"), 135:6-25 (it was known before 2014 to "process[] sound with HRTFs," to "process[] sound to move virtual sound images to different locations," and "generat[e] binaural sound through digital signal processors processing HRTFs"), 196:12-197:4. It is my opinion that Schröder's disclosure of HRTFs disclose "functions that describe transformation of sound that modify the spectrum and time of arrival at a listener's ears" under 8KHZ's position.

132. While Billinghamurst does not explicitly disclose that the voice is processed *with head related transfer functions (HRTFs)* to localize the voice at a sound localization point, Billinghamurst does disclose that the "audio is implemented utilizing Microsoft's Direct Sound technology" and that the "conferencing space uses custom developed telephony libraries that incorporate the Microsoft Direct Sound libraries." *Id.*, 6-7. As the POSA would have known, Microsoft released DirectSound as a set of "sound standards" implemented in the 1990s, where it was

a collection of software libraries that was a part of Windows 95. Ex. 1043, 175. A key feature of DirectSound was 3D sound (called DirectSound3D or DS3D) that was based on HRTFs for positioning of sound sources in 3D space. *Id.* Thus, while Billinghamurst itself does not refer to HRTFs, the POSA would have understood that Billinghamurst's use of DirectSound to implement the spatialized audio signals would have included processing *with HRTFs* to localize the voice at an SLP.

133. Nonetheless, processing *with HRTFs* to localize the voice at an SLP would have been obvious to the POSA from the teachings of Billinghamurst in view of Schröder.

b) *Schröder discloses processing with HRTFs to localize at an SLP*

134. Schröder is generally related to “the binaural reproduction of aural events using headphones.” Ex. 1012, 1:5-6. Specifically, in Schröder's system, “headphones [] detect the rotational movements of the head and are used to keep the position of a virtual sound source outside the head constant.” *Id.*, 1:7-10. As Schröder recognizes, conventional headphones that provide the same signal to both ears give rise to “the problem of in-head localization.” *Id.*, 1:12-14. The POSA would have understood that in-head localization means that, as Schröder explains, “sound sources [are] mapped virtually inside the head.” *Id.*, 1:14-19. As the POSA would have understood, in-head localization is undesirable because mapping sound

sources virtually inside the head is not realistic and does not provide an immersive virtual experience, particularly in a videoconference call setting. To be more realistic, Schröder teaches mapping the sound sources to a location outside the head: “sound sources would be mapped on a virtual stage in front of the listener [as] if the aural events were reproduced on loudspeakers situated in front of the listener.” *Id.*

135. Schröder acknowledges that in-head localization is something to be avoided, and teaches processing the signal to provide a more realistic, immersive experience: “it is possible to process the signal...such that the reproduction on the headphones corresponds to the reproduction of the same signal on a stereo loudspeaker arrangement.” *Id.*, 2:32-37. To achieve reproduction of a stereo loudspeaker arrangement over headphones, Schröder describes with respect to Fig. 5 (which I have reproduced below, *see also* Fig. 7) that headphones 8, 9 “supply the left and right ears of the listener 1 with not only the signals L, R...but also portions of the signals L(HRTF), R(HRTF),” which “are produced by applying an HRTF to the respective original signal L, R.” *Id.*, 3:5-14. In providing left and right ears with these HRTF-processed signals, Schröder is able to avoid in-head localization so that “[t]he listener perceives the sound event as coming from outside the head, illustrated in the figure by arrow 6.” *Id.*, 3:14-15. In Schröder, there are multiple *HRTFs*, including “variable HRTF filters” that Schröder applies to the sound in order to

adjust the perceived direction and position of the aural event. *Id.*, 7:13-16, 12:20-24, 13:6-11, 14:16-21.

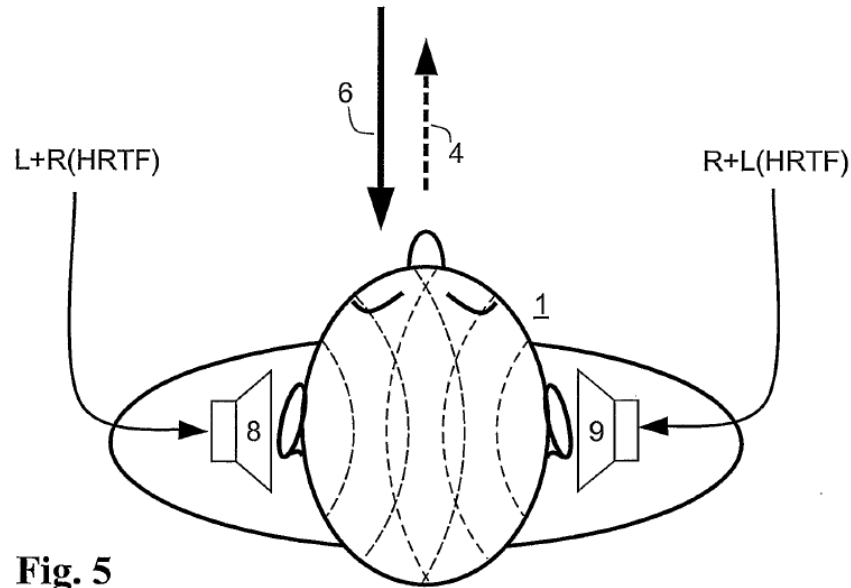


Fig. 5

136. In my opinion, Schröder’s “applying an HRTF” to the respective left and right signals is *processing, ...with a processor*, a sound signal with multiple *HRTFs* (such as one left HRTF and one right HRTF, which Schröder calls variable HRTF filters). The result of this processing is that the sound signal *localizes to the* listener (*called party*) from a specific direction and at a perceived position of the aural event (*at a sound localization point*). Schröder specifically teaches that a “control circuit containing variable HRTF filters adjusts the audio signals for the left and right ears of the listener 1 such that the aural event is perceived as coming unchanged from the direction 6.” Ex. 1012, 12:20-24. As the POSA would have understood and appreciated, the HRTFs that are applied to the L and R signals must

necessarily be different, in order to achieve different resulting L and R signals. This is necessarily for the sound to be perceived to come from outside the listener's head. Therefore, because two different signals are provided to the left and right ears over headphones, Schröder describes "binaural reproduction of aural events using headphones" and therefore provides *binaural sound*. See Ex. 1012, 1:5-6.

137. In my opinion, upon reading Billinghamurst and Schröder, a POSA would have been motivated to implement Billinghamurst's wearable videoconferencing system with Schröder's HRTF processing. Doing so would have provided binaural sound that is localized to the listener outside the listener's head. The POSA would have realized that such an implementation is desirable and beneficial as, consistent with Billinghamurst's goals, it provides a more realistic and immersive virtual experience over headphones for a listener and avoids in-head localization. See also Section IX.A (Motivation to Combine Billinghamurst with Schröder).

c) *Billinghamurst discloses the SLP is one meter to two meters away from the called party*

138. Billinghamurst teaches an *SLP that is one meter to two meters away*. As Billinghamurst depicts in Fig. 2.0, and as was common in videoconferencing, the images of the videoconference participants appear *one meter to two meters away* from the wearer. One common goal of all videoconferences is to produce video displays of participants that replicate the experience of in-person, face-to-face

interactions. In general, videoconferences were widely known in the relevant time frame to mimic a conference room setting, where participants are seated around a physical conference table and could meet and collaborate face-to-face. While conference tables vary in size and dimension, the POSA would have readily appreciated that the typical size of a conference table would have been so that at least some participants are seated *one meter to two meters* away from each other. To mimic such a setting, the POSA would have understood that the images and avatars in Billingham's videoconference system would have appeared as if the remote participants were seated around a conference table, such that they appear *one meter to two meters* away from the user.

139. 8KHZ's expert in the parallel litigation agrees that a distance of *one meter to two meters* was known and obvious. I note that in his testimony, he admitted that the "typical way" to collect HRTFs around a KEMAR dummy was "in an anechoic chamber" with a "dummy head [] located centrally" and "a sound source [] located *1 to 2 meters away* from that dummy head." Ex. 1032, 126:5-128:11 (emphasis added).

140. More generally, a distance of *one meter to two meters* was well-known for social interaction and would have been a typical perceived distance for videoconferencing and social interaction. As an example, in 1957, Edward Hall introduced the concept known as proxemics, which is the study of social interaction

spacing. *See* Ex. 1030, 6-7 (describing Hall's work); *see also* Ex. 1028, 797 (explaining Hall's classification of personal social interactions as having a distance of 1 to 2 meters). As Hall recognized, the spacing between communicating parties plays a large role in the meaning and effectiveness of the messages being conveyed. Ex. 1030, 2, 10. Proximity allows participants to engage in mutual gaze which helps to regulate conversation flow, allow monitoring for understanding, communicate facial emotions, communicate a nature of the relationship, communicate status, prevent distraction, signal interest, and coordinating turn-taking. *Id.*, 10. Proximity also improves engagement, comprehension, and familiarity between participants. *Id.* A specific distance helps to define the relationship of the participants and establishes the proper parameters for the type of conversation. *Id.*, 5, 6. For example, close friends can effectively converse at a close distance without any issue, but communication between colleagues at the same distance would be distracting and uncomfortable. *Id.*

141. In 1957, Hall suggested that a personal conversation between friends occurs between 0.5 and 1.25 meters whereas impersonal business dealings occur between approximately 1.25 and 3.5 meters. *Id.* Based on these known distances, the POSA would have found it obvious to have a distance of *one meter to two meters* for interaction over a videoconference system, such as that described in in Billinghamurst.

142. As the POSA would have understood, providing effective communication over a video conferencing system requires an effective spacing between participants. Ex. 1028, 795-6. A realistic face-to-face video interaction can achieve the same benefits associated with an in-person interaction. *Id.*, 796. For example, one study found that “[f]or natural one-to-one conversation the typical distance is the personal distance around one meter[,]” where any closer “may be perceived inconvenient or arousing.” *Id.*, 797. That said, a distance “larger than 1.7 meters” was deemed too far for conversation. *Id.* In another study, replicating a conference table setting in a video conference involved displaying remote participants as life-size images at a distance of approximately 4 feet, or 1.22 meters. Ex. 1029, 387, *see also id.*, 385. Accordingly, in my opinion, it was known for the optimal distance in one-to-one conversation, and therefore for the perceived life-like display of a participant in a one-on-one video conversation, to be *one to two meters*.

143. Therefore, in view of the above, the POSA would have appreciated that the audio streams in Billinghamurst are “spatialized in real time so that they appear to emit from the corresponding avatar[,]” and the perceived distance between the listener and the avatars would have been at typical distances for communication and social and business interactions, within the range of *one to two meters*. Ex. 1010, 4. Thus, Billinghamurst teaches an *SLP that is one meter to two meters away from the called party* because the avatars appear *one meter to two meters* away from the

listener (*called party*), and the audio streams are spatialized to appear to emit from the avatars.

144. As I explained above for 9[pre], Billinghurst teaches using its system *during a telephone call*. See Section IX.C.1 (Ground 1, 9[pre]). Implementing Billinghurst's videoconferencing system for a *telephone call* would have provided spatial auditory cues to the user. This would have made it easier for the user to distinguish amongst the remote collaborators, and improved the listener's experience of a *telephone call* by making it feel more like a face-to-face conversation.

d) Conclusion

145. Thus, it is my opinion that a POSA would have readily understood that Billinghurst in view of Schröder teaches *processing, during the telephone call and with a CPU (processor), an audio stream of the remote collaborator (voice of the calling party) with the HRTFs taught in Schröder so the audio stream of the remote collaborator (voice of the calling party) localizes to the listener (called party) as spatialized sound (binaural sound) at the avatar's position (sound localization point) that is one meter to two meters away from the listener (called party)*.

3. 9[b]: “*displaying, during the telephone call and to the called party with a wearable electronic device worn on a head of the called party, an image that represents the calling party that is seen by the called party as being at the SLP that is one meter to two meters away from the called party;*”

146. In my opinion, Billinghurst discloses this limitation.

147. Billinghamhurst's wearable hardware includes a display that is part of the wearable device, as is depicted in Fig. 3.0 (reproduced below). Ex. 1010, 4, ("Figure 3.0 shows a user wearing the display and computer."). Billinghamhurst's "display" is "a pair of Virtual i-O iglasses! converted into a monoscopic display by the removal of the left eyepiece." *Id.*



Figure 3.0 The Wearable Hardware.

148. As Billinghamhurst depicts in Fig. 3.0, the monoscopic display is part of a set of wearable hardware that is *a wearable electronic device worn on a head* of the listener (*called party*). As Billinghamhurst depicts in Fig. 2.0 (reproduced below), the *wearable electronic device display[s]* to the listener (*called party*) *an image that represents* a remote collaborator (*calling party*). Moreover, Billinghamhurst's wearable display is a *wearable electronic device* even under 8KHZ's position that a *wearable electronic device* is a "portable wireless electronic device that is worn on or attached to a person," as Fig. 2.0 depicts the *wearable device* as having no wires. *See also*

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

Ex. 1010, 1 (referencing wearables as having “wireless communications hardware”),
5 (referencing the wearable computer as having wireless capability).

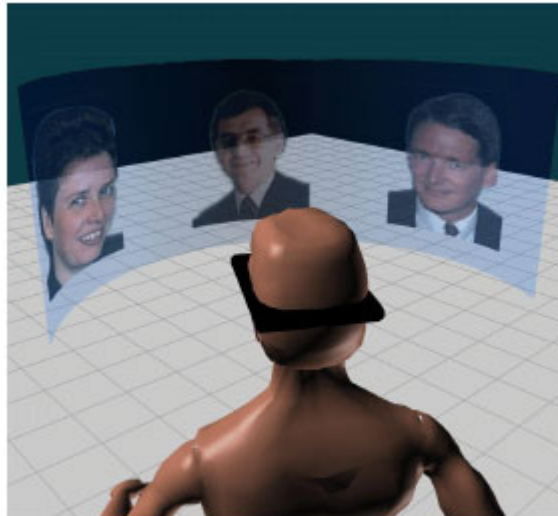


Figure 2.0 A Spatial Conferencing Space.

149. As I explained above for 9[a], Billinghurst’s Fig. 2.0 depicts the images of the remote collaborators as being *one meter to two meters* away from the wearer, which were typical distances for inter-personal communication and interaction. *See* Section IX.C.2 (Ground 1, 9[a]). In addition, Billinghurst co-locates the voice and the image so that the wearer perceives both to originate from the same location, the *SLP*. Specifically, Billinghurst states the audio streams are “spatialized in real time so that they to emit from the corresponding avatar.” Ex. 1010, 4. Thus, it is my opinion that Billinghurst teaches the avatars appear *one to two meters* away from the listener (*called party*), and the audio streams are spatialized to emit from the avatars, corresponding to the *SLP*. Moreover, as 8KHZ’s own expert in the parallel district court litigation admitted, it was known before 2014 for audio and visual events to be

connected. Ex. 1032, 244:16-245:12, 236:2-237:4 (describing co-locating an auditory object with a visual object in one location would improve the sense of motion as the object moves); *see also* Ex. 1036, 95 (“most sounds are somehow connected to visual events”).

150. Moreover, as I explained above for 9[pre], Billinghurst’s videoconference system is a *telephone call* under Patent Owner’s interpretation, and the Billinghurst videoconference system displays the avatar. *See* Section IX.C.1 (Ground 1, 9[pre]). Billinghurst’s implementation would have been a natural extension of a *telephone call*, by providing spatial auditory cues to the listener during a *telephone call*. This would have made it easier for the listener to distinguish amongst the remote collaborators, and improved the listener’s experience by making it feel more realistic and natural.

151. Thus, it is my opinion that the POSA would have understood Billinghurst discloses *displaying, during the telephone call and to the listener (called party) with a wearable electronic device worn on a head of the listener (called party), an avatar of (image that represents) the remote collaborator (calling party) as being at the avatar’s position (SLP) that is one meter to two meters away from the listener (called party).*

4. 9[c]: “*tracking, during the telephone call and with the wearable electronic device, head movements of the called party with respect to the image that represents the calling party and*

that is seen by the called party as being at the SLP; and”

152. In my opinion, Billinghamurst discloses this limitation.

153. Billinghamurst’s system *track[s] head movements* of the user (*called party*). Billinghamurst depicts in Fig. 1.0b (which I have reproduced below) a “body-stabilized” display in which, as Billinghamurst describes, “information is fixed relative to the users [sic] body position and varies as the user changes viewpoint orientation.... This requires the users [sic] viewpoint orientation to be tracked.” Ex. 1010, 3. In other words, and as depicted in Fig. 1.0b, Billinghamurst’s “body-stabilized” display is one in which “the user experiences a hemispherical information surround.” *Id.*, 3-4. The Billinghamurst system *tracks* the user’s viewpoint orientation (which is based on and affected by the user’s *head movements*). *Id.*, (to “track[] the user’s head orientation as they look around the space” only requires “a simple one degree of freedom tracker” to provide “cylindrical spatial information displays.”).

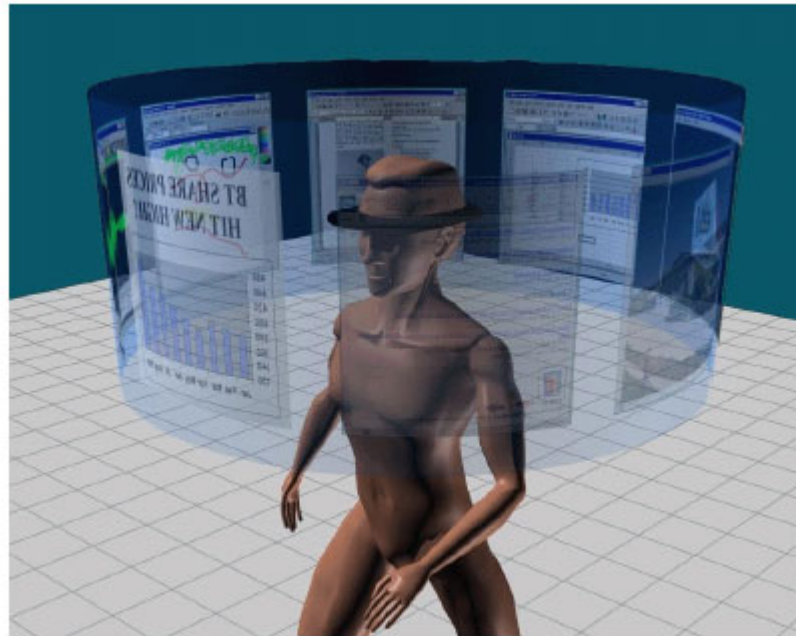


**Figure 1.0b One Degree of Freedom
Body-Stabilized Display**

154. Billinghurst even goes one step further by describing a “world-stabilized” display in which “information is fixed to real world locations and varies as the user changes viewpoint orientation and position. This requires the users [sic] viewpoint position and orientation to be tracked.” *Id.* In the “world-stabilized” display, therefore, Billinghurst not only tracks the head orientation, but also the user’s head position, which also meets the limitation *tracking head movements*.

155. In my opinion, therefore, either the “body-stabilized” display or the “world-stabilized” display meets this limitation, because both involve *tracking head movements* of the user.

156. According to Billinghurst, and as would have been readily appreciated by a POSA, such displays are “attractive” because they are immersive. *Id.*, 4.

Specifically, the Billingham displays are “very natural to use since most head and body motion is about the vertical axis, making it very difficult for the user to become disoriented.” *Id.*, 4. A POSA would understand that tracking an orientation and/or position of the user’s head would result in a tracking of the user’s head movements, as changes in orientation are indications of head movement.

157. Billingham’s *tracking* of the *head movements* is performed *with the wearable electronic device*. Specifically, as explained above, Billingham’s *wearable electronic device* includes the iglasses! display, which has “a sourceless two-axis inclinometer and a magnetometer used as a three degree of freedom orientation tracker.” *Id.*, 5. It is this combination of the inclinometer and the magnetometer in the Billingham system that allows it to track the head movements of the listener wearing the wearable hardware.

158. Billingham’s *track[ed] head movements* of the user are *with respect to the image* of the remote collaborator (*calling party*). This is because the user “can simply turn to face the speakers they are interested in.” *Id.*, 5. As Billingham states, “[j]ust as in face-to-face collaboration, users could turn to face the collaborators they wanted to talk to while still being aware of other conversations taking place.” *Id.*, 4. In both the body-stabilized and world-stabilized displays, Billingham enables the user to turn to face the avatar of a collaborator, so that the avatars appear “fixed relative to the users [sic] body position and varies as the user changes viewpoint

orientation.” Therefore, in Billinghamurst, either body-stabilized or world-stabilized displays meets this limitation, as the *head movements are with respect to the image that represents the calling party*.

159. As I explained above for 9[pre], Billinghamurst’s videoconference system is a *telephone call* under Patent Owner’s interpretation, and its videoconference system also includes tracking the head movements of the listener. *See* Section IX.C.1 (Ground 1, 9[pre]). As explained, implementing head movement tracking for a *telephone call* would have made the environment feel more realistic like a face-to-face conversation, improving the listener’s experience of the *telephone call* by making it feel more like a face-to-face conversation.

160. As I explained above for 9[a] and 9[b], Billinghamurst’s avatar’s position is an *SLP* because the sound appears to emit from the avatar. *See* Section IX.C.2-3 (Ground 1, 9[a]-9[b]).

161. Thus, it is my opinion that a POSA would have understood Billinghamurst discloses *tracking, during the telephone call and with the wearable electronic device, head movements of the listener (called party) with respect to the avatar of (image that represents) the remote collaborator (calling party) and that is seen by the listener (called party) as being at the avatar’s position (SLP)*. *See* Section IX.C.2-3 (Ground 1, 9[a]-9[b]).

5. 9[d]: “*changing, during the telephone call and in response to*

the head movements of the called party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the SLP that is seen by the called party as being the image that represents the calling party that is one meter to two meters away from the called party while the head of the called party moves with respect to the image that represents the calling party.”

162. In my opinion, Billinghamurst in view of Schröder discloses this limitation.

a) *Billinghamurst discloses this limitation but is not explicit that it changes HRTFs*

163. In Billinghamurst, the audio stream (*voice*) of the remote collaborator (*calling party*) continues to localize to the SLP that is seen by the listener (*called party*) as being the avatar (*image*) that represents the remote collaborator (*calling party*) while the head of the listener (*called party*) moves with respect to the avatar (*image*) that represents the remote collaborator (*calling party*) that is one meter to two meters away from the listener (*called party*). See Sections IX.C.2-4 (Ground 1, 9[a]-9[c]).

164. As I explained above for 9[pre], Billinghamurst’s videoconference is a *telephone call* under 8KHZ’s interpretation, and it changes the processing of audio streams. See Section IX.C.1 (Ground 1, 9[pre]). As I explained above for in 9[c], Billinghamurst discloses a “body-stabilized” system and a “world-stabilized” system,

either of which changes the images and audio streams based on the listener's head movements and therefore meets this limitation. *See* Section IX.C.4 (Ground 1, 9[c]).

165. Specifically, in Billingham's "body-stabilized" system, "information is fixed relative to the users [sic] body position and varies as the user changes viewpoint orientation" and "requires the users [sic] viewpoint orientation to be tracked." Ex. 1010, 3. Upon the user connecting to Billingham's "body-stabilized" system, the audio streams of the remote collaborators are "spatialized according to the distance and direction between speaker and listener. As users face or move closer to different speakers the speaker volume changes due to the sound spatialization." *Id.*, 4. By spatializing the remote collaborators' audio streams and adjusting them according to the user's direction and movement, the "body-stabilized" system in Billingham continues to localize the voice of a remote collaborator at the avatar of that remote collaborator while the head of the listener moves.

166. Similarly, in Billingham's "world-stabilized" system, "information is fixed to real world locations and varies as the user changes viewpoint orientation and position" and "requires the users [sic] viewpoint positions and orientation to be tracked." *Id.* The information that is "fixed" while the "user changes viewpoint orientation" and/or "position" includes "audio streams [from remote collaborators] spatialized in real time so that they appear to emit from the corresponding avatar." *Id.* Again, with the "world-stabilized" system, by spatializing the remote

collaborators' audio streams and adjusting them according to the user's viewpoint orientation and position, the "world-stabilized" system in Billinghamurst also continues to localize the voice of a remote collaborator at the avatar of that remote collaborator while the head of the listener moves.

167. While Billinghamurst does not expressly provide detail regarding how *the voice continues to localize to the SLP*, I note that PO's expert in the parallel district court litigation admitted that this was known as early as the 1980s. Specifically, PO's expert admitted that it was known for "head-track virtual acoustic systems [to] keep a sound source fixed in a location in virtual space independent of head movement"). Ex. 1032, 250:2-9; *see also* Ex. 1036, 12, 99, 132-34, Figs. 4.37-4.38. Moreover, this was a detail that was well-known in the art, and is expressly taught by Schröder.

b) *Schröder discloses changing HRTFs so the sound continues to localize to the SLP even while the head of the listener moves*

168. Specifically, Schröder processes sound signals that are provided over headphones so that "they have a constant spatial direction." Ex. 1012, Title. I have reproduced below Fig. 7 from Schröder. As Schröder explains in describing Figs. 5 and 7 (which are identical such that the quoted text of Schröder in relation to Fig. 5 also applies to Fig. 7), Schröder's system "appl[ies] an HRTF" to process sound signals provided to the left and right ears over headphones so that the "listener

perceives the sound event as coming from outside the head, illustrated in the figure by arrow 6.” *Id.*, 3:12-14, *see also id.*, 4:18-19 (stating Figure 7 corresponds to Figure 5).

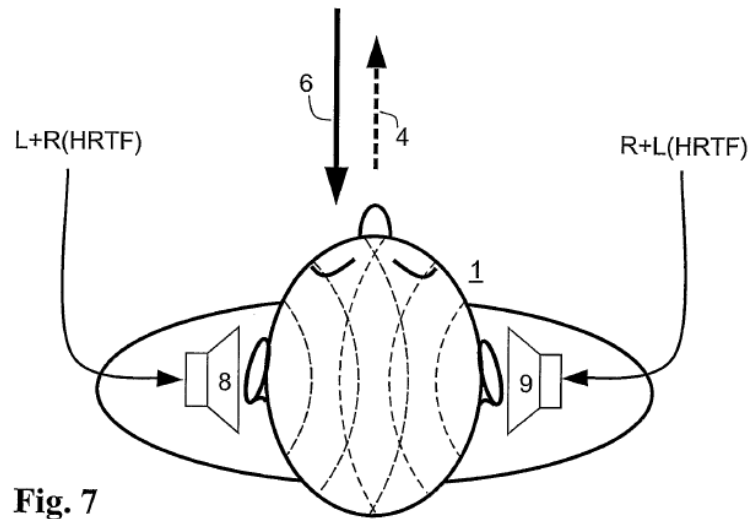


Fig. 7

169. I have reproduced below Fig. 8 from Schröder. As Schröder illustrates in Fig. 8, the user has turned his head 90 degrees to his right (compared to Fig. 7), yet the direction arrow 6 is unchanged from Fig. 7. Therefore, when the user “change[s] his position,” the “aural event is perceived as coming from the original direction 6.” *Id.*, 4:12-25.

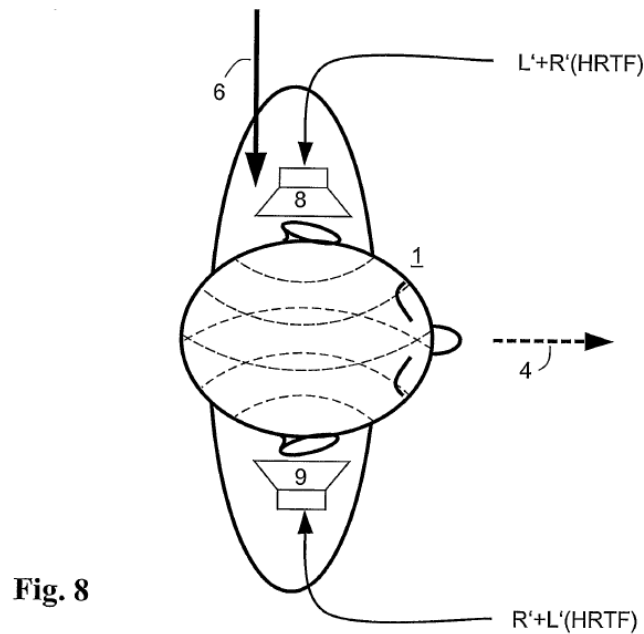


Fig. 8

170. Therefore, in Schröder’s system, the sound appears to remain spatially fixed from direction arrow 6 even when “the listener turns his head.” *Id.*, 5:27-30. Causing the sound to remain spatially fixed even as the listener turns his head is the same concept as used in the body-stabilized and world-stabilized displays in Billinghamurst. Schröder provides specific details on how to implement its system, with the use of HRTFs) to ensure the sound remains spatially fixed, which the POSA would have known to implement in Billinghamurst’s system.

171. Specifically, to cause the sound to come from a “constant spatial direction,” Schröder states “[t]he signals for the left and right ears are adjusted, on the basis of the position of the listener or of his head, using variable filters for the HRT function such that the direction from which the aural event appears to come remains stable in space.” *Id.*, 5:30-35. For example, “the variable HRT filters are

[] adjusted such that the aural event is perceived as coming from the direction 6 again.” *Id.*, 13:8-11. Accordingly, it is my opinion that Schröder discloses *changing, in response to the head movements* of a listener (or wearer/user of the headphones, *called party*), *the HRTFs processing* the sound so that the sound *continues to localize* to the same fixed direction.

172. In my opinion, a POSA would have recognized that Schröder’s teaching of *changing the HRTFs* would be the conventional way to implement Billingham’s described capability of localizing the voices of the virtual avatars to a stabilized or fixed location even while the user moves his head. A POSA would have found it obvious to include Schröder’s teaching of changing HRTFs in Billingham’s system, as HRTFs were well-known and commonly used in the art to manipulate the perceived origination of sound, and it was known to change HRTFs as the user’s head moves to fix the perceived location of the sound. *Id.*

173. Moreover, as I explained above for 9[pre], Billingham’s videoconference system is a *telephone call* under Patent Owner’s interpretation, including changing the processing of the audio streams. *See* Section IX.C.1 (Ground 1, 9[pre]). Such an implementation would have provided spatial auditory cues to the listener during a *telephone call*, making it easier for the listener to distinguish amongst the remote collaborators, and improving the listener’s experience by making it feel more like a face-to-face conversation.

174. Additionally, as explained above for 9[a], Billinghurst also teaches that the continual localization to the SLP representing the calling party appears a distance that is *one meter to two meters away from the called party*. See Section IX.C.3 (Ground 1, 9[b]).

c) Conclusion

175. Thus, it is my opinion that a POSA would have understood Billinghurst discloses *changing, during the telephone call and in response to the head movements of the listener (called party), the HRTFs processing taught in Schröder the audio stream of the remote collaborator (voice of the calling party) so the audio stream of the remote collaborator (voice of the calling party) continues to localize to the avatar's position (SLP) that is seen by the listener (called party) as being the avatar of the remote collaborator (image that represents the calling party) that is one meter to two meters away from the listener (called party) while the head of the listener (called party) moves with respect to the avatar of the remote collaborator (image that represents the calling party)*.

D. Dependent Claims 10-14¹

1. Claim 10: “*The method of claim 9 further comprising:*

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

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

tracking, during the telephone call and with a wearable electronic device worn on a head of the calling party, head movements of the calling party;

moving, during the telephone call, the image that represents the calling party to copy the head movements of the calling party so the called party sees, with the display of the wearable electronic device worn on the head of the called party, the head movements of the calling party; and

changing, during the telephone call and in response to the head movements of the calling party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the image that represents the calling party that is seen by the called party as being at the SLP that is one meter to two meters away from the called party while the image that represents the calling party moves.”

176. In my opinion, Billinghurst in view of Schröder discloses this limitation.

177. In Billinghurst’s system, not only does the user depicted in Fig. 2.0 wear the wearable hardware, but the remote collaborators also wear their own wearable hardware. In this manner, it is my opinion that Billinghurst discloses *tracking, during the telephone call and with a wearable electronic device worn on a head of the remote collaborator (calling party), the head movements of the remote collaborator (calling party).* Specifically, Billinghurst’s “remote users may also be using wearable computers and head mounted displays” creating a wearable communications space that supports “numerous collaborative applications in which some participants are either not sitting at desks or need mobility.” Ex. 1010, 4, 7.

“In this way the conference space can support dozens of simultaneous users, similar to current collaborative virtual environments.” Ex. 1010, 4; *see also id.*, 1. Therefore, the POSA would have known that the remote users are wearing wearable hardware and head mounted display devices similar to the main user depicted in Billinghamhurst’s Fig. 2.0. As discussed in 9[c], the *wearable electronic device worn on a head* of a user, including the remote collaborator (*calling party*), *track[s] the head movements* of the user with the combination of the inclinometer and the magnetometer in the Billinghamhurst wearable hardware. *Id.*, 5; *see* Section IX.C.4 (Ground 1, 9[c]).

178. In my opinion, Billinghamhurst further discloses *moving, during the telephone call*, the avatar of the remote collaborator (*image that represents the calling party*) *to copy the head movements of the remote collaborator (calling party) so the listener (called party) sees, with the display of the wearable hardware (wearable electronic device) worn on the head of the listener (called party), the head movements of the remote collaborator (calling party)*. When the remote collaborator in Billinghamhurst’s videoconferencing system moves their head, the image displayed to the user also moves in the same manner, since the image is a live video stream of the remote collaborator. Ex. 1010, 4 (the “avatars could be live video streams.”); *see also id.*, 7 (“We will incorporate live video texture mapping...enabling users to see their remote collaborators as they speak.”). According to Billinghamhurst, users

prefer “both the audio and visual cues together.” *Id.*; *see also id.*, 1 (“The user can use natural head motions to attend to the remote collaborators, can communicate freely while being aware of other side conversations and can move through the communication space.”). Billinghurst recognizes the benefits of having live video streams of the remote collaborators, as it resembles live conversations. *Id.*, 2 (“In face-to-face conversations, speech, gesture, body language and other non-verbal cues combine to show attention and interest.”). As the POSA would have understood, without these cues, “users often find it difficult to know when people are paying attention to them, to hold side conversations, and to establish eye contact.” *Id.* Moreover, as I explained above for 9[b], the avatar of the remote collaborator, including the live video, appears in *the display of the wearable hardware (wearable electronic device) worn on the head of the listener (called party)*. *See* Section IX.C.3 (Ground 1, 9[b]).

179. Moreover, in my opinion, Billinghurst also discloses *changing, during the telephone call and in response to the head movements of the remote collaborator (calling party), the processing of the audio stream of the remote collaborator (voice of the calling party) so the audio stream of the remote collaborator (voice of the calling party) continues to localize to the avatar of the remote collaborator (image that represents the calling party) that is seen by the listener (called party) as being at the avatar’s position (SLP) that is one meter to two meters away from the listener*

(called party) while the avatar of the remote collaborator (image that represents the calling party) moves. Specifically, Billinghurst states that “[t]he avatars could be live video streams and as they speak their audio stream spatialized in real time so that they appear to emit from the corresponding avatars.” *Id.*, 4. In Billinghurst’s system, this spatialization is changed “according to the distance and direction between speaker and listener,” such that “[a]s users face...different speakers the speaker volume changes due to the sound spatialization.” *Id.*, 5. The POSA, upon reading Billinghurst, would have understood that allowing the avatar to reflect the head movement of the remote collaborator, allows users to “turn to face the collaborators they wanted to talk to while still being aware of other conversations taking place.” *Id.*, 5. Therefore, Billinghurst’s users are able to move and turn to face different speakers, and the audio adjusts accordingly to maintain the SLPs of remote collaborators at fixed locations to reflect a face-to-face conversation.

180. As I explained above for 9[d], Billinghurst does not expressly provide detail regarding *changing the HRTFs processing the voice of the calling party*, where this processing allows the audio stream to *continue[] to localize to the SLP*, but this detail is expressly taught by Schröder. *See* Section IX.C.5 (Ground 1, 9[d]).

181. As explained in 9[pre], Billinghurst’s videoconference system is a *telephone call* under Patent Owner’s interpretation, and it further includes tracking

head movements, moving images, and changing the processing of the audio stream.

See Section IX.C.1 (Ground 1, 9[pre]).

2. Claim 11: “*The method of claim 9 further comprising:*

tracking, during the telephone call and with a wearable electronic device worn on a head of the calling party, head movements of the calling party; and

changing, during the telephone call and in response to the head movements of the calling party, the voice of the calling party so the voice of the calling party emulates a sound that the called party would hear if the calling party were physically present at the SLP and talking to the called party.”

182. In my opinion, Billingham discloses this limitation.

183. As I explained above for claim 10, Billingham discloses *tracking, during the telephone call and with a wearable electronic device worn on a head of the remote collaborator (calling party), the head movements of the remote collaborator (calling party)*. Ex. 1010, 4 (“These remote users may also be using wearable computers.”). See Section IX.D.1 (Ground 1, Claim 10).

184. As I explained above for claim 10, Billingham also discloses *changing, during the telephone call and in response to the head movements of the remote collaborator (calling party), the audio stream of the remote collaborator (voice of the calling party) so that the audio stream of the remote collaborator (voice of the calling party) is spatialized based on the position and orientation of the user, including changing the sound in each ear when the remote collaborator turns to face the*

listeners, so that the audio stream *emulates a sound that the called party would hear if the calling party were physically present at the SLP and talking to the called party.*

See Section IX.D.1 (Ground 1, Claim 10).

3. **Claim 12:** “*The method of claim 9 further comprising: displaying, during the telephone call and to the called party with the wearable electronic device worn on the head of the called party, the image that represents the calling party as a virtual image of the calling party.*”

185. In my opinion, Billingham discloses this limitation.

186. In the '628 patent, the only mention of a displayed “virtual image” is that a “sound localization point 2050 can appear to the listener 2020 as a virtual image of the speaker 2010.” Ex. 1001, 33:20-22. As an example of this, the patent explains, “For example, the wearable electronic 2015 is a pair of electronic glasses with a display that provides an image or video of the speaker 2010 in a field of view 2080 of the listener 2020.” Therefore, it is my opinion that Billingham’s description of a live video stream of the remote collaborator meets this limitation of a *virtual image*. Thus, as I explained above for 9[b] and claim 10, Billingham discloses *displaying, during the telephone call and to the listener (called party) with the wearable electronic device worn on the head of the listener (called party), an avatar that is the live video stream of the remote collaborator (the image that represents the calling party as a virtual image of the calling party).* Ex. 1010, 4 (“The avatars

could be live video streams”), Fig. 2 (reproduced below); *see* Sections IX.C.3 and IX.D.1 (Ground 1, 9[b] and Claim 10).

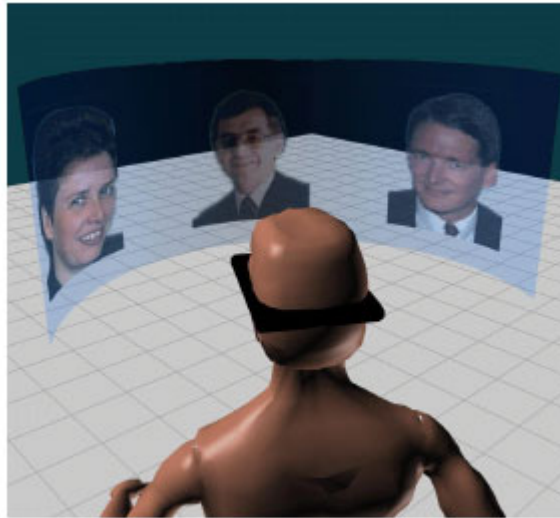


Figure 2.0 A Spatial Conferencing Space.

Ex. 1010, Figure 2.0

4. **Claim 13:** “*The method of claim 9 further comprising: displaying, during the telephone call and to the called party with the wearable electronic device worn on the head of the called party, the image that represents the calling party as an augmented reality image of the calling party.*”

187. In my opinion, Billinghurst discloses this limitation. *See* Section IX.C.3 (Ground 1, 9[b]).

188. Billinghurst states that its wearable system is an “augmented reality communication space.” Ex. 1010, 1, 4-5. That system has display hardware in the form of a “Virtual i-O head mounted display” having a “see-through [] mode” so that “the user could also see the real world at the same time, enabling the remote collaborators to help them with real world tasks.” *Id.*; *see also id.*, 7 (“A user could,

for example, choose to view avatars of conference participants overlaid and attached to a physical notice board.”). In my opinion, a POSA would understand that in Billinghamurst, because a “see-through mode” is used to display the avatars over the real world, the Billinghamurst display includes *displaying, during the telephone call and to the user (called party) with the wearable hardware (wearable electronic device) worn on the head of the user (called party), the avatar (image) that represents the remote collaborator (calling party) as a see-through image displayed over the real world (an augmented reality image) of the remote collaborator (calling party).* See Section IX.C.3 (Ground 1, 9[b]).

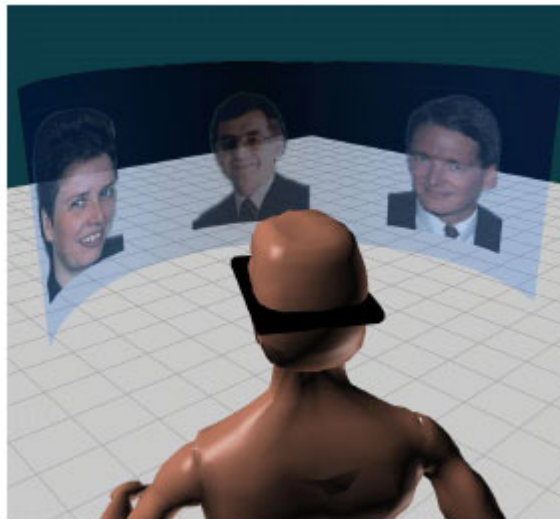


Figure 2.0 A Spatial Conferencing Space.

5. Claim 14: “*The method of claim 9 further comprising:*

moving, during the telephone call, the image that represents the calling party around a room where the called party is located;
and

changing, during the telephone call and in response to

movements in the room of the image that represents the calling party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the image that represents the calling party even as the image that represents the calling party moves around the room.”

189. In my opinion, Billinghamurst in view of Schröder discloses this limitation.

190. In Billinghamurst’s videoconferencing system, users are allowed to move around the space, corresponding to a *room where the called party is located*, as Billinghamurst depicts in Fig. 4.0 (reproduced below). Specifically, Billinghamurst teaches that “[a]s a user moves through the space, their avatar’s ID tag, positional and orientation information is broadcast onto the transformation multicast group.” Ex. 1010, 6. As Billinghamurst further describes, “[w]hen received by each client in the group it is used to update the relevant avatar’s position and orientation.” *Id.* “Each user determines the position and orientation of their own avatar in space, which changes as they move or look about [the] environment.” *Id.*, 5, Fig. 4.

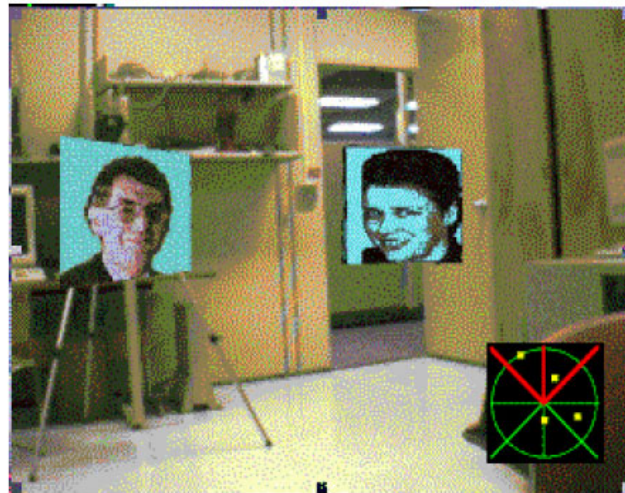


Figure 4.0 The User's View of the Wearable Conferencing Space

191. As Billinghurst further explains, the audio stream will continue to localize to the avatar of the remote collaborator (*image that represents the calling party*) even as the avatar *moves around the room* because “[t]he transformational information is also used in spatialising the user’s audio stream relevant to the user’s position.” *Id.*, 6. Therefore, as the remote collaborator moves around the virtual space, the image of the corresponding avatar also moves, and the remote collaborator’s audio stream continues to be spatialized to the avatar’s location as it moves about. Specifically, as I explained above for 9[d], Billinghurst does not expressly provide detail regarding *changing... the HRTFs processing the voice of the calling party*, to enable the audio stream to *continue[] to localize to the image*, but this detail was well known in 2014 and is expressly taught by Schröder. See Section IX.C.5 (Ground 1, 9[d]). Therefore, it is my opinion that Billinghurst-

Schröder render obvious *moving, during the telephone call, the avatar (image) that represents the remote collaborator (calling party) around a room where the user (called party) is located; and changing, during the telephone call and in response to movements in the room of the avatar (image) that represents the remote collaborator (calling party), the HRTFs processing the voice of the remote collaborator (calling party) so the voice of the remote collaborator (calling party) continues to localize to the avatar (image) that represents the remote collaborator (calling party) even as the avatar (image) that represents the remote collaborator (calling party) moves around the room.* See Sections IX.C.3, 5 (Ground 1, 9[b], 9[d]).

E. Independent Claim 16:

1. **16[pre]:** “*A non-transitory computer readable storage medium storing instructions that one or more electronic devices in an electronic system execute as a method during a telephone call between a calling party and a called party, the method comprising:*”

192. In my opinion, to the extent it is limiting, Billingham discloses the preamble.

193. As I already explained above for 9[pre], Billingham discloses wearable hardware (*one or more electronic devices*) in a videoconferencing system (*an electronic system*) that executes as a method during a *telephone call between a remote collaborator (calling party) and a listener (called party).* See Section IX.C.1 (Ground 1, 9[pre]).

194. In my opinion, Billingham further discloses memory (*non-transitory computer readable storage medium*) storing instructions that the wearable hardware (*one or more electronic devices*) executes as the method. Billingham specifically discloses that the wearable computer “is a custom built 586 PC 104 based computer with 20mb of RAM running Windows 95.” Ex. 1010, 4. A POSA would have understood that the RAM is memory that also stores instructions to execute other programs and algorithms, including the programs and algorithms disclosed in Billingham.

2. **16[a]:** “*displaying, during the telephone call and to the called party with a wearable electronic device worn on a head of the called party, an image of the calling party that appears to the called party as being at least one meter away from a head of the called party;*”

195. In my opinion, Billingham discloses this limitation.

196. As I explained above for 9[b], Billingham teaches *displaying, during the telephone call and to the listener (called party) with wearable hardware on the head of the listener (a wearable electronic device worn on a head of the called party), the avatar of the remote collaborator (an image of the calling party) that appears to the listener (called party) as being at least one meter away from a head of the listener (called party).* See Section IX.C.3 (Ground 1, 9[b]).

3. **16[b]:** “*processing, during the telephone call, a voice of the calling party with head related transfer functions (HRTFs) so the voice of the calling party localizes to the called party as*

binaural sound at the image of the calling party that appears at least one meter away from the head of the called party; ”

197. In my opinion, Billinghurst in view of Schröder discloses this limitation.

198. As I explained above for 9[a], the combination of Billinghurst and Schröder teaches *processing, during the telephone call*, an audio stream of the remote collaborator (*voice of the calling party*) with HRTFs so the audio stream of the remote collaborator (*voice of the calling party*) localizes to the listener (*called party*) as spatialized sound (*binaural sound*) at the avatar of the remote collaborator (*image of the calling party*) that appears at least one meter away from the head of the listener (*called party*). See Section IX.C.2 (Ground 1, 9[a]).

4. 16[c]: *“tracking, during the telephone call, head movements of the called party; and.”*

199. In my opinion, Billinghurst discloses this limitation.

200. As I explained above for 9[c], Billinghurst discloses *tracking, during the telephone call, head movements of the listener (called party)*. See Section IX.C.4 (Ground 1, 9[c]).

5. 16[d]: *“changing, during the telephone call and in response to the head movements of the called party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the image of the calling party that appears at least one meter away from the head of the called party as the head of the called party moves and changes head orientations.”*

201. In my opinion, Billinghurst in view of Schröder discloses this limitation.

202. As I explained above for 9[d], Billinghurst in view of Schröder teaches *changing, during the telephone call and in response to the head movements of the listener (called party), the HRTFs taught by Schröder processing the audio stream of the remote collaborator (voice of the calling party) so the audio stream of the remote collaborator (voice of the calling party) continues to localize to the avatar of the remote collaborator (image of the calling party) that appears at least one meter away from the head of the listener (called party) as the head of the listener (called party) moves and changes head orientations. See Section IX.C.5 (Ground 1, 9[d]).*

F. Dependent Claims 17-20:

1. **Claim 17:** *The non-transitory computer readable storage medium storing instructions of claim 16 further to cause the one or more electronic devices in the electronic system to execute the method comprising:*

tracking, during the telephone call, head movements of the calling party; and

changing, during the telephone call and in response to the head movements of the calling party, the HRTFs processing the voice of the calling party so the voice of the calling party continues to localize to the image of the calling party that appears at least one meter away from the head of the called party as a head of the calling party moves.

203. In my opinion, Billinghurst in view of Schröder discloses this limitation.

204. As I explained above for claim 10, Billinghurst in view of Schröder teaches *tracking, during the telephone call, head movements of the remote collaborator (calling party) and changing, during the telephone call and in response to the head movements of the remote collaborator (calling party), the HRTFs as taught by Schröder processing the audio stream of the remote collaborator (voice of the calling party) so the audio stream of the remote collaborator (voice of the calling party) continues to localize to the avatar of the remote collaborator (image of the calling party) that appears at least one meter away from the head of the listener (called party) as a head of the remote collaborator (calling party) moves. See Section IX.D.1 (Ground 1, Claim 10).*

2. **Claim 18:** *“The non-transitory computer readable storage medium storing instructions of claim 16, wherein a location of the image of the calling party also functions as a virtual microphone point (VMP) for sound that is captured by the wearable electronic device worn on the head of the called party, and the sound captured by the wearable electronic device worn on the head of the called party is provided to the calling party.”*

205. In my opinion, Billinghurst discloses this limitation.

206. The '628 Patent states that a “virtual microphone point” is “a virtual location or virtual position where sound is captured, recorded, or monitored.” Ex. 1001, 42:33-35. That is, as the '628 Patent explains, the system can “designate a virtual microphone point in empty space” “that simulates a location of a listener with

whom the talker speaks.” *Id.*, 10:32-35. Thus, the POSA would have understood that while an SLP in the patent refers to “emanation of sound” (*e.g.*, a simulated location of a talker), a *virtual microphone point* in the patent refers to “sound... capture” (*e.g.*, a simulated location of a listener). *Id.*, 42:30-35.

207. In my opinion, Billinghamurst discloses the SLP and the *virtual microphone point* are co-located at the avatar’s position (*location of the image of the calling party*), as the users (including the local user and remote users) all participate in and share the same conferencing system. That is, the avatar’s position functions as both an SLP and a *virtual microphone point*.

208. Specifically, the local user and the remote collaborators in Billinghamurst’s system share the same virtual space (as depicted in the bottom right corner of Fig. 4.0 of Billinghamurst, reproduced below). Therefore, the remote collaborator’s perception of the local user mirrors the local user’s perception of the remote collaborator. Ex. 1010, 5 (the common conferencing space “enabl[es] users to find each other easily” and their audio “is spatialized according to the distance and direction between speaker and listener.”). In other words, if a first user perceives a second user’s avatar to the right of the first user, then the second user in the same virtual space would perceive the first user’s avatar to the left of the second user, mirroring the first user’s perception.

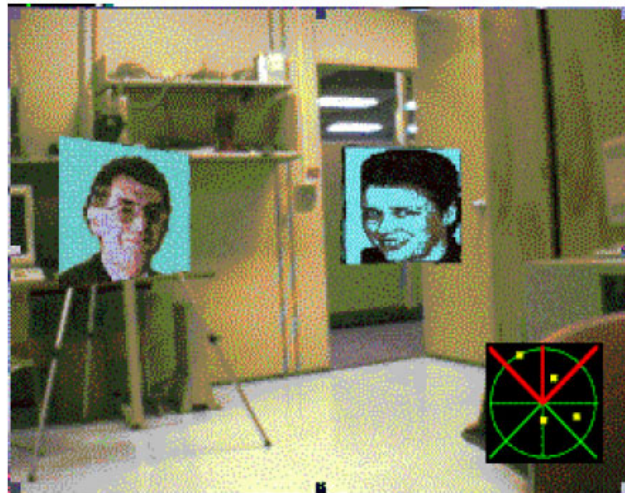


Figure 4.0 The User's View of the Wearable Conferencing Space

209. Thus, it is my opinion that in Billinghamhurst, just as the voice of the remote collaborator (*calling party*) is spatialized at an SLP for the local user (*called party*), the voice of the local user (*called party*) is spatialized at a *virtual microphone point* for the remote collaborator (*calling party*). In this manner, Billinghamhurst therefore discloses that the *location of the avatar (image) of the remote collaborator (calling party) also functions as a virtual microphone point for sound that is captured by the wearable hardware (wearable electronic device) worn on the head of the listener (called party)*.

210. As to *provid[ing] sound captured by the wearable electronic device worn on the head of the called party... to the calling party*, it is my opinion that Billinghamhurst also discloses this limitation. Specifically, in Billinghamhurst, the remote users have similar wearable hardware as the local user. Ex. 1010, 4-5, 7 (“remote

users may also be using wearable computers and head mounted displays” that have “headmounted microphone[s]”). In particular, as Billinghamurst explains, “[w]hen received by the other clients the senders IP address identifies the avatar that the audio belongs to and its position and orientation.” *Id.*, 6. “The audio is then spatialized in real time.” *Id.* “The transformational information [of the avatar] is also used in spatialising the user’s audio stream relevant to the receiving user’s position,” meaning that the voice of the listener (*sound captured by the wearable electronic device worn on the head of the called party*) is spatialized before being *provided to the remote collaborator (calling party)*. *Id.*

3. **Claim 19:** “*The non-transitory computer readable storage medium storing instructions of claim 16 further to cause the one or more electronic devices in the electronic system to execute the method comprising: changing, during the telephone call, the HRTFs processing the voice of the calling party so the voice of the calling party moves and follows the image of the calling party as the image of the calling party moves with respect to the called party.*”

211. In my opinion, Billinghamurst in view of Schröder discloses this limitation.

212. As I explained above for claim 14, Billinghamurst in view of Schröder teaches *changing, during the telephone call, the HRTFs* as taught by Schröder *processing the audio stream of the remote collaborator (voice of the calling party) so the audio stream of the remote collaborator (voice of the calling party) moves and*

follows the avatar of the remote collaborator (image of the calling party) as the avatar of the remote collaborator (image of the calling party) moves with respect to the listener (called party). Id., 6; see Section IX.D.5 (Ground 1, claim 14).

4. **Claim 20:** “*The non-transitory computer readable storage medium storing instructions of claim 16 further to cause the one or more electronic devices in the electronic system to execute the method comprising:*

providing the image of the calling party as being a virtual microphone point; and

reducing, during the telephone call, an amplitude of a voice of the called party that is provided to the calling party when the called party moves away from the virtual microphone point.”

213. In my opinion, Billingham discloses this limitation.

214. As I explained above for claim 18, Billingham discloses *providing the avatar of the remote collaborator (image of the calling party) at a location that also functions as being a virtual microphone point. Id., 6; see Section IX.F.2 (Ground 1, Claim 18).*

215. It is my opinion that Billingham further discloses *reducing, during the telephone call, the volume of the audio stream of the listener (an amplitude of a voice of the called party) that is provided to the remote collaborator (calling party) when the listener (called party) moves away from the location of the avatar of the remote collaborator (virtual microphone point). Ex. 1010, 5 (“As users face or move closer to different speakers the speaker volume changes due to the sound*

specialization... Audio culling is also used so that only the audio streams from the speakers closest to the listener are broadcast and spatialized.”), 6.

X. GROUND 2: BILLINGHURST AND SCHRÖDER IN VIEW OF ZHOU AND HADDICK RENDERS OBVIOUS CLAIMS 9-14 AND 16-20

216. To the extent that Billinghamurst does not explicitly disclose that the avatar’s position (*SLP*) is one meter to two meters away from the listener (called *party*), it is my opinion that this would have been obvious in view of Zhou.

217. To the extent that Billinghamurst does not disclose that the method is performed *during a telephone call*, it is my opinion that this would have been obvious in view of Haddick.

A. Motivation to Combine Billinghamurst and Schröder with Zhou

218. As I explained above in Section IX.A, a POSA would have been motivated to combine the teachings of Billinghamurst and Schröder. Furthermore, a POSA would have been motivated to further modify the system resulting from the combination of Billinghamurst and Schröder to localize the *SLP* to be 1.5 meters away from the listener as taught by Zhou. As Zhou already explained, the Billinghamurst paper used a most common paradigm for augmented reality – a single camera for capturing physical scenes. Ex. 1018, 7. This implementation is advantageous because it is cheap, has little computation cost, and is easily implemented. *Id.* However, according to Zhou, the set-up in Billinghamurst “can lead to the difficulty of

depth perception.” *Id.* For example, when a single camera is used to capture a physical scene, augmented objects can occlude physical objects behind them, which makes it difficult for the user to perceive the depth of the object. Zhou set out to solve this depth perception problem by introducing a 3D sound at a distance of 1.5 m, which Zhou concluded to significantly improve the accuracy of the user’s depth perception of an augmented object. *Id.*, 9. Therefore, a POSA would have recognized that the teachings of Zhou regarding the distance of an SLP would have advanced the objective of both Billinghamst and Schröder to simulate face-to-face collaboration. Ex. 1018, 1 (“[T]o access the impact of 3D sound on depth perception in a single-camera AR environment,” such as Billinghamst’s augmented reality system.).

219. Zhou is similar to Billinghamst and Schröder in that it also recognizes that 3D sound “has proven itself to be an effective addition to human-computer interfaces” and “provide[s] a more realistic virtual environment.” *Id.*, 2. Zhou indeed builds on the work of Billinghamst. Zhou expressly cited to Billinghamst as an example of an application that has “applied 3D sounds in AR to enhance user experiences,” and described Billinghamst as “a prototype wearable communication space that uses spatial and visual audio cues to enhance communication between remote groups of people.” *Id.* In my opinion, a POSA would have recognized “that binaural cues play an important role in auditory distance perception for nearby

sources.” *Id.*, 5. Thus, in considering appropriate distances for distance perception, a POSA implementing the system resulting from the Billingham and Schröder would have naturally looked at the distance in Zhou, as Zhou determined the distance where binaural cues would be effective in enhancing the user experience of the listener. Ex. 1018, 5 (“In our first experiment the source sources are always placed around 1.5 m away in the front of the listener.”) .

B. Motivation to Combine Billingham and Schröder with Haddick

220. As I explained above in Section IX.A, a POSA would have been motivated to combine the teachings of Billingham and Schröder, and as I explained above in Section IX.C.1, Billingham already discloses a *telephone call*, under PO’s understanding or a narrower interpretation requiring a phone call. However, to the extent that this is not explicit in Billingham, it is my opinion that this would have been obvious in view of Haddick.

221. Billingham already contemplates how portable computer technology was converging towards telephones, as issues of enhancing wearable videoconferencing systems are “becoming increasingly important as telephones incorporate more computing power and more portable computers become more like telephones.” Ex. 1010, 1-2. Based on this teaching and the knowledge of the POSA, it is my opinion that the POSA would have been motivated to modify the system

resulting from the combination of Billinghurst and Schröder for use during a *telephone call*, which is specifically taught by Haddick.

222. Specifically, when Billinghurst's paper was published in 1998, wearable hardware was bulky, expensive, and power inefficient. However, by the relevant time frame in 2014, mobile devices and associated wearable hardware were ubiquitous and powerful devices that could perform many, if not all, of the functions described in Billinghurst. Hardware in 2014 was more portable, smaller, lighter, and cheaper than Billinghurst's wearable system in 1998. Therefore, a POSA in 2014 would have found it obvious to update Billinghurst's 1998 hardware to implement its functionality on Haddick's wearable hardware capable of receiving telephone calls. As Haddick recognized, its wearable device (in the form of an eyepiece) "may be able to replace the need for having a separate mobile communication device, such as a mobile phone." Ex. 1011, ¶594.

223. Haddick is similar to Billinghurst and Schröder in that it also teaches a wearable device that can "reproduce [a] 3D sound profile" and "realize a direction of sound for the wearer." *Id.*, ¶653. As Haddick explains, a "wearer with headphones, hearing loss, in a loud environment, and the like, may not otherwise be able to tell what's being said and/or the direction of the person speaking, but is provided with a 3D sound enhancement from the eyepiece." *Id.* It would have been straightforward and obvious for a POSA implementing the system resulting from the

Billinghurst and Schröder to have looked at other wearable devices, like that of Haddick, and recognized that the advantages of using 3D sound during a videoconference call would have also applied during a *telephone call*, as it would enhance the user experience of the *telephone call*.

C. Reasonable Expectation of Success

224. In my opinion, a POSA would have reasonably expected to successfully combine the system resulting from the combination of Billinghurst and Schröder with the teachings of Zhou. Specifically, the POSA implementing the Billinghurst and Schröder system would have had to make a design choice as to at what distance for localizing the sound localization point, and the 1.5 meters taught by Zhou would have been one of the limited number of design choices, and would have had the benefit of enhanced depth perception, as taught in Zhou. Indeed, even in Figure 2.0 in Billinghurst, the avatars where the audio streams are localized appear to be one to two meters from the listener.

225. Moreover, the POSA would have known that to mimic a conference room setting, a typical social or business interaction spacing, or a life-like display of a participant in a video conversation, a distance of one to two meters would have been optimal. For example, Grämer in 2010 discussed measuring HRTFs for a distance of 1.5 meters, which was a typical distance for measuring HRTFs during the relevant time frame. Ex. 1039; *see also* Ex. 1036, 74-75 (describing a nominal

level of sound as corresponding to normal speech “at about 1 meter from a listener); *id.*, 77 (describing how a closely proximate sound source has a curved wavefront, such that below one meter, low frequency sound energy indicates the approach of sound, and “[a]s a sound source moves beyond a distance of approximately two meters, the wavefront becomes planar rather than spherical, and this rule begins to not apply.”); *id.*, 79 (describing the spectral changes caused by HRTF to be constant as a function of distance when beyond one meter). Furthermore, adjusting the audio streams to localize 1.5 meters away would just be a matter of programming the Billingham processor.

226. Moreover, it is my opinion that a POSA would have also reasonably expected to successfully combine the system resulting from the combination of Billingham and Schröder with the teachings of Haddick. Specifically, in my opinion, a POSA would have recognized that the benefits of using 3D sound in the Billingham and Schröder videoconferencing system would have also been advantageous during a *telephone call*. Billingham expressly recognized that the use of “a wearable computer [to] be used to support collaboration” was “increasingly important as telephones incorporate more computing power and portable computers became more like telephones.” Ex. 1010, 1-2. Furthermore, Haddick explicitly describes using known surround sound techniques to create 3D sound and realize a direction of the sound. Ex. 1011, ¶655 (“a ‘virtual’ surround sound system that uses

the two ear buds to give the perception of 3D space (such as the SRS True Surround Technology)”). Although Haddick does not describe the details of such techniques, those were well-known in the field and are explicitly described by Billinghurst and Schröder.

D. Claims 9-14 and 16-20

227. As I explained in Section IX, it is my opinion that Billinghurst in view of Schröder teaches each and every limitation of claims 9-14 and 16-20. To the extent that Billinghurst does not explicitly disclose an *SLP that is one meter to two meters away from the called party* (claim 9) and an image *that appears at least one meter away* (claim 16), it is my opinion that these limitations would have been obvious in view of Zhou. Specifically, in my opinion, a POSA implementing the Billinghurst system would have found it obvious to localize the SLP at the position of the avatar of the remote collaborator around 1.5 meters away, as taught in Zhou. *Id.*

228. Zhou’s experiments were conducted, in part, “to study the impact of 3D sound on task performance and feeling of ‘human presence and collaboration.’” Ex. 1018, 1. As Zhou explains, “[i]n the first experiment the sound sources are always placed around 1.5 m away in the front of the listener.” *Id.*, 5. According to Zhou, at this distance, the listener’s depth perception improved significantly, and “the addition of 3D sound can be a very effective way to complement the user experiences

in AR environments.” *Id.*, 13-14, 26. Given that Zhou places the sound source 1.5 meters away from the listener, a POSA would have understood that implementing Billingham’s SLPs and avatars so that they appear 1.5 meters away from the user would have similarly enhanced depth perception and complemented the user experience. In my opinion, this is especially true, since it was well-known that traditional spatialization methods, such as HRTFs, are not as effective for distances less than a meter. *See, e.g.*, Ex. 1018, 5 (“[A] design problem of AAR [occurs] when the listener is very close to or even overlaid on the virtual sound source,” resulting in a “very strange auditory experience” and leads to “confus[ion].”) .

229. To the extent that Billingham does not explicitly disclose executing the *method during a telephone call involving a calling party and a called party*, it is my opinion that this would have been obvious in view of Haddick. Haddick is similar to Billingham in that it also discloses a wearable head mounted display device that provides both visual and audio sensory input for live communication with one or more other parties, such as in a videoconference setting. Haddick includes specific detail as to how a user might join that videoconference. Specifically, Haddick’s wearable “augmented reality eyepiece” with ear buds 120 (Ex. 1011, ¶11, Fig. 1, reproduced below) provides an incoming *telephone call* notification to the eyepiece worn by a *called party* (*id.*, ¶662, Fig. 7, reproduced below).

Declaration of Dr. Chris Kyriakakis in Support of
Petition for *Inter Partes* Review of USP No. 9,674,628

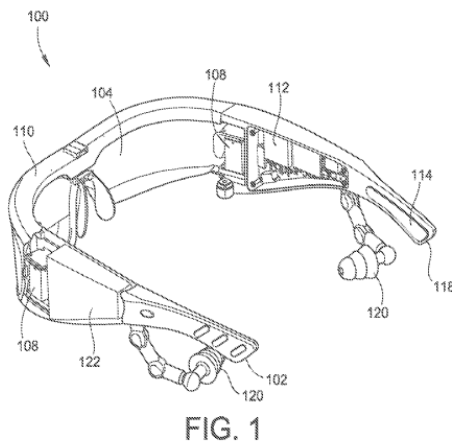


FIG. 1

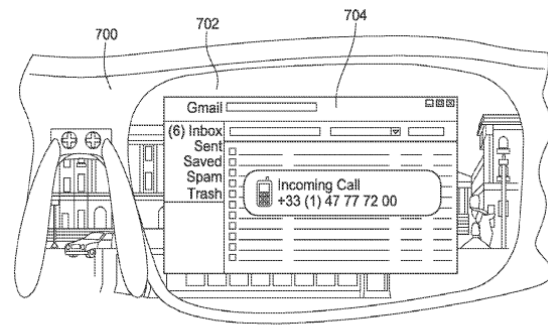
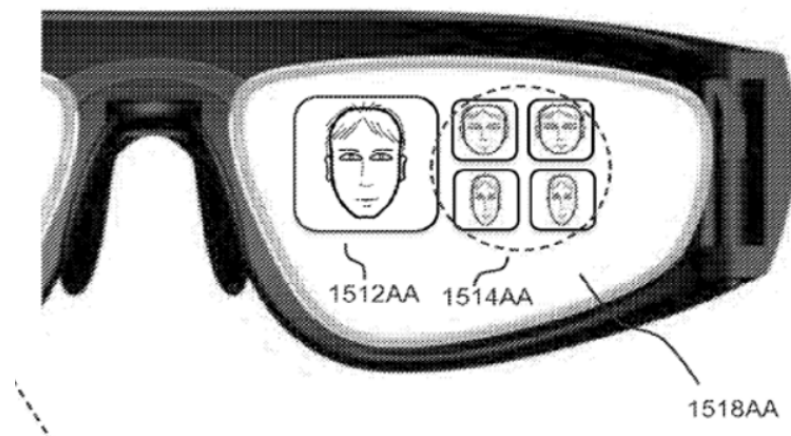


FIG. 7

230. In Haddick's system, upon answering the call, the *called party* joins a videoconference, and Haddick's eyepiece displays "images 1514AA of other participants on the videoconference call." *Id.*, ¶636, Fig. 15AA (I have reproduced below the relevant portion of Fig. 15AA); *see also id.*, ¶¶637, 730, 995.



231. According to Haddick, the eyepiece also receives "audio sensor input" as "the voice of a particular participant of the video conference." *Id.*, ¶730. Thus, it is my opinion that the POSA would have understood Billingham in view of Haddick to disclose executing the *method during a telephone call*.

**XI. GROUND 3: BILLINGHURST IN VIEW OF BALLAS RENDERS
OBVIOUS CLAIMS 9-14 AND 16-20**

A. Motivation to Combine Billinghurst with Ballas

232. For the same reasons that I already explained for Ground 1, it would have been obvious to a POSA to modify the CPU in Billinghurst to process audio streams using the HRTFs taught by Ballas. Specifically, a POSA would have been motivated to combine the teachings of Billinghurst and Ballas because in my opinion, a POSA would have recognized that Billinghurst's audio streams would have been spatialized using the conventional HRTF techniques described in Ballas. Although Billinghurst does not describe the spatialization techniques in explicit detail, a POSA would have understood that these details were described in the HRTF techniques in Ballas.

233. Ballas shares Billinghurst's goal of "measur[ing] the location and orientation of a user's head" to "produc[e] spatialized audio" to the user. Ex. 1008, ¶¶24-26. Indeed, Ballas's primary goal is to "maintain[] virtual location of the audio signals, thus allowing the listener to change locations and head orientation without degradation of the audio signal." *Id.*, ¶¶17-20. To accomplish this, Ballas expressly teaches that it applies "a pair of filters defined by a pair of head related transfer functions (HRTFs) for a particular location." *Id.*, ¶16. Thus, Ballas's use of HRTFs advanced the common spatialization goal of both Billinghurst and Ballas, and further

advanced the collaboration goal of Billinghurst. Moreover, as Ballas was a publication assigned to the U.S. Department of Navy, it is my opinion that the POSA would have been further motivated to look to the military's technology and approaches. Specifically, the POSA would have known that the military was a major investor into virtual reality systems, and when implementing Billinghurst's spatialized audio system, would have looked to the military for potential techniques and systems.

B. Reasonable Expectation of Success

234. In my opinion, a POSA would have reasonably expected to successfully combine the teachings of Billinghurst and Ballas. Both systems used virtual reality wearable devices that provide visual and auditory displays including a processor to process the sound output through headphones or speakers in the wearable device. Moreover, consistent with Billinghurst's videoconferencing system with head tracking, Ballas's system also includes a head tracking system that "measures the location and orientation of a user's head" and "produc[es] spatialized audio." Ex. 1008, ¶¶25-26. In Ballas, the sound is filtered with "a pair of head related transfer functions (HRTFs) for a particular location," thus producing spatialized audio, and the system therefore is able to "maintain[] the virtual location of the audio signals/sound, thus allowing the listener to change locations and head orientation

without degradation of the audio signal.” *Id.*, Abstract, *see also id.*, ¶¶16-20, claims 1, 7, 13.

235. Accordingly, it is my opinion that a POSA would have recognized that Billinghurst and Ballas are structurally and functionally similar. Specifically, each system uses a processor to adjust an audio stream to provide spatialized binaural sound. Ex. 1010, 5; Ex. 1008, ¶¶18-19, 27. Each system also tracks the head movement of the listener, and adjusts the audio stream to maintain the sound localization point. Ex. 1010, 4-5; Ex. 1008, Abstract, ¶¶17-20, claims 1, 7, 13. Therefore, the POSA implementing the Billinghurst system would have recognized that Billinghurst’s spatialized sound could be achieved by programming the Billinghurst processor with the HRTF functionality described in Ballas. Such an implementation would have merely involved a combination of known prior art elements (sound processor) according to known methods and techniques (HRTF filters) to yield predictable results (spatialized sound), and involves use of a known technique (spatialization) to improve a similar device (wearable communication device) in the same way (filtering sounds to introduce spatial cues).

C. Claims 9-14 and 16-20

236. As I already explained above for Ground 1, Billinghurst teaches nearly each and every limitation of claims 9-14 and 16-20, but does not explicitly disclose processing sound with *HRTFs* for localizing sound to a sound localization point. *See*

Section IX (Ground 1). In my opinion, this would have been obvious in view of Ballas.

237. Ballas discloses “producing spatialized audio signals,” including “producing virtual sound sources that are externally perceived and positioned at any orientation in azimuth and elevation from a listener.” Ex. 1008, Title, Abstract. Ballas’s system resembles Billinghamhurst’s wearable hardware in that Ballas discloses “a virtual reality head mounted visual display system,” such as an eyeglass frame or helmet. *Id.*, ¶25, *see also* Figs. 3, 4 (reproduced below).

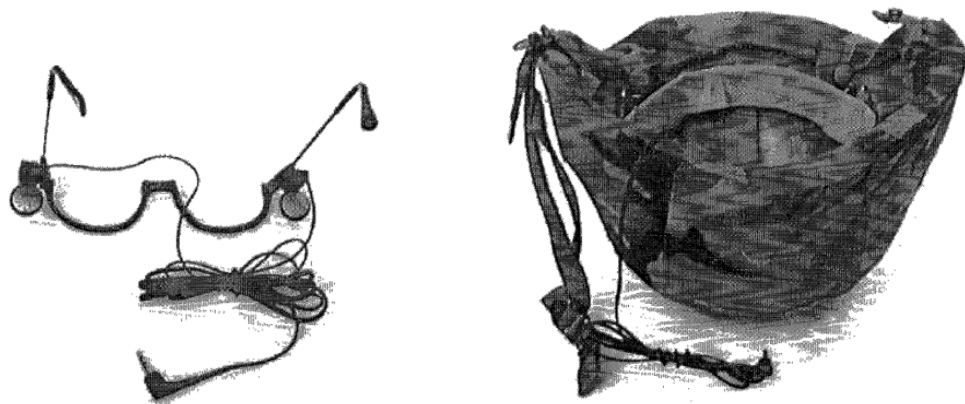


Figure 3. Eyeglass frame mount prototype Figure 4. Helmet mount prototype

238. In Ballas, the wearable device also resembles Billinghamhurst as it includes 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.” Ex. 1008, ¶25 (*see also* Abstract, ¶¶17-20, and claims 1, 7, 13).

239. Ballas provides explicit detail regarding how the spatialized audio is produced, by using HRTFs. Specifically, Ballas discloses “a pair of head related transfer functions (HRTFs)” (*HRTFs*) that are “for a particular location” (*at a sound localization point*). *Id.*, ¶16. In Ballas, the 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.

240. In my opinion, a POSA would have been motivated to implement Billinghurst’s wearable videoconferencing system to perform Ballas’s HRTF processing to provide binaural sound that is localized to the listener outside the listener’s head, as doing so would have provided a more realistic virtual experience over headphones for a listener by avoiding in-head localization consistent with Billinghurst’s goals. *See also* Section XI.A (Motivation to Combine Billinghurst with Ballas).

XII. GROUND 4: BILLINGHURST AND BALLAS IN VIEW OF ZHOU AND HADDICK RENDERS OBVIOUS CLAIMS 9-14 AND 16-20

241. As I already explained above, Billinghurst in view of Ballas teaches each and every limitation of claims 9-14 and 16-20. *See* Section XI (Ground 3).

242. For the same reasons as discussed for Ground 2, to the extent that Billinghurst does not explicitly disclose that the avatar’s position (*SLP*) is one meter

to two meters away from the listener (called party), it is my opinion that this would have been obvious in view of Zhou. *See* Section X (Ground 2).

243. For the same reasons as discussed for Ground 2, to the extent that Billingham and Ballas do not disclose that the method is performed *during a telephone call*, it is my opinion that this would have been obvious in view of Haddick. *See* Section X (Ground 2).

XIII. SECONDARY CONSIDERATIONS

244. I am not aware of any secondary considerations of non-obviousness for the claims of the '628 patent. If Patent Owner raises any secondary considerations in its briefing or expert's declaration, I reserve the right to respond to those considerations.

XIV. CONCLUSION

245. For these reasons, it is my opinion that claims 9-14, 16-20 of the '628 patent are obvious.

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



Dr. Chris Kyriakakis