

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

**PETITION FOR *INTER PARTES* REVIEW OF
U.S. PATENT NO. 9,674,628**

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	MANDATORY NOTICES	2
	A. Real Parties-in-Interest	2
	B. Related Matters	2
	C. Notice of Counsel and Service Information	3
	D. Fee for <i>Inter Partes</i> Review	4
	E. Certification of Grounds for Standing	4
III.	CHALLENGES UNDER 35 U.S.C. § 103	5
IV.	BACKGROUND	6
	A. Technology Overview	6
	B. The '628 Patent	10
	C. The Challenged Claims	12
	D. Prosecution History	12
V.	OVERVIEW OF THE PRIOR ART	13
	A. Billinghamst	13
	B. Schröder	15
	C. Zhou	16
	D. Haddick	17
	E. Ballas	18

VI.	LEVEL OF ORDINARY SKILL IN THE ART	19
VII.	CLAIM CONSTRUCTION	19
VIII.	GROUND 1: BILLINGHURST IN VIEW OF SCHRÖDER RENDERS OBVIOUS CLAIMS 9-14 AND 16-20	20
A.	Motivation to Combine Billinghamurst with Schröder	20
B.	Reasonable Expectation of Success	23
C.	Independent Claim 9	24
1.	9[pre]: Method That Improves Experience During A Telephone Call	24
2.	9[a]: Processing with HRTFs so Voice Localizes as Binaural Sound at SLP One to Two Meters Away	29
3.	9[b]: Display with WED an Image as at the SLP	38
4.	9[c]: Track Head Movements	40
5.	9[d]: Change HRTFs to Continue to Localize to SLP	43
D.	Dependent Claims 10-14	49
1.	Claim 10: Track Head Movements, Move Image, and Change HRTFs	49
2.	Claim 11: Track Head Movements and Change Voice	52
3.	Claim 12: Virtual Image	52
4.	Claim 13: Augmented Reality Image	54
5.	Claim 14: Moving Image and Changing HRTFs	55
E.	Independent Claim 16:	57
1.	16[pre]: Electronic System Executes Method During Telephone Call	57

2.	16[a]: Display with WED an Image At Least One Meter Away	58
3.	16[b]: Processing with HRTFs so Voice Localizes as Binaural Sound at Image.....	58
4.	16[c]: Track Head Movements	59
5.	16[d]: Change HRTFs to Continue to Localize to Image	59
F.	Dependent Claims 17-20:.....	60
1.	Claim 17: Track Head Movements and Change HRTFs	60
2.	Claim 18: Location of Image Functions As Virtual Microphone Point.....	60
3.	Claim 19: Change HRTFs to Move and Follow Image	63
4.	Claim 20: Provide Image at Virtual Microphone Point and Reduce Amplitude.....	63
IX.	GROUND 2: BILLINGHURST AND SCHRÖDER IN VIEW OF ZHOU AND HADDICK RENDERS OBVIOUS CLAIMS 9-14 AND 16-20	64
A.	Motivation to Combine Billinghamurst and Schröder with Zhou	64
B.	Motivation to Combine Billinghamurst and Schröder with Haddick	65
C.	Reasonable Expectation of Success	67
D.	Claims 9-14 and 16-20	68
X.	GROUND 3: BILLINGHURST IN VIEW OF BALLAS RENDERS OBVIOUS CLAIMS 9-14 AND 16-20	71
A.	Motivation to Combine Billinghamurst with Ballas	71
B.	Reasonable Expectation of Success	73
C.	Claims 9-14 and 16-20	74

XI.	GROUND 4: BILLINGHURST AND BALLAS IN VIEW OF ZHOU AND HADDICK RENDERS OBVIOUS CLAIMS 9-14 AND 16-20	76
XII.	SECONDARY CONSIDERATIONS	76
XIII.	THE BOARD SHOULD REACH THE MERITS OF THIS PETITION.....	77
XIV.	CONCLUSION.....	77

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Apple Inc. v. AGIS Software Development, LLC</i> , IPR2018-00819, Paper 9 (Nov. 7, 2018)	20
<i>Cienna Corp. v. Oyster Optics, LLC</i> , IPR2021-00565, Paper 11 (Aug. 23, 2021)	<i>passim</i>
<i>Eight KHZ, LLC v. Meta Platforms, Inc. et al.</i> , No. 2:22-cv-000575 (W.D. Tex.)	3
<i>Eight kHz, LLC v. Meta Platforms, Inc.</i> , No. 6:22-cv-00575-ADA-DTG, Dkt. Nos. 46, 48, 90 (W.D. Tex.)	77
<i>Western Digital Corp. v. SPEX Techs., Inc.</i> , IPR2018-0084, Paper 14 (April 25, 2018)	19
Statutes	
35 U.S.C. § 102(a)(1)	5, 6
35 U.S.C. § 102(a)(1)/(2)	5, 6
35 U.S.C. § 103	5, 6, 9
35 U.S.C. § 314(a)	77
Other Authorities	
37 C.F.R. § 42.10(b)	4
37 C.F.R. § 42.15(a)	4
37 C.F.R. § 42.24	82
37 C.F.R. § 42.24(a)(1)	82
37 C.F.R. § 42.24(a)(1)(i)	82
37 C.F.R. § 42.104(a)	4

37 C.F.R. § 42.104(b)	5
-----------------------------	---

EXHIBIT LIST

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”)
1003	Declaration of Dr. Kyriakakis (“Kyriakakis”)
1004	<i>Curriculum Vitae</i> of Dr. Kyriakakis
1005	RESERVED
1006	RESERVED
1007	RESERVED
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”)
1014	RESERVED
1015	Rumsey, Francis “Whose head is it anyway? Optimizing binaural audio.” J. Audio Eng. Soc., Vol. 59, No. 9, 2011 (“Rumsey”)
1016	RESERVED
1017	RESERVED

Exhibit	Description
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”)
1019	Declaration of Dr. Hsieh-Yee
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”)
1024	Moller, Henrik, et al. “Head-Related Transfer Functions of Human Subjects.” <i>J. Audio Eng. Soc.</i> , Vol. 43, No. 5, 1995, pp. 300-321 (“Moller”)
1025	Akeroyd, Michael. “The psychoacoustics of binaural hearing.” <i>International Journal of Audiology</i> , 45:sup1, 25-33 (2006) (“Akeroyd”)
1026	Middlebrooks, John, et al. “Sound Localization by Human Listeners.” <i>Annual Review of Psychology</i> , 1991. 42:135-39 (“Middlebrooks”)
1027	<i>Eight KHZ, LLC v. Meta Platforms, Inc. et al.</i> , 6:22-cv-575-ADA-DTG (W.D. Tex.), Plaintiff’s Claim Construction Reply Brief (ECF 92) (“8KHZ Claim Construction Brief”)
1028	Harma, Aki. “Ambient Human-to-Human Communication.” <i>Handbook of Ambient Intelligence and Smart Environments</i> , 2010, pp. 795-823 (“Harma”)
1029	Okada, Ken-ichi, et al. “Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design.” <i>Association of Computing Machinery</i> , 1994, pp. 385-393 (“Okada”)
1030	Nova, Nicolas. “A Review of How Space Affords Socio-Cognitive Processes during Collaboration.” <i>PsychNology Journal</i> , 2005 (“Nova”)
1031	RESERVED
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-

Exhibit	Description
	DTG (W.D. Tex.)
1033	U.S. Patent No. 257,453 to Ader (“Ader”)
1034	Olson, Harry. “Sound Reproducing Systems - Monaural, Binaural, Monophonic, and Stereophonic.” Audio, 1958 (“Olson”)
1035	Kyriakakis, Chris. “Fundamental and Technological Limitations of Immersive Audio Systems.” Proceedings of the IEEE, Vol. 86, No. 5, 1998 (“Kyriakakis 1998”)
1036	Begault, Durand. “3-D Sound for Virtual Reality and Multimedia.” 2000 (“Begault”)
1037	RESERVED
1038	RESERVED
1039	Grämer, T “Efficient modeling of head movements and dynamic scenes in virtual acoustics.” University of Zurich, Faculty of Medicin, 2010 (“Grämer”)
1040	RESERVED
1041	RESERVED
1042	RESERVED
1043	Laing, Gordon “Sound effects.” Personal Computer World, June 1997 (“Laing”)
1044	U.S. Patent No. 2,955,156 to Heilig (“Heilig”)

I. INTRODUCTION

Meta Platforms, Inc. (“Meta” or “Petitioner”) respectfully requests *inter partes* review of claims 9-14 and 16-20 of U.S. Patent No. 9,674,628 (EX1001, “’628 Patent”).

The foundational technology underlying virtual reality was developed in the 1960s with the Telesphere Mask, Headsight device, and Ultimate Display. As virtual reality advanced, so did sound simulation, particularly binaural sound. Humans perceive sounds binaurally because both ears receive auditory stimuli allowing the brain to compare the sound from each ear to determine the sound’s source location. By 2014, capturing, processing, and generating binaural sound so a listener localizes its origin at a sound localization point, or SLP (i.e., point from where sound originates) were well-known, well-understood, and routinely used.

The ’628 Patent simulates face-to-face conversation during a call by positioning one participant’s voice at an SLP near another participant and maintaining that SLP even when the other participant’s head moves. But rather than disclose some new innovation, the ’628 Patent merely describes known techniques during calls. Notably, nowhere does the ’628 Patent describe the difference between the claimed subject matter and prior art. Indeed, there is no patentable difference. For example, in 1998, Professor Mark Billinghurst had already developed a wearable electronic device (WED) that spatialized sound to SLPs to simulate face-to-face

conversations during a teleconference, and maintained the SLP by tracking the user's head movements.

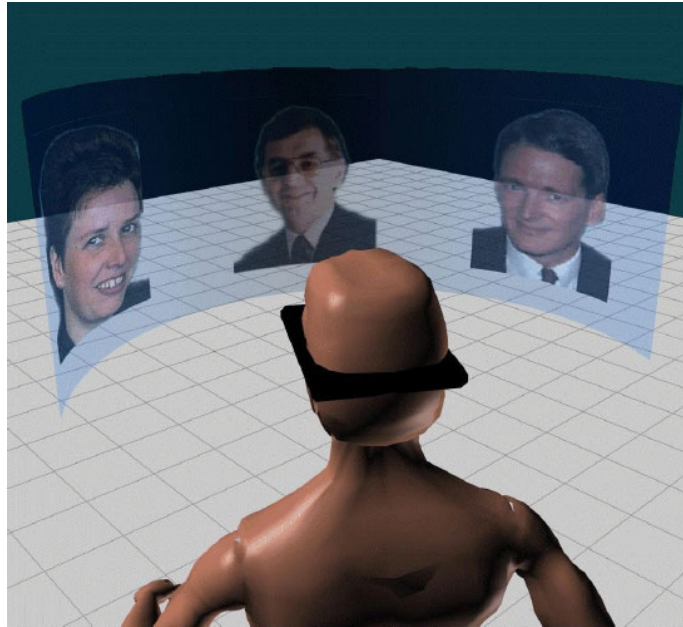


Figure 2.0 A Spatial Conferencing Space.

EX1010, Fig. 2

Accordingly, Petitioner respectfully requests that the Board institute review and find all challenged claims of the '628 Patent unpatentable.

II. MANDATORY NOTICES

A. Real Parties-in-Interest

The real parties-in-interest are Meta Platforms, Inc., Meta Platforms Technologies, LLC, and Twisted Pixel Games, LLC.

B. Related Matters

U.S. Patent Office records indicate that the '628 Patent is assigned to Eight KHZ, LLC ("PO"), which is currently asserting the '628 Patent in the following

U.S. Patent No. 9,674,628
Petition for *Inter Partes* Review

concurrent litigation filed on June 3, 2022: *Eight KHZ, LLC v. Meta Platforms, Inc. et al.*, No. 2:22-cv-000575 (W.D. Tex.).

Petitioner has filed, at substantially the same time that this Petition was filed, petitions for *inter partes* review against related family members U.S. Patent Nos. 9,226,090 and 9,282,196, both of which are also asserted in the concurrent litigation. Petitioner has also filed, at substantially the same time that this Petition was filed, petitions for *inter partes* review against the following other patents asserted in the concurrent litigation: U.S. Patent No. 10,448,184; U.S. Patent No. 11,172,316; U.S. Patent No. 11,290,836; U.S. Patent No. 10,368,179; U.S. Patent No. 10,798,509; and U.S. Patent No. 10,917,737.

C. Notice of Counsel and Service Information

Lead Counsel	Back-Up Counsel
Lisa K. Nguyen (Reg. No. 58,018) lisa.nguyen@allenoverly.com Postal & Hand-Delivery Address: Allen & Overy LLP 550 High Street Palo Alto, CA 94301 Telephone: (650) 388-1724	David M. Tennant (Reg. No. 48,362) david.tennant@allenoverly.com Postal & Hand-Delivery Address: Allen & Overy LLP 1101 New York Ave NW Washington, DC 20005 Telephone: (202) 683-3891 Grace I. Wang (Reg. No. 69,892) grace.wang@allenoverly.com Postal & Hand-Delivery Address: Allen & Overy LLP 1221 Ave of the Americas New York, NY 10020 Tel: (212) 756-1143 David A. Hubbard (Reg. No. 73,621)

	david.hubbard@allenoverly.com Postal & Hand-Delivery Address: Allen & Overy LLP 1221 Ave of the Americas New York, NY 10020 Tel: (212) 610-6357 Alan M. Billharz (Reg. No. 79,532) alan.billharz@allenoverly.com Postal & Hand-Delivery Address: Allen & Overy LLP 1101 New York Avenue, NW Washington, DC 20005 Tel: (202) 683-3862 Eric E. Lancaster (<i>pro hac vice</i> to be filed) eric.lancaster@allenoverly.com Postal & Hand-Delivery Address: Allen & Overy LLP 550 High Street Palo Alto, CA 94301 Telephone: (650) 388-1700
--	--

A Power of Attorney is being filed concurrently with this Petition in accordance with 37 C.F.R. § 42.10(b). Petitioner consents to electronic service by email, using the email addresses identified above.

D. Fee for *Inter Partes* Review

The Director is authorized to charge the fee specified by 37 C.F.R. § 42.15(a) to Deposit Account No. 60-4184.

E. Certification of Grounds for Standing

Petitioner certifies pursuant to 37 C.F.R. § 42.104(a) that the '628 Patent is

available for *inter partes* review and that Petitioner is not barred or estopped from requesting an *inter partes* review challenging the patent claims on the grounds identified in this Petition.

III. CHALLENGES UNDER 35 U.S.C. § 103

Grounds	Claims	Prior Art
1	9-14, 16-20	Billinghurst ¹ in view of Schröder ²
2	9-14, 16-20	Billinghurst in view of Schröder in

¹ Billinghurst, Mark, et al. “A wearable spatial conferencing space,” Digest of Papers. Second International Symposium on Wearable Computers. IEEE, 1998 (EX1010, “Billinghurst”) is prior art under at least 35 U.S.C. § 102(a)(1). *See* EX1019, ¶¶1-46.

² International Pub. WO 2005/032209 to Schröder (EX1012, “Schröder”) is prior art under at least 35 U.S.C. § 102(a)(1)/(2).

		further view of Zhou ³ and Haddick ⁴
3	9-14, 16-20	Billinghurst in view of Ballas ⁵
4	9-14, 16-20	Billinghurst in view of Ballas in further view of Zhou and Haddick

IV. BACKGROUND

A. Technology Overview

Wearable virtual reality technology dates back to the 1960s, when early pioneers invented head mounted displays (“HMDs”) to provide immersive and realistic simulated experiences. EX1003 (“Kyriakakis”) ¶¶38-42. The earliest WEDs had displays synchronized with sound. More sophisticated devices adjusted

³ Zhou, Zhiying, et al. “An experimental study on the role of software synthesized 3D sound in augmented reality environments,” *Interacting with Computers* 16.5 (2004): 989-1016 (EX1018, “Zhou”) is prior art under at least 35 U.S.C. § 102(a)(1). *See* EX1019, ¶¶47-72.

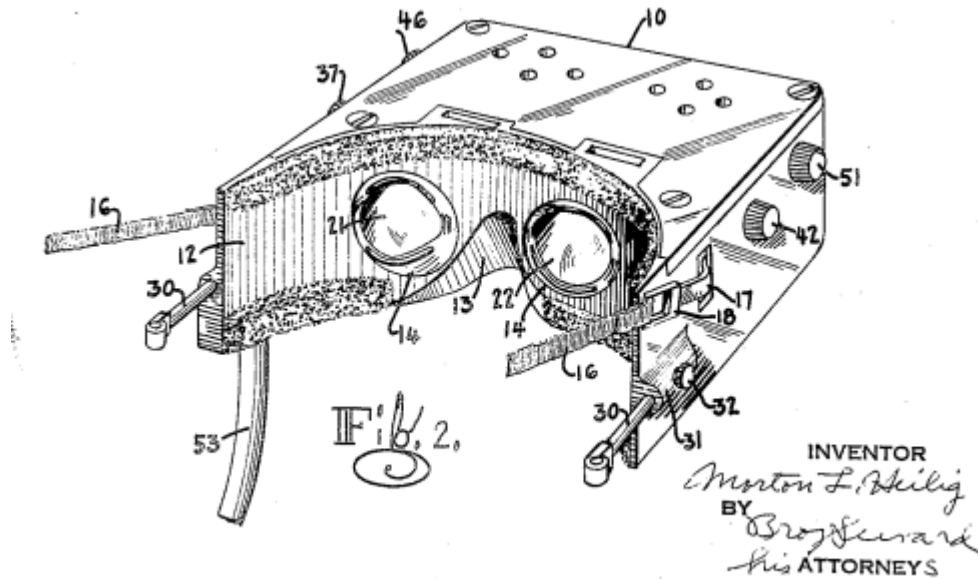
⁴ International Pub. No. WO 2013/049248 to Haddick, et al. (EX1011, “Haddick”) is prior art under at least 35 U.S.C. § 102(a)(1)/(2).

⁵ U.S. Patent Pub. No. 2003/0059070 to Ballas (“Ballas”) is prior art under at least 35 U.S.C. § 102(a)(1)/(2).

the displays and sound in response to tracking the wearer's head movements. *Id.*

¶¶38-42, 60-66.

In 1960, Morton Heilig patented the first HMD having earphones and a screen for each eye. EX1044, Fig. 2 reproduced below, *see also* EX1020; Kyriakakis ¶39.



In 1961, Charles Comeau and James Bryan developed “Headsight,” the first HMD to incorporate head tracking and provide images and sounds synchronized with the wearer's head movements. *Id.*, 5; EX1021, 3-7; Kyriakakis ¶40. Later, Ivan Sutherland described an HMD that 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.” EX1023, 1; *see also* EX1022, 2; Kyriakakis ¶¶41-42.

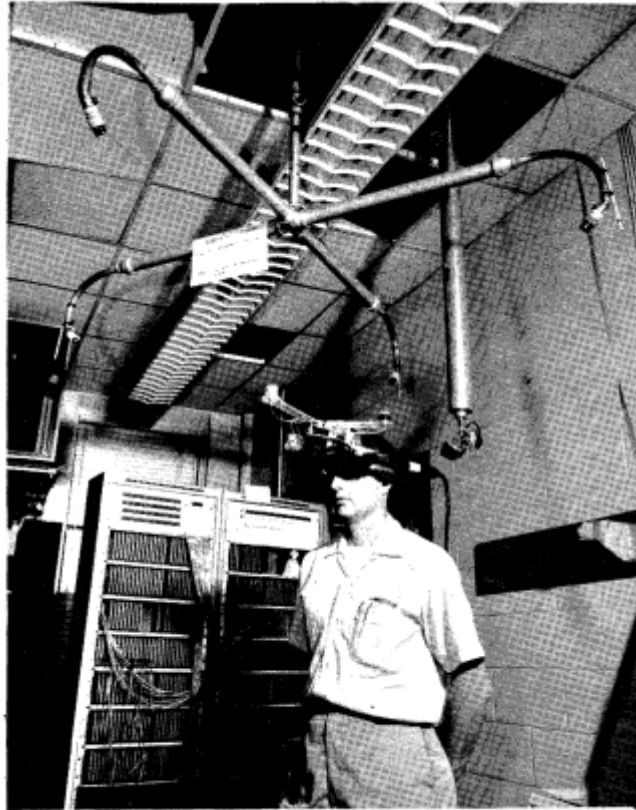


FIGURE 3—The mechanical head position sensor in use

As wearable HMD technology developed, so did binaural sound simulation. Kyriakakis ¶¶43-47. To determine sound source location, humans process and compare sound cues at each ear. *Id.* ¶¶48-54. For a sound source directly in front of the human listener, the sound generally arrives at each ear at the same time. *Id.* ¶50.

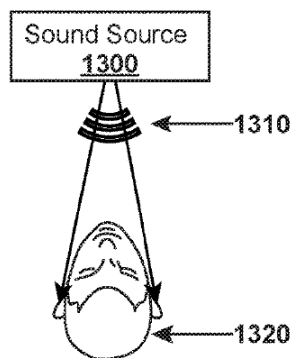


Figure 13A

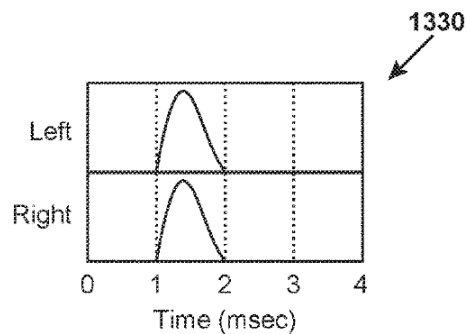


Figure 13B

EX1001, Figs. 13A-13B

But when the sound source is to the left, the sound arrives at the left ear slightly before the right ear, and the sound at the right ear is less intense. Kyriakakis ¶¶50-51. The difference in arrival time across the ears is the interaural time difference (ITD). *Id.*

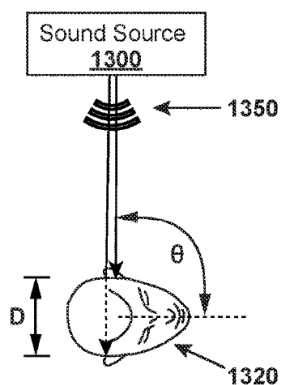


Figure 13C

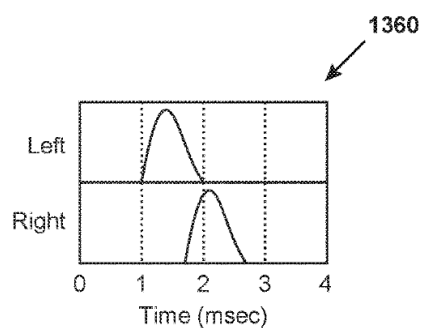


Figure 13D

EX1001, Figs. 13C-13D

Other factors impact how an ear receives sound: reverberation, reflection, refraction, attenuation. *Id.* ¶52. Even the shape of the listener’s ears, head, shoulders, and torso can change the sound. *Id.* ¶53.

To capture these factors, for decades, a head related transfer function (HRTF) has been used to characterize how an ear receives a sound from a point in space. EX1024, 1; EX1025, 25; Kyriakakis ¶52. To create a “virtual auditory space,” sounds are processed (convolved in the time domain) with a left HRTF and a right HRTF and played over headphones. EX1025, 25; Kyriakakis ¶¶55-59. As head tracking systems developed, the user’s head position was used in selecting “which transfer function should be used in the headphone presentation to maintain an apparently stable source location.” EX1026, 155; Kyriakakis ¶60.

B. The ’628 Patent

The ’628 Patent describes well-known concepts of providing sound “comparable in quality to sound that listeners would hear if they were communicating face-to-face with each other.” EX1001, 1:15-16. The system provides binaural sound of a person’s voice “to localize at an image during a telephone call.” *Id.*, Abstract, Fig. 9 (below); *see also id.*, 21:36-40; Kyriakakis ¶¶67-69.

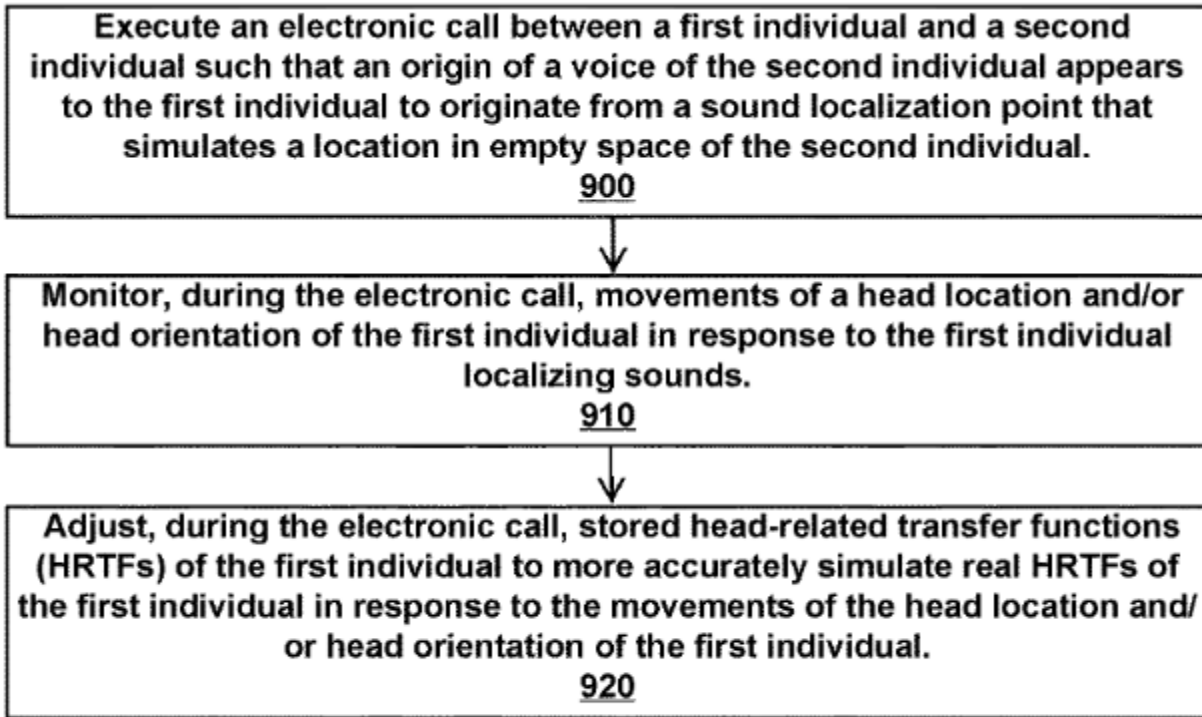


Figure 9

The WED tracks the listener's head movements of the listener relative to the SLP. EX1001, 21:51-54, 21:57-66. In response to the head movements, the system adjusts the voice to allow the listener to continue localizing the voice at the SLP. Kyriakakis ¶70.

The WED 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 SLP. EX1001, 33:22-24; Kyriakakis ¶71.

Neither the specification nor the file history describe how these embodiments improve upon the prior art. Notably, as discussed below, the prosecution history indicates that the Examiner allowed the wrong claims, and did not consider the

issued claims against any prior art. *See* Section IV.D (Prosecution History); Kyriakakis ¶72.

C. The Challenged Claims

The effective filing date of the challenged claims is no earlier than June 23, 2014.

D. Prosecution History

The prosecution history indicates the Examiner mistakenly allowed the wrong claims. There is no definitive evidence that the Office ever examined the issued claims.

The application for the '628 Patent was filed on February 23, 2017, with twenty original claims. EX1002, 1-116. 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.

Four days later, the Examiner addressed only the cancelled claims. On April 21, 2017, the Examiner signed a Notice of Allowability, and in the “Reasons for Allowance,” the Examiner indicated that claims 1-20 (previously cancelled by the

Applicant) were allowed and made no mention of the pending new claims 21-40.⁶ *Id.*, 149-150. Indeed, the Examiner's statement of reasons for allowance referred solely to the cancelled claims 1, 8 and 15 as requiring the following limitation:

... 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 [SLP] in e [sic] space that is away from and proximate to the listening person. (emphasis omitted).

Id. The examiner further stated: "These features, in combination with the other features of Claims 1, 8 and 15 are not anticipated by, nor made obvious over, the prior art of record." *Id.* But, the pending new claims 21-40 did not include the limitation quoted above.

The '628 Patent issued on June 6, 2017, with pending claims 21-40 renumbered as issued claims 1-20.

V. OVERVIEW OF THE PRIOR ART

A. Billinghamurst

Billinghurst generally discloses all limitations of the independent challenged claims, except the explicit use of HRTFs for localization (called spatialisation in

⁶ The Office mailed the Notice of Allowance on April 26, 2017. EX1002, 149-150.

Billinghurst). Kyriakakis ¶85. Billinghurst was not cited or discussed during the '628 Patent prosecution.

Billinghurst discloses systems and methods of providing sounds that remote users would hear if they were communicating face-to-face with each other. EX1010, 2 (“In collaborative virtual environments (CVEs) spatial visual and audio cues can combine in natural ways to aid communication.”). Billinghurst explains “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. (emphasis omitted); Kyriakakis ¶¶86-88.

Like the '628 Patent, during a call between users, the voice of the calling party appears to the called party to originate from a SLP that simulates a location in empty space of the calling party. EX1010, 5. “Each user determines the position and orientation of their own avatar in space, which changes as they move or look about [the] environment.” *Id.* Billinghurst 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 (emphasis omitted); Kyriakakis ¶89.

Billinghurst discloses that the wearable device further 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.*; Kyriakakis ¶90.

The Billinghurst 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. “With 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.*; Kyriakakis ¶91.

B. Schröder

Schröder discloses the use of HRTFs to localize binaural sound at an SLP. Specifically, Schröder teaches that spatialized sound (like taught in Billinghurst) can be generated using HRTFs. Kyriakakis ¶92. Schröder was not cited or discussed during the ’628 Patent prosecution.

Like the ’628 Patent and Billinghurst, Schröder generally relates to the reproduction of binaural sound using a wearable device. EX1012, 1:5-6 (“the binaural reproduction of aural events using headphones.”). Specifically, Schröder

discloses “a method for locating aural events, which are reproduced by headphones.” *Id.*, Abstract. Schröder teaches 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. The headphones in Schröder also “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; Kyriakakis ¶93.

C. Zhou

Zhou generally discloses using HRTFs to provide binaural sound that localizes the sound *at an SLP one or two meters away*. Kyriakakis ¶94. Zhou was not cited or discussed during the '628 Patent prosecution.

Like the '628 Patent, Zhou generally relates to use of 3D sound in augmented reality environments. EX1018, 1. Specifically, an objective of Zhou is to “understand the role of 3D sound in...human-human interactions.” *Id.* Zhou ultimately concludes that “3D sound contributes significantly to the feeling of ‘human presence and collaboration’ and helps the subjects to ‘identify spatial objects.’” *Id.*, 2. Zhou further concludes that 3D sound “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. Zhou reaches these conclusions through 3D sound experiments where “the sound sources are

always placed around 1.5 m away in the front of the listener.” *Id.*, 5; Kyriakakis ¶95.

D. Haddick

Haddick generally discloses a wearable HMD device that provides both visual and audio input for live communication with one or more other parties over a *telephone call*. Kyriakakis ¶96. Haddick was not cited or discussed during the ’628 Patent prosecution.

Like the ’628 Patent, Haddick discloses a wearable device that can “reproduce [a] 3D sound profile.” EX1011, ¶653. 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.*

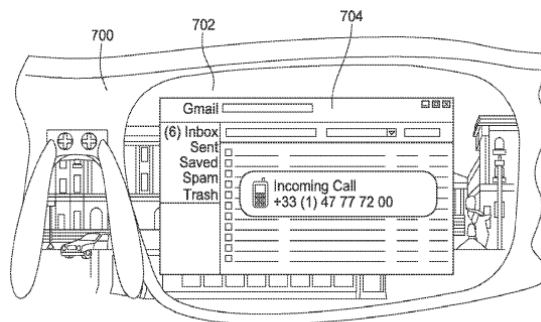


FIG. 7

When answering, Haddick’s wearer joins a videoconference call where “images 1514AA of other participants on the videoconference call.” *Id.*, ¶636, Fig. 15AA, portion reproduced above (*see also id.*, ¶¶637, 730, 995); Kyriakakis ¶97.

E. Ballas

Ballas discloses the use of HRTFs for localization not explicitly described in Billinghamurst. Specifically, Ballas teaches that spatialized sound (like taught in Billinghamurst) can be generated using HRTFs. Kyriakakis ¶98. Ballas was not cited or discussed during the '628 Patent prosecution.

Like the '628 Patent and Billinghamurst, Ballas teaches “[s]patialized audio 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.” EX1008, ¶2. Assigned to the U.S. Department of Navy, Ballas describes wearable devices (e.g., an eyeglass frame or helmet with mounted speakers) “mounted on the user’s head” that produces 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; Kyriakakis ¶99.

Notably, the wearable device in Ballas 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.” *Id.*, ¶¶17-18, 25-26. In Ballas, the use of the HRTF filter “maintains 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; Kyriakakis ¶99.

VI. LEVEL OF ORDINARY SKILL IN THE ART

A person of ordinary skill in the art (“POSA”) in 2014 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. Additional experience could substitute for less education, and additional education could likewise substitute for less experience. Kyriakakis ¶¶32-34.

This Petition does not turn on this precise definition, and the challenged claims would be unpatentable from the perspective of any reasonable person of ordinary skill in the art at the relevant time. *Id.* ¶¶35-37.

VII. CLAIM CONSTRUCTION

The prior art relied on in this Petition discloses the subject matter of the challenged claims under any reasonable construction, including their plain meaning. Notably, PO contends in district court that “a non-transitory computer readable storage medium” (i.e., Element 16[pre]),⁷ should be construed by its plain meaning and “[s]ection 112(f) does not apply.” EX1027, 24-38. Solely for this IPR, Petitioner adopts PO’s plain-meaning positions for such limitations.⁸ *See Western*

⁷ See Claim Listing Appendix for claim language and labels.

⁸ Petitioner reserves the right to argue alternative constructions in other proceedings, including that claims are indefinite.

Digital Corp. v. SPEX Techs., Inc., IPR2018-0084, Paper 14, at 10-12 (April 25, 2018) (petitioner not required “to express its subjective belief regarding the correctness of its proposed constructions”); *Apple Inc. v. AGIS Software Development, LLC*, IPR2018-00819, Paper 9, at 9-11 (Nov. 7, 2018); *Cienna Corp. v. Oyster Optics, LLC*, IPR2021-00565, Paper 11, at 11-13 (Aug. 23, 2021). Thus, no express construction is required.

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

A. Motivation to Combine Billinghamhurst with Schröder

It would have been obvious to a POSA to modify the CPU in Billinghamhurst to process audio streams using the HRTFs taught by Schröder. Kyriakakis ¶¶100-106. A POSA would have been motivated to combine the teachings of Billinghamhurst and Schröder because a POSA would have recognized that Billinghamhurst’s audio streams would have been spatialized using conventional HRTF techniques, such as those described in Schröder. *Id.*

At the relevant time period, a POSA would have understood that “[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)*.” EX1018, 3 (emphasis added). Thus, although Billinghamhurst does not detail the spatialization techniques used to localize the audio stream of the remote collaborators to their respective avatars, a POSA would have

known to look for such details in known HRTF techniques such as those taught by Schröder. Kyriakakis ¶101.

The primary goal of Billingham is to support collaboration between remote users. EX1010, 1 (“The problem we are interested in is how a wearable computer can be used to support collaboration between multiple people.”). To this end, Billingham identifies (1) visual and audio enhancements that can be used to aid communication, (2) how a collaborative communications space can be created between remote users, and (3) how remote users can be represented in such a space, all with a focus on spatialization using wearable hardware. *Id.*; Kyriakakis ¶102.

Spatialization is key because it emulates face-to-face interaction. Kyriakakis ¶103. “In collaborative virtual environments (CVEs) spatial visual and audio cues can combine in natural ways to aid communication.” EX1010, 2. Billingham emphasizes that “[e]ven a simple virtual avatar representation and spatial audio model enables users to discriminate between multiple speakers.” *Id.*, 3. As such, a POSA implementing the Billingham system would have looked to other references, such as Schröder, to spatialize the audio streams provided through the headphones of the Billingham wearable hardware such that the audio streams are spatialized in real time and emit from the corresponding avatar even when the users move. Kyriakakis ¶103.

Schröder shares this goal of spatialization so that a sound source over headphones appears constant in space even when the listener moves. Indeed, Schröder’s goal is also “to perceive 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.” EX1012, 5:5-10. Schröder further explains that an objective of its system and method is “to specify an arrangement for locating aural events such that they have a constant spatial direction using headphones.” *Id.*, 5:10-13; Kyriakakis ¶¶104-105.

Like Billinghamurst, Schröder contemplates that “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.” *Id.*, 4:13-17. Specifically, “[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, the use of HRTFs advanced the common spatialization goal of both Billinghamurst and Schröder, and further advanced the collaboration goal of Billinghamurst. Kyriakakis ¶106.

B. Reasonable Expectation of Success

A POSA would have reasonably expected to successfully combine the teachings of Billingham and Schröder. Kyriakakis ¶¶107-111. The systems of Billingham and Schröder both used wearable devices that have a processor to process and adjust audio streams to deliver binaural sound through headphones in the wearable device. In Billingham, the wearable computer “is a custom built 586 PC 104 based computer with 20mb of RAM running Windows 95.” EX1010, 4. Using the processor in the computer, the audio “is spatialized according to the distance and direction between speaker and listener.” *Id.*, 5; Kyriakakis ¶108. This means that “the speaker volume changes due to the sound spatialisation.” EX1010, 5.

Consistent with the Billingham system, in Schröder, “[a] listener 1 is wearing headphones 11 for binaural reproduction of aural events.” EX1012, 11:22-24. 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. “The control circuit HRTF 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; Kyriakakis ¶109.

Accordingly, a POSA would have recognized that Billingham and Schröder are structurally and functionally similar. Kyriakakis ¶110. Each system uses a processor to adjust an audio stream to provide spatialized binaural sound. EX1010, 5; EX1012, 15:17-21. Each system further tracks the head movement of the listener, and adjusts the audio stream to maintain the perceived position of the sound source (the SLP). EX1010, 4-5; EX1012, 12:3-24. Although Billingham does not explicitly disclose how the audio stream is spatialized, this detail is provided in Schröder. Kyriakakis ¶110. A POSA implementing the Billingham system would have recognized that the spatialized sound described in Billingham could be achieved by programming the Billingham processor with the HRTF functionality described in Schröder. *Id.* 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). *Id.* ¶111.

C. Independent Claim 9

1. 9[pre]: Method That Improves Experience During A Telephone Call

To the extent limiting, Billingham discloses the preamble. Kyriakakis ¶¶112-121.

Billinghurst discloses a *method* of implementing a “wearable spatial conferencing space.” EX1010, Title. The Billinghurst *method* is *executed by* wearable hardware (*one or more electronic devices*) in a videoconferencing system (*an electronic system*). Kyriakakis ¶113. Billinghurst shows an example of the “wearable hardware” in Fig. 3.0 (reproduced below), which includes a “wearable computer,” a display as “a pair of Virtual i-O iglasses! converted into a monoscopic display by the removal of the left eyepiece” and having a “inclinometer and a magnetometer used as a three degree of freedom orientation tracker,” and “a soundBlaster compatible soundboard with headmounted microphone.” EX1010, 4-5; Kyriakakis ¶113.



Figure 3.0 The Wearable Hardware.

Billinghurst explains that the “wearable conferencing space...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.”

EX1010, 4, Fig. 2.0 (reproduced below); Kyriakakis ¶114.

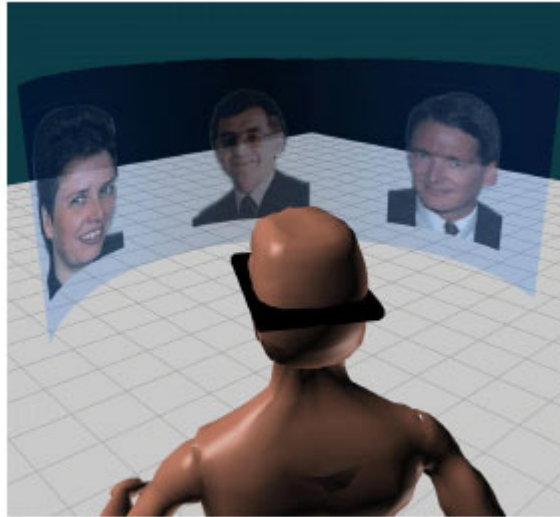


Figure 2.0 A Spatial Conferencing Space.

In Billingham’s videoconferencing system (*electronic system*), the wearable hardware (*one or more electronic devices*) works together to execute a *method* that *improves an experience* for users of a videoconference, because “[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 the other conversations taking place.” *Id.*; Kyriakakis ¶115. The system also *improves an experience* for videoconference participants by allowing “[t]he user [to] move about the space enabling them to choose their own viewpoint and the spatial relationships between the collaborators,” and by “allow[ing] the faces of remote users to appear life-size, a crucial factor for

establishing equal relationships in remote collaboration.” EX1010, 4; Kyriakakis ¶116.

Other *improvements to the experience* for videoconference participants in Billingham’s *electronic system* include: (1) enabling “[u]sers []to easily discriminate between three simultaneous speakers when their audio streams are spatialized,” (2) “allow[ing] [participants] to use some of the non-verbal cues commonly used in face-to-face communication such as gaze modulation and body motion,” which “[p]articipants preferred... to just hearing their speech,” (3) allowing users to “continue doing real world tasks while talking to collaborators in the conferencing space,” and (4) providing an “interface [that] is easy and intuitive to use,” and improved “head tracking on the wearable” compared to a “mouse only desktop interface.” EX1010, 6; Kyriakakis ¶117.

Billingham discloses a *telephone call* with a *calling party* and a *called party* *talking to each other*. EX1010, 2 (describing telephones with computing power and a conference phone call), *see also id.*, 5 (describing the wearable hardware supporting “3D spatialized Internet telephony”); Kyriakakis ¶118. Under PO’s position that a *telephone call* should mean “voice communication over an electronic network between at least two parties” (EX1027, 9), Billingham’s videoconferencing system meets a *telephone call* because it provides an interface for remote parties to use their voices to communicate over the Internet. Kyriakakis ¶118. Moreover, to

the extent that the *calling party* and *called party* include two parties participating in a conference call, the user or any of the remote collaborators in Billingham's videoconferencing system would have been the *calling party* or *called party*. *Id.*

In addition, Billingham contemplates a *telephone call* even under a different interpretation requiring a "call between at least two parties using at least one phone." EX1027, 9; Kyriakakis ¶119. Billingham recognized in 1998 that telephones and portable computer technologies were merging, as "telephones incorporate more computing power and portable computers become more like telephones...a conference phone call may be just as effective as [computer-mediated communication]." EX1010, 2. Therefore, even if *telephone call* were interpreted to require a call... using at least one phone, a POSA would have recognized that Billingham's teachings for a videoconference were applicable to a *telephone call* between the *calling party* (e.g., the remote collaborator in Billingham's system) and the *called party* (e.g., the user or wearer of Billingham's wearable hardware).⁹ Kyriakakis ¶119. It would have been a natural extension of Billingham to implement the wearable system during a call using a phone, as doing so would have

⁹ For clarity and despite being unnecessary under PO's interpretation, Petitioner maps the *calling party* to Billingham's remote collaborator and the *called party* to Billingham's user or wearer of the wearable hardware).

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. *Id.* Consequently, the POSA would have appreciated that Billinghurst discloses *a calling party and a called party talking to each other during a telephone call*. *Id.* ¶120.

Thus, Billinghurst 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 Billinghurst system (*of a calling party and a called party talking to each other during a telephone call*). *Id.* ¶121.

2. 9[a]: Processing with HRTFs so Voice Localizes as Binaural Sound at SLP One to Two Meters Away

Billinghurst-Schröder renders obvious this limitation. Kyriakakis ¶¶122-145.

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

As explained in 9[pre], Billinghurst teaches using its system *during a telephone call*. Kyriakakis ¶123; see Section VIII.C.1 (Ground 1, 9[pre]). Billinghurst discloses “a wearable conferencing space...that allows remote collaborators [to] appear as virtual avatars distributed about the user (figure 2.0).” EX1010, 4. As discussed in 9[pre], a POSA would have understood that the wearer

of the head mounted display in Fig. 2.0 is the *called party* and any remote collaborator is a *calling party*. Kyriakakis ¶123.

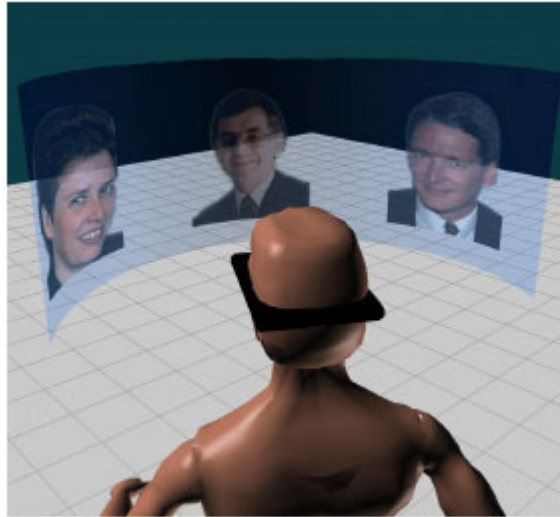


Figure 2.0 A Spatial Conferencing Space.

Further, Billinghamurst explains the auditory and positional relationship between the called and calling parties. *See* EX1010, 4 (explaining that 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.”). Indeed, a POSA would have understood that each audio stream in Billinghamurst’s conferencing space is a *voice* from a remote collaborator (i.e., represented by the virtual avatar). Kyriakakis ¶124.

A POSA would have also understood that the *voice* is provided as a *binaural sound* so that the sound *localizes* to the user *at a SLP* that corresponds to the position

of the avatar. *See, e.g.*, EX1010, 4, Fig. 2.0; Kyriakakis ¶124.¹⁰ Specifically, Billinghamurst explains that “current wearable computers use see-through or see-around monoscopic head mounted displays with stereo headphones.” EX1010, 3. “Each user determines the position and orientation of their own avatar in space.” *Id.*, 5. A POSA would have understood that when users connect to the Billinghamurst system, “their audio is broadcast to all the other users in the space” using the headphones, such that the audio “is spatialized according to the distance and direction between speaker [such as the remote collaborator who is the *calling party*] and listener [such as the user of the wearable hardware who is the *called party*]” as *binaural sound* that localizes the *calling party* at the avatar’s position (*SLP*). *Id.*; Kyriakakis ¶¶125-126.¹¹

¹⁰ *See* EX1032:176:3-5 (PO’s district court expert admitting it was known to provide binaural sound to listeners through earphones before 2014).

¹¹ Billinghamurst’s system has remote users that connect to the conferencing space, where any of the users are speakers or listeners. For clarity, this petition relies on Billinghamurst’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 Billinghamurst’s

A POSA would have also understood that Billinghurst discloses that the audio stream of the remote collaborator (*voice of the calling party*) is processed with a CPU (*with a processor*). Kyriakakis ¶129.¹² Billinghurst explains that the wearable

system can be the listener or speaker at any time during the videoconference.

Kyriakakis ¶¶127-128.

¹² As explained in Section VII, for the purposes of this IPR, Petitioner adopts PO's plain-meaning positions, including that section 112(f) does not apply. Under PO's understanding of the claims, Billinghurst's CPU is sufficient for performing this limitation. But even if limitation 9[a] were construed as a means-plus-function term, it would have been obvious over Billinghurst-Schröder because Schröder described in detail converting propagation time, amplitude, phase and frequency into HRTFs (EX1012, 2:3-12), and because PO's district court expert admitted that processing with HRTFs was well-known (EX1032, 59:4-16 (admitting there were "many references"—"in the 1990s"—"discussing using convolution processes to process head related transfer functions with digital signal processors for binaural audio"), 135:6-25 (admitting that before 2014, it was known 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).

hardware includes a wearable computer in the form of “a custom built 586 PC 104 computer with 20mb of RAM running Windows 95.” EX1010, 4. Billinghurst explains that when the audio is broadcast, “culling is also used so that only the audio streams from the speakers closest to the listener are broadcast and spatialized.” *Id.*, 5. A POSA would have understood that the CPU in the computer (*processor*) processes the audio stream of the speaker (*voice of the calling party*) given that Billinghurst explains that the culling “significantly reduces the CPU load.” *Id.*; Kyriakakis ¶129.

Thus, Billinghurst discloses *processing, during the telephone call*¹³ *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 (SLP).* Kyriakakis ¶130.

Billinghurst implemented its audio using “Microsoft’s Direct Sound technology” and “the Microsoft Direct Sound libraries” (EX1010, 6-7), which the POSA would have known to use HRTFs to position sound sources in 3D space. Kyriakakis ¶132 (explaining that a key feature of DirectSound was using HRTFs for

¹³ As explained for 9[pre], Billinghurst’s videoconference meets a *telephone call* under PO’s position.

positioning sound sources in 3D space and citing EX1043). Nonetheless, to the extent that Billinghurst does not explicitly disclose that the voice is processed *with head related transfer functions (HRTFs)* to localize at an SLP, this would have been obvious in view of Schröder. *Id.*, ¶¶131, 133.

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

Schröder relates to “the binaural reproduction of aural events using headphones.” EX1012, 1:5-6. Specifically, Schröder “relates to headphones which 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. Schröder recognizes that using “conventional headphones” means that “the problem of in-head localization frequently arises.” *Id.*, 1:12-14. According to Schröder, this means that “sound sources [are] mapped virtually inside the head,” whereas to be more realistic, “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.*, 1:14-19; Kyriakakis ¶¶134-135.

To achieve this, Schröder discloses in Fig. 5 (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.” EX1012, 3:5-14. Accordingly, “[t]he listener perceives the sound event as coming from

outside the head, illustrated in the figure by arrow 6.” *Id.*, 3:14-15. Schröder further discloses multiple HRTFs in the form of “variable HRTF filters” that are used to adjust the perceived direction and position of the aural event. *Id.*, 7:13-16, 12:20-24, 13:6-11, 14:16-21.

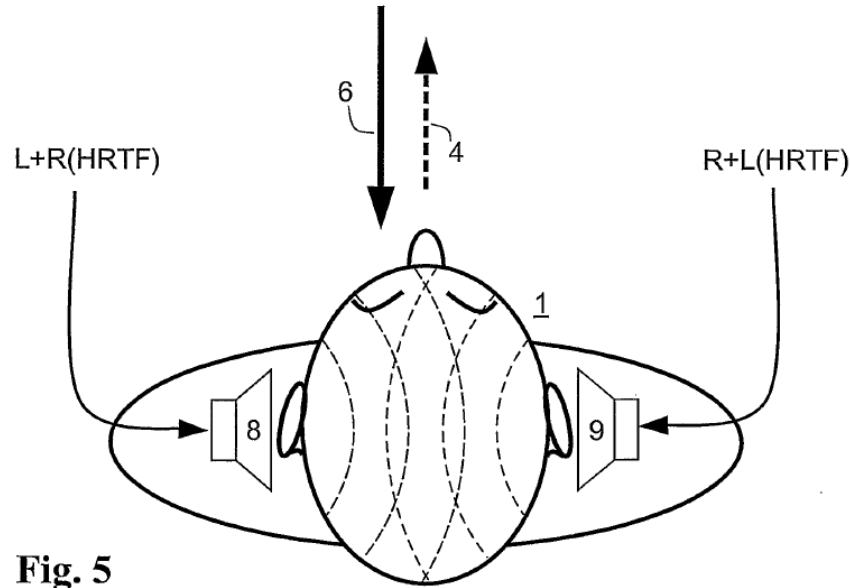


Fig. 5

Schröder’s teaching of “applying an HRTF” to the respective signals L, R discloses *processing, ...with a processor, a sound signal with multiple HRTFs (e.g., one left HRTF and one right HRTF, or variable HRTFs) so that the sound signal localizes to the listener (called party) from a specific direction and at a perceived position of the aural event (at an SLP).* Kyriakakis ¶¶135-136. Schröder 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.” EX1012, 12:20-24. As a POSA would have

understood, the HRTFs applied to the L and R signals must be different so that the sound is perceived to come from outside the head. Kyriakakis ¶136. Because two separate signals are provided over headphones to the left and right ears that cause the user to spatialize sound outside the head, Schröder relates to “binaural reproduction of aural events using headphones” and therefore provides *binaural sound*. See EX1012, 1:5-6.

A POSA would have been motivated to implement Billinghamhurst’s wearable videoconferencing system with Schröder’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 Billinghamhurst’s goals. Kyriakakis ¶137; see also Section VIII.A (Motivation to Combine Billinghamhurst with Schröder).

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

As Billinghamhurst 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. Kyriakakis ¶¶138-144 (explaining how EX1028 describes natural human-to-human communication as having a typical distance of “around one meter,” EX1029 describes the virtual distance during video teleconferencing as including 4 feet, which is 1.2 meters, and EX1030 describes 1.25 m as a distance for

conversations between friends and for impersonal business dealings).¹⁴ The audio streams in Billinghamurst are “spatialized in real time so that they appear to emit from the corresponding avatar.” EX1010, 4. Thus, the POSA would have readily appreciated that Billinghamurst teaches an *SLP that is one meter to two meters away from the called party* because, as was common in videoconferencing, the avatars appear at typical communication distances, including *one to two meters* away from the listener (*called party*), and the audio streams are spatialized to appear to emit from those avatars.

d) Conclusion

A POSA would have understood Billinghamurst-Schröder renders obvious *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*

¹⁴ PO’s expert in the parallel district court litigation 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.” EX1032, 126:5-128:11.

avatar's position (*SLP*) that is one meter to two meters away from the listener (*called party*). Kyriakakis ¶145.

3. 9[b]: Display with WED an Image as at the SLP

Billinghurst discloses this limitation. Kyriakakis ¶¶146-151.

Billinghurst discloses that the wearable hardware includes a display. EX1010, 4, Fig. 3.0 (reproduced below) (“Figure 3.0 shows a user wearing the display and computer.”). Billinghurst describes the “display” as “a pair of Virtual i-O iglasses! converted into a monoscopic display by the removal of the left eyepiece.” *Id.*; Kyriakakis ¶147.



Figure 3.0 The Wearable Hardware.

As depicted in Fig. 3.0 of Billinghurst, the monoscopic display is part of a set of wearable hardware that is a *WED* worn on a head of the listener (*called party*). In Billinghurst, the *WED display[s]* to the listener (*called party*) an image that

represents a remote collaborator (*calling party*), as depicted in Fig. 2.0 (reproduced below). Kyriakakis ¶148.

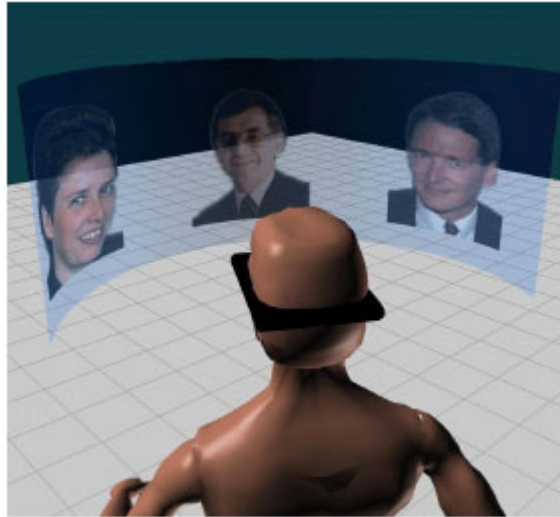


Figure 2.0 A Spatial Conferencing Space.

As explained in 9[a], Billingham's Fig. 2.0 depicts the images of the remote collaborators as being *one to two meters* away from the wearer. Kyriakakis ¶149; *see* Section VIII.C.2 (Ground 1, 9[a]). Moreover, Billingham aligns the voice and the image so that the wearer perceives both to originate from the same location, the *SLP*. EX1010, 4 (The audio streams are "spatialized in real time so that they to emit from the corresponding avatar."). Thus, Billingham 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' positions (*SLP*). Kyriakakis ¶149; *see also* EX1036, 95 ("most sounds are somehow connected to visual events"); EX1032, 244:16-245:12, 236:2-24 (PO's expert in the parallel district court litigation

describing co-locating an auditory object with a visual object in one location would improve the sense of motion as the object moves).

As explained in 9[pre], Billinghurst's videoconference is a *telephone call* under PO's interpretation, and includes displaying the avatar. Kyriakakis ¶150; see Section VIII.C.1 (Ground 1, 9[pre]).

Thus, a POSA would have understood Billinghurst discloses *displaying, during the telephone call and to the listener (called party) with a WED 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)*. Kyriakakis ¶151.

4. 9[c]: Track Head Movements

Billinghurst discloses this limitation. Kyriakakis ¶¶152-161.

Billinghurst discloses *tracking...head movements* of the user. *Id.* ¶¶153-156. As Billinghurst depicts in Fig. 1.0b (reproduced below), a "body-stabilized" display is one in which "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." EX1010, 3.

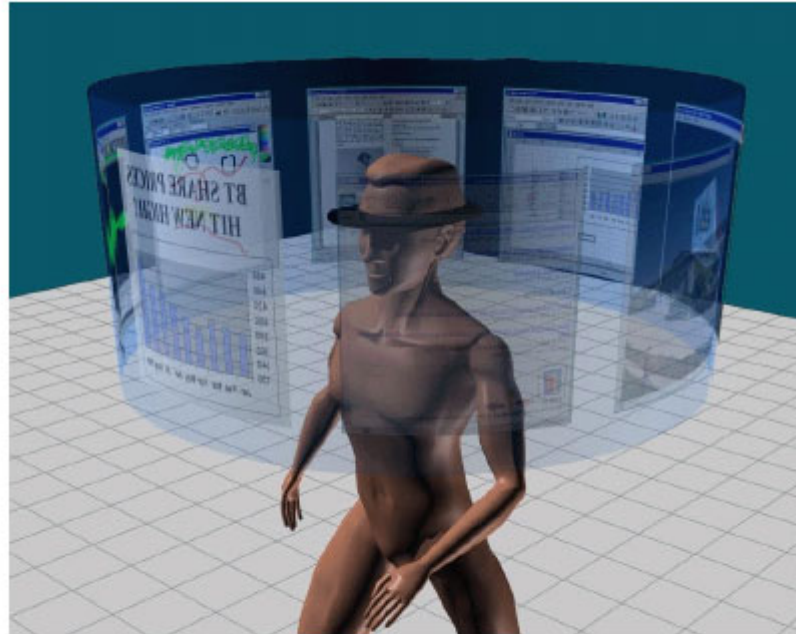


**Figure 1.0b One Degree of Freedom
Body-Stabilized Display**

Billinghurst further describes 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.* Billinghurst further discloses “[b]y tracking head orientation the user experiences a hemispherical information surround” and that 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.” *Id.*, 3-4. According to Billinghurst, such displays are “attractive” because they 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. Both “viewpoint orientation” and “viewpoint position” of the user change as the user’s head moves.

Therefore, by *tracking* one or both, either of Billingham's "body-stabilized" or "world-stabilized" displays meets this limitation because each involves *tracking head movements*. Kyriakakis ¶¶153-156.

In Billingham, the *tracking* of the *head movements* is performed with the *WED*, which includes the *iglasses! Display*. In Billingham, this display has "a sourceless two-axis inclinometer and a magnetometer used as a three degree of freedom orientation tracker." EX1010, 5. The combination of the inclinometer and the magnetometer in the Billingham wearable hardware allows the system to track the head movements of the listener wearing the wearable hardware. Kyriakakis ¶157.

In Billingham, the *track[ed] ...head movements* of the user are *with respect to the image* of the remote collaborator because the user "can simply turn to face the speakers they are interested in." EX1010, 5. "Just 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. By providing a body-stabilized or world-stabilized display that enables the user to turn to face the avatar of a collaborator, the avatars appear "fixed relative to the users [sic] body position and varies as the user changes viewpoint orientation," such that the *head movements* are *with respect to the image*. Kyriakakis ¶158. In Billingham, either teaching of the

body-stabilized display or world-stabilized display meets this limitation because each display varies as the user's head moves. EX1010, 4.

As explained in 9[pre], Billinghurst's videoconference is a *telephone call*, at least under PO's interpretation, and it further includes tracking the head movements of the listener. Kyriakakis ¶159; *see* Section VIII.C.1 (Ground 1, 9[pre]). As explained in 9[a] and 9[b], Billinghurst's avatar's position is an *SLP* because the sound appears to emit from the avatar. Kyriakakis ¶160; *see* Sections VIII.C.2-3 (Ground 1, 9[a]-9[b]).

Thus, a POSA would have understood Billinghurst discloses *tracking, during the telephone call and with the WED, 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)*. Kyriakakis ¶161.

5. 9[d]: Change HRTFs to Continue to Localize to SLP

Billinghurst-Schröder renders obvious this limitation. Kyriakakis ¶¶162-175.

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

Billinghurst discloses that 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). Kyriakakis ¶¶163-167; *see* Sections VIII.C.2-4 (Ground 1, 9[a]-9[c]).

As explained in 9[pre], Billinghamhurst’s videoconference is a *telephone call* at least under PO’s interpretation, and it further includes changing the processing of the audio streams. Kyriakakis ¶164; *see* Section VIII.C.1 (Ground 1, 9[pre]).

As discussed in 9[c], Billinghamhurst discloses a “body-stabilized” system and a “world-stabilized” system, each of which changes the images and audio streams based on the listener’s head movements. Kyriakakis ¶164; *see* Section VIII.C.4 (Ground 1, 9[c]).

In Billinghamhurst’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.” EX1010, 3. Upon connecting to Billinghamhurst’s “body-stabilized” system, the audio streams 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. Therefore, Billinghamhurst’s body-stabilized system varies the display as the user’s head movement changes. Kyriakakis ¶165.

Similarly, in Billinghamhurst’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.” EX1010, 4. 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.* Therefore, Billinghurst’s world-stabilized system also varies the display as the user’s head movement changes. Kyriakakis ¶166.

Either system (“body-stabilized” system or “world-stabilized” system) therefore uses spatialized audio streams that change while the user moves. *Id.* ¶¶165-166. Accordingly, either system in Billinghurst continues to localize the voice of the remote collaborator at the avatar of that remote collaborator while the head of the listener moves. *Id.*

Billinghurst does not expressly provide detail regarding how *the voice continues to localize to the SLP*. However, PO’s expert in the parallel district court litigation admitted that it was known in the 1980s for “head-track virtual acoustic systems [to] keep a sound source fixed in a location in virtual space independent of head movement.” EX1032, 250:2-9; *see also* EX1036, 12, 99, 132-34, Figs. 4.37-4.38. Moreover, Schröder expressly teaches this detail. Kyriakakis ¶167.

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

Schröder processes sound signals over headphones so that “they have a constant spatial direction.” EX1012, Title. Specifically, Schröder depicts in Fig. 7 “applying an HRTF” to process sound signals provided to the left and right ears so that the “listener perceives the sound event as coming from outside the head, illustrated in the figure by arrow 6.” *Id.*, 3:12-14.¹⁵

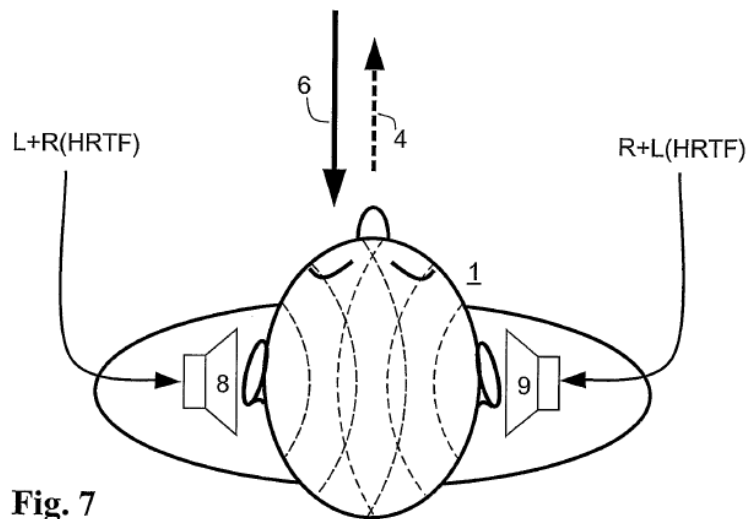


Fig. 7

As Schröder illustrates in Fig. 8 (reproduced below), when the user “change[s] his position,” the “aural event is perceived as coming from the original direction 6.” *Id.*, 4:12-25.

¹⁵ The quoted text of Schröder is in reference to Fig. 5, which is identical to Fig. 7.

See EX1012, 4:18-19 (stating “Figure 7 corresponds to Figure 5.”).

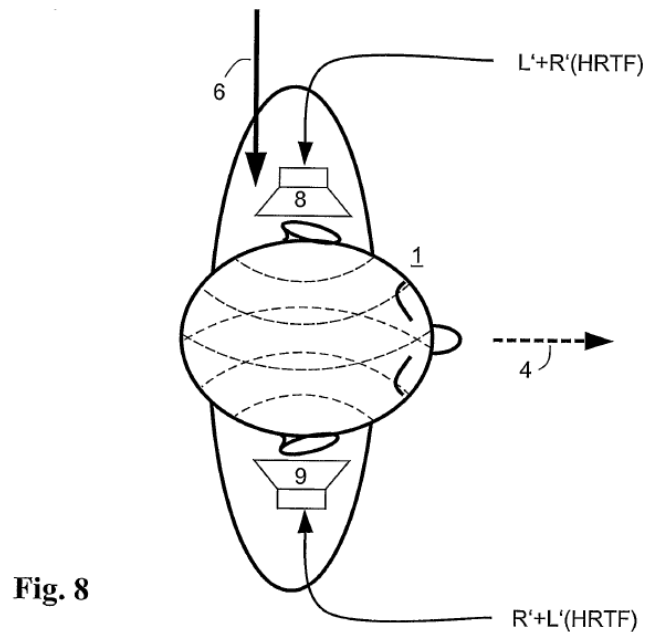


Fig. 8

According to Schröder, the sound appears to remain spatially fixed even when “the listener turns his head.” *Id.*, 5:27-30. To implement the “constant spatial direction” for the sound, Schröder discloses “[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, Schröder states “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, Schröder discloses *changing, in response to the head movements of a listener, the HRTFs processing the sound so that the sound continues to localize to the SLP.* Kyriakakis ¶¶168-171.

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 while the user moves their head. *Id.* ¶172. A POSA would have used Schröder's teaching of changing HRTFs in Billinghamurst's system, as HRTFs were well-known and commonly used in the art to manipulate the perceived origination of sound. *Id.*

c) Conclusion

Thus, a POSA would have understood Billinghamurst 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).* Kyriakakis ¶¶173-175.

D. Dependent Claims 10-14¹⁶

1. Claim 10: Track Head Movements, Move Image, and Change HRTFs

Billinghurst-Schröder renders obvious this limitation. *Id.* ¶¶176-181.

Billinghurst discloses *tracking, during the telephone call and with a WED worn on a head of the remote collaborator (calling party), the head movements of the remote collaborator (calling party).* *Id.* ¶177. Specifically, not only does the main user depicted in Billinghurst’s figures wear a set of wearable hardware, but also Billinghurst discloses that the “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.” EX1010, 4, 7. “In this way the conference space can support dozens of simultaneous users, similar to current collaborative virtual environments.” *Id.*, 4; *see also id.*, 1. As discussed in 9[c], the *WED 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 Billinghurst wearable hardware. *Id.*, 5; Kyriakakis ¶177; *see* Section VIII.C.4 (Ground 1, 9[c]).

¹⁶ All dependent claims incorporate the analysis of the claims from which they depend.

Billinghurst 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 (WED) worn on the head of the listener (called party), the head movements of the remote collaborator (calling party)*. Kyriakakis ¶178. Specifically, Billinghurst discloses that the “avatars could be live video streams.” EX1010, 4; *see also id.*, 7 (“We will incorporate live video texture mapping...enabling users to see their remote collaborators as they speak.”). Billinghurst teaches that 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.”). “In face-to-face conversations, speech, gesture, body language and other non-verbal cues combine to show attention and interest.” *Id.*, 2. 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.* As discussed in 9[b], the avatar, including live video, appears in *the display of the wearable hardware (WED) worn on the head of the listener (called party)*. Kyriakakis ¶178; *see* Section VIII.C.3 (Ground 1, 9[b]).

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. Kyriakakis ¶179. Specifically, “[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.” EX1010, 4. This spatialization is changed “according to the distance and direction between speaker and listener.” *Id.*, 5. “As users face...different speakers the speaker volume changes due to the sound spatialization.” *Id.* 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.

As explained in 9[d], Billinghurst does not expressly provide detail regarding *changing the HRTFs processing the voice of the calling party*, to allow the audio stream to *continue[] to localize to the SLP*, but this detail is expressly taught by Schröder. Kyriakakis ¶180; see Section VIII.C.5 (Ground 1, 9[d]).

As explained in 9[pre], Billinghurst's videoconference is a *telephone call* under PO's interpretation, and it further includes including tracking head

movements, moving images, and changing the processing of the audio stream. Kyriakakis ¶181; *see* Section VIII.C.1 (Ground 1, 9[pre]).

2. Claim 11: Track Head Movements and Change Voice

Billinghurst discloses this limitation. Kyriakakis ¶¶182-184.

As explained in claim 10, Billinghurst discloses *tracking, during the telephone call and with a WED worn on a head of the remote collaborator (calling party), the head movements of the remote collaborator (calling party)*. EX1010, 4 (“These remote users may also be using wearable computers.”); Kyriakakis ¶182; *see* Section VIII.D.1 (Ground 1, Claim 10).

As explained in claim 10, Billinghurst 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*. Kyriakakis ¶¶183-184; *see* Section VIII.D.1 (Ground 1, Claim 10).

3. Claim 12: Virtual Image

Billinghurst discloses this limitation. Kyriakakis ¶¶185-186.

The '628 Patent's only mention of a "virtual image" that is displayed is that a "[SLP] 2050 can appear to the listener 2020 as a virtual image of the speaker 2010." EX1001, 33:20-22. 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." Billingham's description of a live video stream therefore meets this limitation. Thus, as explained in 9[b] and claim 10, Billingham discloses *displaying, during the telephone call and to the listener (called party) with the WED worn on the head of the listener (called party), an avatar that is the live video stream of the remote collaborator (an image that represents the calling party as a virtual image of the calling party)*. EX1010, 4 ("The avatars could be live video streams"), Fig. 2 (reproduced below); Kyriakakis ¶186; see Sections VIII.C.3 and VIII.D.1 (Ground 1, 9[b] and Claim 10).

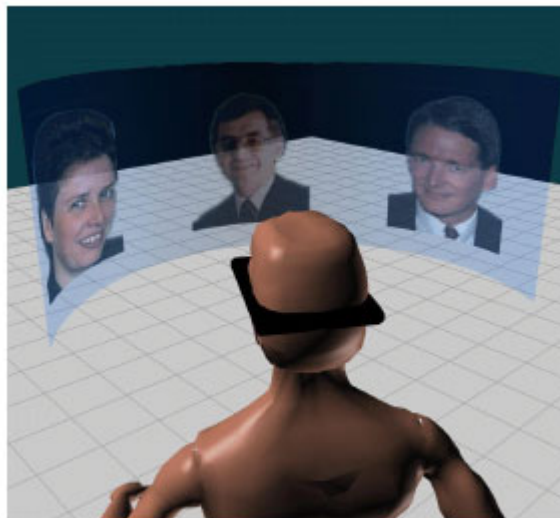


Figure 2.0 A Spatial Conferencing Space.

EX1010, Figure 2.0

4. Claim 13: Augmented Reality Image

Billinghurst discloses this limitation. Kyriakakis ¶¶187-188; *see* Section VIII.C.3 (Ground 1, 9[b]).

In Billinghurst, the wearable system is an “augmented reality communication space,” where the display hardware “Virtual i-O head mounted display” has 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.” EX1010, 1, 4-5; *see also id.*, 7 (“A user could, for example, choose to view avatars of conference participants overlaid and attached to a physical notice board.”). Thus, a POSA would understand that the Billinghurst avatars can be displayed over the real world in a “see-through mode,” and therefore include *displaying, during the telephone call and to the user (called party) with the wearable hardware (WED) 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)*. Kyriakakis ¶188; *see* Section VIII.C.3 (Ground 1, 9[b]).

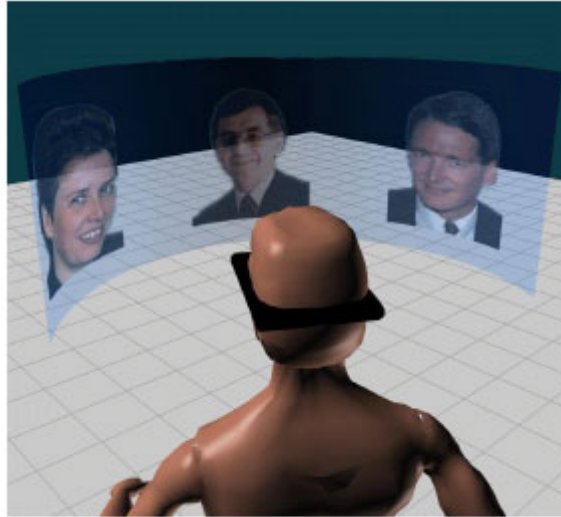


Figure 2.0 A Spatial Conferencing Space.

5. Claim 14: Moving Image and Changing HRTFs

Billinghurst-Schröder renders obvious this limitation. Kyriakakis ¶¶189-191.

Billinghurst 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.” EX1010, 6. “When 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 (reproduced below).

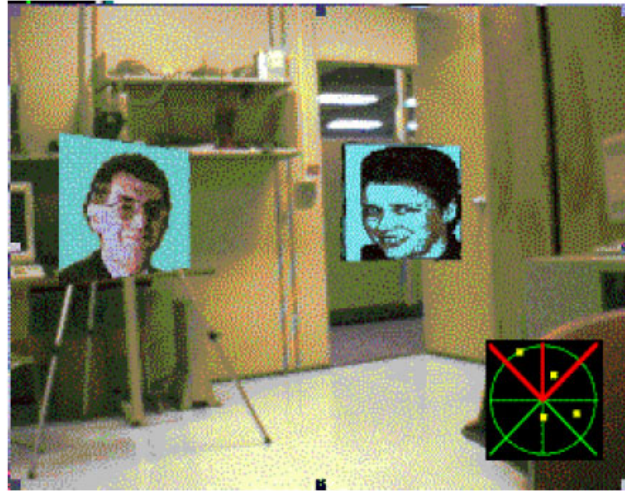


Figure 4.0 The User's View of the Wearable Conferencing Space

Billinghurst further explains that the audio stream will continue to localize to the avatar (*image that represents the calling party*) even as it *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; Kyriakakis ¶190.

As explained in 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 is expressly taught by Schröder. Kyriakakis ¶191; *see* Section VIII.C.5 (Ground 1, 9[d]). Therefore, 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. Kyriakakis ¶191; see Sections VIII.C.3, 5 (Ground 1, 9[b], 9[d]).

E. Independent Claim 16:

1. 16[pre]: Electronic System Executes Method During Telephone Call

To the extent limiting, Billingham discloses the preamble. Kyriakakis ¶¶192-194.

As discussed in 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)*. Kyriakakis ¶193; see Section VIII.C.1 (Ground 1, 9[pre]).

Billinghurst further discloses memory (*non-transitory computer readable storage medium*) storing instructions that the wearable hardware (*one or more electronic devices*) execute as the method. Specifically, Billingham discloses that the wearable computer “is a custom built 586 PC 104 based computer with 20mb of RAM running Windows 95.” EX1010, 4. A POSA would have understood that the

RAM is memory that also stores instructions to execute other programs and algorithms. Kyriakakis ¶194.¹⁷

2. 16[a]: Display with WED an Image At Least One Meter Away

Billinghurst discloses this limitation. Kyriakakis ¶¶195-196.

As discussed in 9[b], Billinghurst teaches *displaying, during the telephone call and to the listener (called party) with wearable hardware on the head of the listener (a WED 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)*. Kyriakakis ¶196; Section VIII.C.3 (Ground 1, 9[b]).

3. 16[b]: Processing with HRTFs so Voice Localizes as Binaural Sound at Image

Billinghurst-Schröder renders obvious this limitation. Kyriakakis ¶¶197-198.

As discussed in 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*

¹⁷ As explained in Section VII, for the purposes of this IPR, Petitioner adopts PO's plain-meaning positions, including that section 112(f) does not apply. Under PO's understanding of the claims, Billinghurst's RAM is sufficient for performing this limitation.

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*). Kyriakakis ¶198; Section VIII.C.2 (Ground 1, 9[a]).

4. 16[c]: Track Head Movements

Billinghurst discloses this limitation. Kyriakakis ¶¶199-200.

As discussed in 9[c], Billinghurst discloses *tracking, during the telephone call, head movements of the* listener (*called party*). Kyriakakis ¶200; Section VIII.C.4 (Ground 1, 9[c]).

5. 16[d]: Change HRTFs to Continue to Localize to Image

Billinghurst-Schröder renders obvious this limitation. Kyriakakis ¶¶201-202.

As discussed in 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*. Kyriakakis ¶202; *see* Section VIII.C.5 (Ground 1, 9[d]).

F. Dependent Claims 17-20:

1. Claim 17: Track Head Movements and Change HRTFs

Billinghurst-Schröder renders obvious this limitation. Kyriakakis ¶¶203-204.

As discussed in claim 10, Billinghurst-Schröder renders obvious *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.* Kyriakakis ¶204; see Section VIII.D.1 (Ground 1, Claim 10).

2. Claim 18: Location of Image Functions As Virtual Microphone Point

Billinghurst discloses this limitation. Kyriakakis ¶¶205-210.

According to the '628 Patent, “virtual microphone point” is “a virtual location or virtual position where sound is captured, recorded, or monitored.” EX1001, 42:33-35. In the '628 Patent, 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, 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; Kyriakakis ¶206. In Billinghamurst, the avatar’s position (*location of the image of the calling party*) functions as both an SLP and a *virtual microphone point*. Kyriakakis ¶207.

Specifically, because the local user and the remote collaborators in Billinghamurst’s system share the same virtual space (*see* bottom right corner of Fig. 4.0), the remote collaborator’s perception of the local user mirrors the local user’s perception of the remote collaborator. Kyriakakis ¶208; EX1010, 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.”).

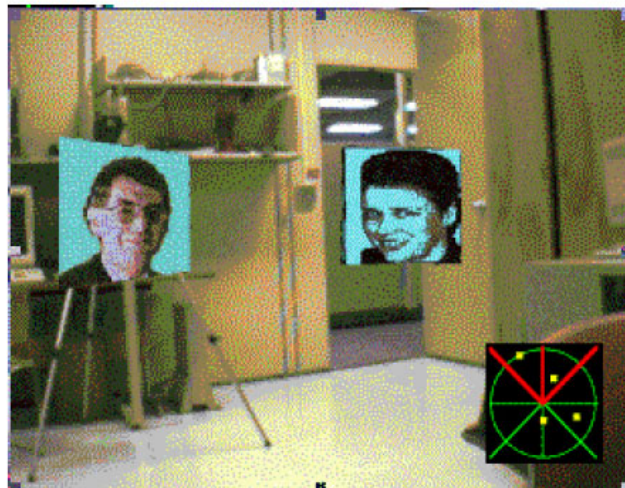


Figure 4.0 The User’s View of the Wearable Conferencing Space

Thus, in Billinghamurst, 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*). Kyriakakis ¶209. In this manner, Billinghamurst 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 (WED) worn on the head of the listener (called party). Id.*

Regarding *provid[ing] sound captured by the WED worn on the head of the called party... to the calling party*, Billinghamurst discloses this limitation. In Billinghamurst, the remote users have similar wearable hardware as the local user. EX1010, 4-5, 7 (“remote users may also be using wearable computers and head mounted displays” that have “headmounted microphone[s]”). Specifically, Billinghamurst explains that “[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 WED worn on the head of the called party*) is spatialized before being *provided to the remote collaborator (calling party). Id.*; Kyriakakis ¶210.

3. Claim 19: Change HRTFs to Move and Follow Image

Billinghurst-Schröder renders obvious this limitation. Kyriakakis ¶¶211-212.

As discussed in claim 14, Billinghurst 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)*. EX1010, 6; Kyriakakis ¶212; *see* Section VIII.D.5 (Ground 1, claim 14).

4. Claim 20: Provide Image at Virtual Microphone Point and Reduce Amplitude

Billinghurst discloses this limitation. Kyriakakis ¶¶213-215.

As discussed in claim 18, Billinghurst 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; Kyriakakis ¶214; *see* Section VIII.F.2 (Ground 1, Claim 18).

Billinghurst 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). EX1010, 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; Kyriakakis ¶215.

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

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)*, this would have been obvious in view of Zhou. Kyriakakis ¶216.

To the extent that Billinghurst does not disclose that the method is performed *during a telephone call*, this would have been obvious in view of Haddick. *Id.* ¶217.

A. Motivation to Combine Billinghurst and Schröder with Zhou

As discussed in Section VIII.A, a POSA would have been motivated to combine the teachings of Billinghurst and Schröder. Furthermore, a POSA would have been motivated to further modify the system resulting from the Billinghurst-Schröder combination to localize the SLP to be 1.5 meters away from the listener as taught by Zhou. *Id.* ¶¶218-219. Specifically, Zhou improved upon Billinghurst’s work and found that placing a 3D sound at an SLP having a distance of 1.5 meters significantly improved the user’s depth perception of an augmented object. EX1018,

9; Kyriakakis ¶218. A POSA would have recognized that the teachings of Zhou regarding the distance of an SLP would have advanced the objective of both Billinghamurst and Schröder to simulate face-to-face collaboration. EX1018, 1 (“[T]o access the impact of 3D sound on depth perception in a single-camera AR environment.”).

Like Billinghamurst and Schröder, Zhou 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. Indeed, Zhou builds on the work of Billinghamurst—Zhou expressly cites to Billinghamurst as an example of an application that has “applied 3D sounds in AR to enhance user experiences,” describing Billinghamurst as “a prototype wearable communication space that uses spatial and visual audio cues to enhance communication between remote groups of people.” *Id.* A POSA would have recognized “that binaural cues play an important role in auditory distance perception for nearby sources.” *Id.*, 5; Kyriakakis ¶219. Thus, a POSA implementing the system resulting from the Billinghamurst and Schröder would have naturally looked at the distance in Zhou as a distance where binaural cues would be effective in enhancing the user experience of the listener, including for enhanced depth perception. EX1018, 5 (“In our first experiment the source sources are always placed around 1.5 m away in the front of the listener.”); Kyriakakis ¶219.

B. Motivation to Combine Billinghamurst and Schröder with Haddick

As discussed in Section VIII.A, a POSA would have been motivated to combine the teachings of Billinghurst and Schröder, and as explained in Section VIII.C.1, Billinghurst 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 Billinghurst, it would have been obvious in view of Haddick. Kyriakakis ¶¶220-223.

Specifically, Billinghurst 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.” EX1010, 1-2. Based on this teaching, the knowledge of the POSA, and the development of more portable, smaller, lighter, and cheaper hardware by 2014, the POSA would have been motivated to modify the Billinghurst-Schröder system for use during a *telephone call* as specifically taught by Haddick. Kyriakakis ¶¶221-222.

Like Billinghurst and Schröder, Haddick teaches a wearable device that can “reproduce [a] 3D sound profile” and “realize a direction of sound for the wearer.” EX1011, ¶653. “[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.* A POSA implementing the Billingham-Schröder system would 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*. Kyriakakis ¶223.

C. Reasonable Expectation of Success

A POSA would have reasonably expected to successfully combine the Billingham-Schröder system with Zhou. Kyriakakis ¶¶224-226. Specifically, a POSA implementing the Billingham-Schröder system would have had to make a design choice as to where to localize the SLP, 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. *Id.* ¶224. Indeed, even in Figure 2.0 in Billingham, the avatars where the audio streams are localized appear to be one to two meters from the listener. *Id.* 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. *Id.* ¶225 (describing EX1039 which measured HRTFs at a distance of 1.5 meters and describing EX1036). Furthermore, adjusting the audio streams to localize 1.5 meters away would just be a matter of programming the Billingham processor. *Id.*

A POSA would have also reasonably expected to successfully combine the Billinghurst-Schröder system with the teachings of Haddick. *Id.* ¶226. Specifically, a POSA would have recognized that the benefits of using 3D sound in the Billinghurst-Schröder videoconferencing system would have also been advantageous during a *telephone call*. *Id.* Billinghurst 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.” EX1010, 1-2. Further, Haddick describes using known surround sound techniques to create 3D sound and realize a direction of the sound. EX1011, ¶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 are described by Billinghurst and Schröder. Kyriakakis ¶226.

D. Claims 9-14 and 16-20

As discussed in Section VIII.C-F, 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), these limitations would have been obvious in view of Zhou. Kyriakakis ¶227. Specifically, a POSA would have implemented the Billinghurst system to localize

the SLP at the remote collaborator's position around 1.5 meters away as taught in Zhou. *Id.*

Zhou describes experiments conducted, in part, “to study the impact of 3D sound on task performance and feeling of ‘human presence and collaboration.’” EX1018, 1. Zhou explains that “[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 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. As the sound source in Zhou is placed 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. Kyriakakis ¶228. This is particularly so, since it was well-known that traditional spatialization methods, such as HRTFs, are not as effective when the virtual sound source is very close to the listener. *See, e.g.*, EX1018, 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].”); Kyriakakis ¶228.

To the extent that Billingham does not explicitly disclose executing the *method during a telephone call involving a calling party and a called party*, this

would have been obvious in view of Haddick. Similar to Billingham, Haddick discloses a wearable head mounted display device that provides both visual and audio sensory input for live communication with one or more other parties. Kyriakakis ¶229. Specifically, Haddick discloses a wearable “augmented reality eyepiece” with ear buds 120 (EX1011, ¶11, Fig. 1, reproduced below) that provides an incoming *telephone call* notification to the eyepiece worn by a *called party* (*id.*, ¶662, Fig. 7, reproduced below).

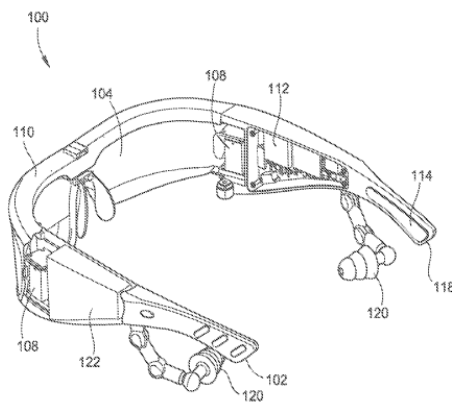


FIG. 1

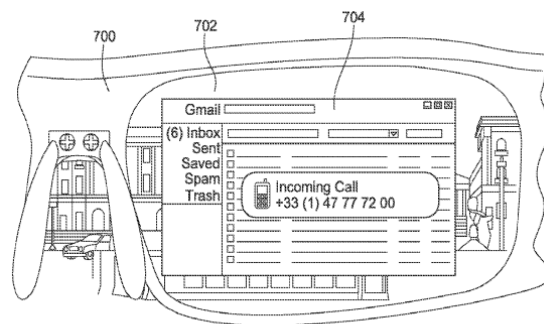
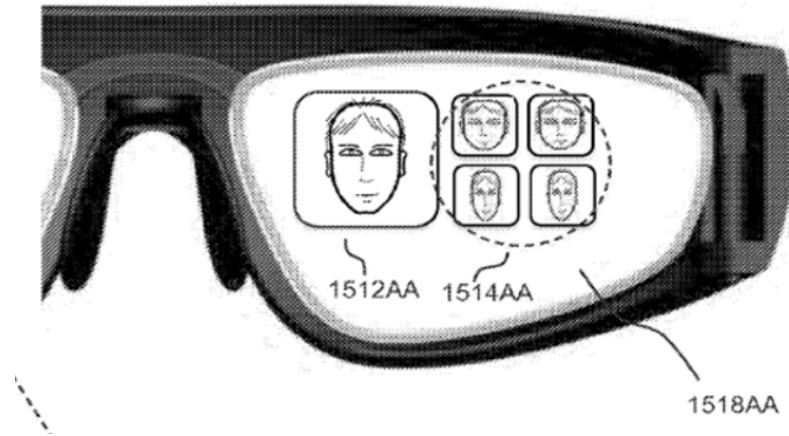


FIG. 7

To the extent the *telephone call* is a call between at least two parties using at least one phone, Haddick’s *telephone call* notification to the *called party*’s eyepiece from the *calling party* discloses this limitation. Kyriakakis ¶229. Upon answering the call, the *called party* joins a videoconference, and Haddick’s eyepiece displays “images 1514AA of other participants on the videoconference call.” EX1011, ¶636, Fig. 15AA, portion reproduced below (*see also id.*, ¶¶637, 730, 995); Kyriakakis ¶230.



The eyepiece also receives “audio sensor input” as “the voice of a particular participant of the video conference.” EX1011, ¶730. Thus, the POSA would have understood Billinghurst in view of Haddick to disclose executing the *method during a telephone call*. Kyriakakis ¶231.

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

A. Motivation to Combine Billinghurst with Ballas

It would have been obvious to a POSA to modify the CPU in Billinghurst to process audio streams using the HRTFs taught by Ballas. Kyriakakis ¶¶232-233. A POSA would have been motivated to combine the teachings of Billinghurst and Ballas because a POSA would have recognized that Billinghurst’s audio streams would have been spatialized using conventional HRTF techniques, such as those described in Ballas. *Id.* ¶232. At the relevant time period, a POSA would have understood that “[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)*.” EX1018, 3. Thus, although Billinghamurst does not describe in detail the 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 the HRTF techniques in Ballas. Kyriakakis ¶232. Although Billinghamurst does not describe the spatialization techniques in detail, a POSA would have understood that these details were described in the HRTF techniques in Ballas. *Id.*

Like Billinghamurst, Ballas shares this goal of “measur[ing] the location and orientation of a user’s head” to “produc[e] spatialized audio” to the user. EX1008, ¶¶24-26. Indeed, Ballas’s 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 do so, Ballas applies “a pair of filters defined by a pair of head related transfer functions (HRTFs) for a particular location.” *Id.*, ¶16. Thus, the use of HRTFs advanced the common spatialization goal of both Billinghamurst and Ballas, and further advanced the collaboration goal of Billinghamurst. Kyriakakis ¶233.

Further, Ballas was a reference assigned to the U.S. Department of Navy. EX1008, (76). A POSA would have been motivated to look at the technology and the approaches implemented by the military. Kyriakakis ¶233. By the relevant time period, a POSA would have understood that the military was one of the major

investors into virtual reality, and a POSA implementing the spatialized audio of Billinghamhurst would have looked to the military for potential techniques. *Id.*

B. Reasonable Expectation of Success

A POSA would have reasonably expected to successfully combine the teachings of Billinghamhurst and Ballas. Kyriakakis ¶¶234-235. The systems of Billinghamhurst and Ballas both 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. Consistent with Billinghamhurst, Ballas’s system includes a head tracking system that “measures the location and orientation of a user’s head” and “produc[es] spatialized audio.” EX1008, ¶¶25-26. Ballas discloses filtering a sound with “a pair of head related transfer functions (HRTFs) for a particular location,” thus producing spatialized audio and “maintain[ing] 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, ¶¶16-20, claims 1, 7, 13.

Accordingly, a POSA would have recognized that Billinghamhurst and Ballas are structurally and functionally similar. Kyriakakis ¶¶234-235. Each system uses a processor to adjust an audio stream to provide spatialized binaural sound. EX1010, 5; EX1008, ¶¶18-19, 27. Each system further tracks the head movement of the listener, and adjusts the audio stream to maintain the SLP. EX1010, 4-5; EX1008,

Abstract, ¶¶17-20, claims 1, 7, 13. A POSA implementing the Billingham system would have recognized that Billingham's spatialized sound could be achieved by programming the Billingham processor with the HRTF functionality described in Ballas. Kyriakakis ¶235. 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). *Id.*

C. Claims 9-14 and 16-20

As discussed above for Ground 1, Billingham teaches nearly each and every limitation of claims 9-14 and 16-20, but does not explicitly disclose processing sound with *HRTFs* so that it localized at an SLP. *See* Section VIII (Ground 1). This would have been obvious in view of Ballas. Kyriakakis ¶236.

Ballas generally relates to “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.” EX1008, Title, Abstract. 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).

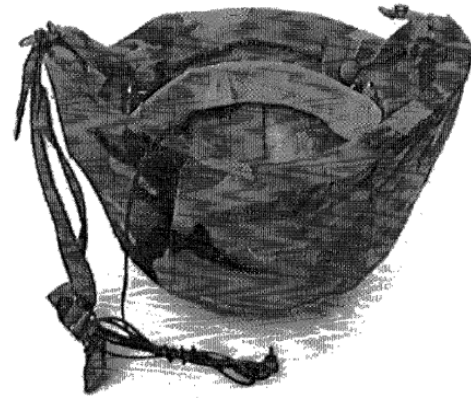


Figure 3. Eyeglass frame mount prototype Figure 4. Helmet mount prototype

The wearable device 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.” EX1008, ¶25 (*see also* Abstract, ¶¶17-20, and claims 1, 7, 13); Kyriakakis ¶¶237-238.

Ballas discloses “a pair of head related transfer functions (HRTFs)” (*HRTFs*) that are “for a particular location” (*at an SLP*). EX1008, ¶16. Specifically, Ballas’s HRTF filters “maintain[] [the] virtual location of the audio signals thereby allowing the user to change location and head orientation without degradation of the virtual location of audio signals.” *Id.*, Abstract, ¶¶17-20, claims 1, 7, 13; Kyriakakis ¶239.

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 Billingham's goals. Kyriakakis ¶240; *see also* Section X.A (Motivation to Combine Billingham with Ballas).

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

As discussed for Ground 3, Billingham in view of Ballas teaches each and every limitation of claims 9-14 and 16-20. *See* Section X (Ground 3); Kyriakakis ¶241.

For the same reasons as discussed for Ground 2, to the extent that Billingham does not explicitly disclose that the avatar's position (*SLP*) is *one meter to two meters away from the listener (called party)*, this would have been obvious in view of Zhou. *See* Section IX (Ground 2); Kyriakakis ¶242.

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*, this would have been obvious in view of Haddick. *See* Section IX (Ground 2); Kyriakakis ¶243.

XII. SECONDARY CONSIDERATIONS

There are no secondary considerations known to Petitioner that affect—let alone overcome—this strong case of obviousness. Should PO proffer any relevant evidence of secondary considerations in its preliminary response, Petitioner will seek leave to reply.

XIII. THE BOARD SHOULD REACH THE MERITS OF THIS PETITION

The Board should reach the merits of this petition and should not exercise its discretion to deny based on 35 U.S.C. § 314(a) or § 325(d). None of the prior art at issue in this petition was cited or relied upon during prosecution of the '628 Patent. As for the district court case, the court is currently deciding a motion to transfer. *See Eight kHz, LLC v. Meta Platforms, Inc.*, No. 6:22-cv-00575-ADA-DTG, Dkt. Nos. 46, 48, 90 (W.D. Tex.).

Petitioner reserves the right to address any discretionary issues raised by Patent Owner based on the facts and law as they stand at that time.

XIV. CONCLUSION

For these reasons, Petitioner respectfully requests *inter partes* review of claims 9-14 and 16-20 of the '628 Patent.

Date: June 9, 2023

Respectfully submitted,

By: /Lisa K. Nguyen/

Lisa K. Nguyen (Reg. No. 58,018)

lisa.nguyen@allenoverly.com

Allen & Overy LLP

550 High Street

Palo Alto, CA 94301

Telephone: 650.388.1724

*Counsel for Petitioner
Meta Platforms, Inc.*

CLAIM LISTING APPENDIX

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:

[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;

[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;

[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

[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.

10. 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;

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.

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.

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.

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.

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.

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:

[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;

[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;

[c] tracking, during the telephone call, head movements of the called party;
and

[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.

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.

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.

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.

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.

CERTIFICATE OF COMPLIANCE WITH 37 C.F.R. § 42.24

I hereby certify that this Petition complies with the word count limitation of 37 C.F.R. § 42.24(a)(1)(i) because the Petition contains a total of 13,950 words, calculated by Microsoft Word's word-count feature. This total excludes the cover page, signature block, and the parts of the Petition exempted by 37 C.F.R. § 42.24(a)(1).

Dated: June 9, 2023

By: /Lisa K. Nguyen/
Lisa K. Nguyen (Reg. No. 58,018)
lisa.nguyen@allenoverly.com
Allen & Overy LLP
550 High Street
Palo Alto, CA 94301
Telephone: 650.388.1724

*Counsel for Petitioner Meta
Platforms, Inc.*

CERTIFICATE OF SERVICE

The undersigned certifies that a complete copy of this Petition for *Inter Partes* Review of U.S. Patent No. 9,674,628 and all Exhibits and other documents filed together with this Petition were served on the official correspondence address for the patent shown in PAIR and a courtesy copy to Eight kHz, LLC's current litigation counsel via FEDERAL EXPRESS next business day delivery on June 9, 2023:

Philip S. Lyren
Philip S. Lyren, PC
289 Woodland Avenue
Wadsworth, OH 44281

By: /Lisa K. Nguyen/
Lisa K. Nguyen (Reg. No. 58,018)
lisa.nguyen@allenoverly.com
Allen & Overly LLP
550 High Street
Palo Alto, CA 94301
Telephone: 650.388.1724

*Counsel for Petitioner Meta
Platforms, Inc.*