

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

META PLATFORMS, INC.,
Petitioner,

v.

EIGHT KHZ, LLC,
Patent Owner.

Case IPR2023-01024
Patent 11,290,836 B2

**DECLARATION OF JOHN C. HART, PH.D.
IN SUPPORT OF PATENT OWNER'S PRELIMINARY RESPONSE**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. QUALIFICATIONS	1
III. SUMMARY OF OPINIONS	4
IV. RELEVANT LEVEL OF ORDINARY SKILL IN THE ART	4
V. The '836 Patent	5
VI. GROUND 1 OF THE PETITION: FINNEGAN AND BEGAULT	9
A. Finnegan	10
VII. GROUND 2 OF THE PETITION: NORRIS, BEGAULT, AND ALGAZI	13
A. Norris	13
VIII. GROUND 3 OF THE PETITION: SHIVAPPA, AND ZHOU	14
A. Shivappa	14
B. Zhou	15
IX. CONCLUSION	17

I, John C. Hart, declare as follows:

I. INTRODUCTION

1. I have been retained as an independent expert consultant in the *inter partes* review (“IPR”) proceeding involving the 11,290,836 Patent by Eight kHz LLC (“8kHz”). I will refer to this patent using common shorthand. I have been asked to consider the arguments raised by the Petition regarding the validity of the ‘836 Patent in this proceeding.

II. QUALIFICATIONS

2. I have been a professor of computer science, focusing on computer graphics including virtual reality systems, since 1992. I am currently a tenured Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign, a department consistently ranked in the top-5 by US News and World Report. I am also Director of the Office for Professional Education of the University of Illinois at Urbana-Champaign.
3. I received a Bachelor of Science degree in Computer Science from Aurora University in 1987, a Master of Science in Electrical Engineering and Computer Science from the University of Illinois at Chicago in 1989, and a Ph.D. in Electrical Engineering and Computer Science from the University of Illinois at Chicago in 1991.

4. I worked as an intern in 1998 at the IBM TJ Watson Research Center in Hawthorne, New York, and in 1999 at AT&T Pixel Machines in Holmdel, New Jersey. I worked as a postdoctoral research associate in 1991-1992 at the Electronic Visualization Laboratory at the University of Illinois at Chicago with funding from the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign. I worked as a visiting researcher in the summer of 2007 for the Graphics Research Group of Adobe Systems in Seattle, Washington.
5. From 1997-2000, I designed, simulated and patented graphics hardware for Silicon Reality and the Evans & Sutherland Computer Corp. I have also consulted for various visual effects companies (e.g., Kleiser-Walczak, Blue Sky | VIFX), defense contractors (e.g., SAIC, Pratt & Whitney) and medical imaging companies (e.g., Intrinsic Medical Imaging).
6. My research in computer graphics has been supported by Adobe, DARPA, Intel, Nokia, NVIDIA, NSF and Microsoft. I have published over 100 technical papers on computer graphics, including papers peer-reviewed, accepted and appearing in the most rigorous and prestigious publications in computer graphics, including the SIGGRAPH Annual Conference Proceedings, the ACM Transactions on Graphics, and the IEEE Transactions on Visualization and Computer Graphics.

7. The Association for Computing Machinery (ACM) is the main professional organization for computer scientists, and its special interest group on computer graphics and interactive techniques (SIGGRAPH) covers its efforts in computer graphics including virtual reality systems. From 2002 to 2008, I served as Editor-in-Chief of the ACM Transactions on Graphics, the top journal in the field of computer graphics. From 1996 to 2000, I served on the ACM SIGGRAPH Executive Committee, and in 1999 I served as Executive Producer for the production of its documentary “The Story of Computer Graphics.” From 2014 to 2022, I served as the founding computer graphics area editor of the ACM Books series, and oversaw the publication of “The VR Book: Human Centered Design for Virtual Reality” (which is cited by Petitioners as Ex. 1033 of the IPR).
8. In 1992, I co-authored “The CAVE audio visual experience automatic virtual environment” which appeared in volume 35, issue 6 of the Communications of the ACM, and cited by over 2,600 subsequent publications on virtual reality systems. This was the seminal first paper describing the “CAVE,” a room-sized virtual reality system where the walls and floor were projection screens, and head-tracking was used to

ensure the images on these screens were rendered from the specific point of view of the occupant within its boundaries.

9. In 1993, I authored a subsequent paper “On recording virtual environments” which I presented at the IEEE Visualization ’93 Symposium on Research Frontiers in Virtual Reality. In 2019, I co-authored “Real-time analytic antialiased text for 3-D environments” which won a best paper award when it was presented at the High-Performance Graphics Conference. This paper described how to render legible text for a head-mounted display in a virtual reality system.

III. SUMMARY OF OPINIONS

10. My opinions in this Declaration are based on my review of the ‘836 Patent, its prosecution history before the U.S. Patent and Trademark Office, the IPR Petition concerning this patent, and references cited therein in support of the Petition. I draw upon my experience and knowledge of acoustical engineering, psychoacoustics, signal processing, human-machine interaction, and virtual reality systems.

IV. RELEVANT LEVEL OF ORDINARY SKILL IN THE ART

11. In my opinion, as relates to the ‘836 Patent, a person having ordinary skill in the relevant art (a POSITA) would have several years of academic and/or professional experience in a technical field relevant to digital audio

signal processing, such as electrical engineering, computer engineering, or computer science; and a familiarity with relevant literature and concepts concerning virtual “3D” audio processing, which were widely available at the time of disclosure.

12. I am at least a POSITA under this definition. My training and experience in computer science, computer graphics and virtual reality systems has included such digital audio signal processing. The field of computer graphics has included publications on audio, including the use of “sonification” as an auditory supplement for data visualization. The simulation of audio wave propagation is also analogous to the radiative light wave simulations used for photorealistic rendering. The use of transfer functions, such as a head-related transfer function (HRTF) for perceiving binaural audio, is also common to computer graphics, which includes the precomputed radiance transfer function (BRTF) for perceiving binocular light. A true and accurate copy of my CV is attached to the Patent Owner Preliminary Response as Exhibit 2009.

V. THE ‘836 PATENT

13. As the ‘836 Patent explains, “[t]he realism and user experience significantly decrease when the location of the origination of the sound does not align with a visible source of the sound.” Ex. 1001, 3:15-17. The

problem “becomes exacerbated,” the ‘836 Patent correctly explains, “when the source of the sound is within a near-field distance of the listener.” *Id.*, 18-20. Through various embodiments, the ‘836 Patent discloses a method (or a system) that addresses this problem.

14. It is known that mere reliance on auditory cues for judging distances leads to distance compression (i.e., underestimation of distance between self and the sound source). This underestimation would dramatically impact user experience. To provide a more realistic user experience, it is desirable to match or align the perceived location of where the sound with a visible source of the sound, such as an image or object.
15. However, when the sound source is located within a distance of one meter from the user (commonly referred to as a near-field distance, *see, e.g.*, ‘Ex. 1001, 2:15-18), the user experience decreases dramatically in quality, partly because aligning auditory cue with visual cue for the sound source in the near-field is extremely difficult—if not impossible. Traditionally, this problem has been addressed by convolving the sound with near-field head-related transfer functions (“HRTFs”) in order for a listener to perceive the sound as if it originates from the near-field visible source of the sound. But aside from being computationally expensive, this approach may fail to accurately localize to a listener due to the lower reliability of

externalization from near-field HRTFs. To my knowledge, convolving the sound with far-field HRTFs has not been proposed or tested prior to the ‘836 Patent.

16. I understand that Petitioner challenges claims 8-17 of the ‘836 Patent, including two independent claims 8 and 15. Both independent claims have similar limitations, thus I focus of claim 8—which is reproduced below:

8. An electronic device, comprising:

a display that displays an image that occurs within a near-field distance from a head of a user; and

one or more processors that process sound for the image to generate binaural sound that has a sound localization point (SLP) with a coordinate location that occurs at a far-field distance located behind the image and along a line-of-sight from the user to the image while the display displays the image to the user at the near-field distance, wherein the one or more processors include a digital signal processor (DSP) that processes the sound for the image with far-field head-related transfer functions (HRTFs) while the electronic device is located at the near-field distance from the head of the user.

17. The key to understand this claim is “an image that occurs within a near-field distance from a head of a user.” On several occasions, Petitioner falsely assumes that this image is necessarily co-located with the claimed electronic device (or the display). In my opinion, this is wrong, Petitioner’s

- assumption is in direct collision with the express language of claim 8 as well as the specification of the '836 Patent, for reasons discussed below.
18. For Petitioner's assumption to be true, the limitation should have stated "a display that displays an image within a near-field distance." But that is not what the claim recites. Instead, the claim limitation recites "a display that displays an image that occurs within a near-field distance." In this context, a POSITA would understand that this limitation points to location of a perceived virtual image not the location of the display.
19. Further, the claim requires that the image occurs "within a near-field" whereas the electronic device (or its display) is located "at the near-field." This difference in describing the locations of image and electronic device supports a conclusion that the two are not necessarily co-located.
20. Moreover, the specification of the '836 Patent repeatedly speaks in terms of "a location where the sound source is visually perceived." For another example, consider FIGs. 6A and 6B of the '836 Patent. In describing these figures, the specification provides that

The listener, however, perceives the images at different distances with respect to the listener. Images 620A and 620B appear to be located at the surface of the display 630. Image 620C, however, appears to be located farther away

than a wall or virtual wall 622 at which the display 630 is located. Ex. 1001, 22:8-15.

21. All these examples communicate one and only one message: the image that occurs within the near-field and the claimed electronic device (or the claimed display) are not necessarily co-located.
22. Further, since the claimed display allows a user to perceive and estimate the distance between the displayed image and the user's self, a POSITA would appreciate that the claimed display necessarily provides the user with important cues to visual depth perception. This is also supported by various examples from the specification involving "stereoscopic videos and images" and "3D images." A POSITA, reading the claims of the '836 Patent in view of its disclosures, would understand that a conventional TV displaying an image with no stereo or motion parallax depth cues does not teach this limitation.

VI. GROUND 1 OF THE PETITION: FINNEGAN AND BEGAULT

23. Petitioner uses Daniel J. Finnegan et al., *Compensating for Distance Compression in Audiovisual Virtual Environments Using Incongruence*, CHI '16: Proc. Of the 2016 CHI Conf. on Hum. Factors in Computing Sys. (2016) ("Finnegan") and Durand R. Begault, *3-D Sound for Virtual Reality and Multimedia*, NASA/TM—2000-209606 (Aug. 2000) ("Begault") in

support of its obviousness challenge. Petitioner has mischaracterized teachings of Finnegan.

A. Finnegan

24. Finnegan reports a phenomenon in distance perception in virtual environments, where the perceived depths of virtual objects are compressed when compared to their actual depths. Based on their analysis and understanding of the phenomenon, Finnegan proposes a hypothesis (H1) that “[r]endering the audio at an incongruent position further from the observer than the visual stimulus ([independent variable]), will result in more accurate distance perception ([dependent variable]) compared to conditions where both stimuli are at the same position (congruent conditions).” To prove this hypothesis, Finnegan performs an experiment where subjects were presented with a first target object, which was then removed and replaced with a second target object, and asked which was closer. This experiment was performed where the target objects’ visual and auditory elements were congruent (at the same depth), and also performed where the target objects’ visual and auditory elements were incongruent, with the auditory element farther away than the visual element. Finnegan concluded (at the bottom of the left column of its 9th page) that the hypothesis (H1) “is supported by the results of our analysis in the far range

[comparing objects from 4.5 to 9.5 meters to a reference object at 7 meters]
but not in the near range [comparing objects from 0.5 to 5.5 meters away
to a reference object at 3 meters away].”

25. Finnegan is later cited by Huang, where audio reverberation is considered as a way to compensate for this distance compression. Huang summarizes Finnegan’s results in several places. “Along these lines, Finnegan et al. [Finnegan et al. 2016] found that the incongruence of auditory and visual perceptual information can compensate distance compression *in longer distances* using anechoic sound.” Ex. 2002, 2, emphasis added. “Using audio to compensate distance compression in VR has been discussed by Finnegan et al. demonstrating that applying an incongruence scheme that misaligns the location of audio and visual stimuli can be used to alter how users perceive distance. They were able to show that shifting the audio stimulus further than the visual object results in lower distance underestimation effects *in far field (4.5 - 9.5 meters)* VR experiences [Finnegan et al. 2016].” *Id.*, 3, emphasis added.

26. Claim 8 requires the display of an image within the near field (closer than one meter) accompanied by binaural sound at the far field (one meter or farther). Finnegan admits that its results in the near range (0.5 – 5.5 meters) were “non-significant,” and a POSITA, for example Huang, would have

understood Finnegan to teach away from using incongruence in the Finnegan “near-range” of 0.5 – 5.5 meters, and especially so at the ’836 Patent’s “near-field” of 1 meter.

27. Further, I note that Finnegan used the SoundScape Renderer to process the binaural audio for its experiments. “Our plugin integrates the SoundScape Renderer (SSR)[], a GPL licensed software implementation for binaural audio [], with Unity for spatial audio rendering over headphones.” Ex. 1005, 6. In an article published by SSR developers (*The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods* (Convention Paper 7330), Audio Eng’g Soc’y 124th Convention (2008)), it is discussed that HRTFs used by SSR for rendering binaural sounds are selected based on “the position of” the sound source. Ex. 1029, 3-4. A POSITA would understand that near-field HRTFs are selected for sound sources located in the near-field, which is the conventional way of rendering binaural sounds for near-field sources. Since Claim 8 requires a far-field HRTF to generate the binaural sound associated with a near-field image, this requirement would not have been supported by the SoundScape Renderer. Moreover, this Claim 8 requirement would not have been implemented by Finnegan’s plugin that “integrates” the SSR.

VII. GROUND 2 OF THE PETITION: NORRIS, BEGAULT, AND ALGAZI

28. Petitioner uses U.S. Patent No. 9,226,090 (“Norris”), Begault, and V.R. Algazi et al., *The CIPIC HRTF Database*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2001) (“Algazi”) in support of its obviousness challenges. Petitioner has mischaracterized teachings of these references.

A. Norris

29. Norris discloses a method to provide a sound localization point to a listener, enabling the listener “to determine a distance and a direction of the sound.” Ex. 1007, 3:37-38. Norris does not disclose “an image that occurs within a near-field” claimed by the ‘836 Patent. To determine the distance and direction of the sound, a user of Norris’s wearable electronic device (“WED”) can only rely on the audio cues provided. Norris’s disclosures of 2-D images and videos (e.g., “virtual image of the speaker” appears on WED (*Id.*, 32:42)) does not provide the user with important cues for perception of the location of virtual image. More important, Norris makes clear that the sound localization point is placed on the image itself—not behind it as required by the ‘836 claims:

This sound localization point 2050 can appear to the listener 2020 as a virtual image of the speaker 2010. For

example, the wearable electronic device 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. This image represents the sound localization point 2050. *Id.*, 32:40-45.

VIII. GROUND 3 OF THE PETITION: SHIVAPPA, AND ZHOU

30. Petitioner uses U.S. Patent No. 10,979,843 (“Shivappa”) and Zhiying Zhou et al., *An Experimental Study on the Role of Software Synthesized 3D Sound in Augmented Reality Environments*, 16 *Interacting with Computs.* 989 (2004) (“Zhou”) in support of its obviousness challenges. Petitioner has mischaracterized teachings of these references.

A. Shivappa

31. Shivappa discloses methods for compensating latency in a 3D audio device. As Shivappa explains, latency is introduced because process of processing “the input spatialized audio signal may be a complex and resource-intensive process for the processor 106.” Ex. 1010, 9:8-10. Thus, in cases where the speed of a moving listener exceeds the processing time required, this latency greatly affects the listener’s ability to determine the correct distance and direction of the sound source. Shivappa discloses a position predictor to compensate for the latency described above. To describe its position predicting algorithm, Shivappa discloses no more than

a well-known VR system. For example, Shivappa uses a conventional display: “The VR system (or the AR system) may be configured to display media content, such as movies, and to provide interactive content, such as video games, presentations, virtual meetings, etc.” *Id.*, 5:42-45. This disclosure does not teach “an image that occurs within a near-field distance.” Shivappa merely discusses benefits of spatialization in general terms but does not suggest how spatialization should be accomplished in order to create the optimum user experience.

B. Zhou

32. A POSITA reading Zhou would only conclude using 3D sound ““improves the accuracy of depth judgment.” Ex. 1011, 1. They used an augmented reality system with a single camera and explored “perceptual feeling of sound depth exaggeration.” *Id.*, 4. The experiment based correct judgments on virtual objects (“which telephone is nearer”) and used a very unnatural “exaggeration” of the sound cues from a virtual telephone. A POSITA could not read Zhou and gain insight into the ‘836 Patent’s distinction between near-field and far-field. Further, Zhou expressly provides that “the 3D sound appears to emanate from the marker,” not behind it as required by claims of the ‘836 patent. *Id.*, 7.

33. Zhou used a single camera system, which was suitable only for monocular vision, thereby eliminating importance cues to visual depth perception. Therefore, experiments in monocular vision are not applicable to stereo-vision head mounted displays capable of providing visual cues for depth perception. I understand Petitioner purports that since Zhou's HMD presented images to the listener on its lenses, that the image was somehow at the near field. The '836 Patent makes clear that images are not necessarily perceived at their display depth, and this was demonstrated with a wall display. A POSITA would have understood that the monocular display of a video image of a sound source, such as Zhou's example of a video call, would not be perceived to be at the display depth, which would be millimeters from the listener's eye. A POSITA would have understood Zhou to teach that the sound source would have been visually perceived by the observer to be a natural speaking distance away, and that Zhou taught that the auditory coordinates of the sound source should be co-located with the perceived location of the visual source.

34. Zhou relied on real-world depth cues, and presented monoscopic virtual images overlayed on the listener's view of the real-world environment. Turner nicely summarizes a POSTA's understanding that Zhou did not teach an appropriate display of an image that occurs within a near-field

distance from the head of a user. “[Zhou’s] results are encouraging in that they show 3D sound can be a depth cue, but the use of headphones and the monoscopic display mean the conclusions of this study cannot be transferred directly to a stereoscopic 3D viewing situation such as TV or cinema.” Ex. 2012, 4. Since Turner represents a peer-reviewed research contribution to build off of Zhou toward even just a stereoscopic display, the ordinary skill of POSITA could not have been expected to combine Zhou with Shivappa to construct anything near to the perceptual discoveries and advancements represented by the inventions claimed by the ’836 Patent.

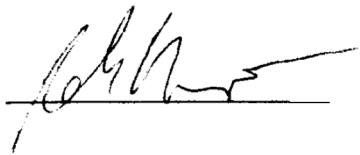
35. Furthermore, the depth exaggeration of the sound presented to the listener did not involve any depth exaggeration in the binaural processing through a HRTF. Zhou’s depth exaggeration was simply a change in the volume of sounds, resulting in the same binaural HRTF presented to a listener, but at a different volume level to better emphasize the differences between two objects at different depths. The simple change in volume based on depth falls well short of what was claimed by the ’836 Patent.

IX. CONCLUSION

36. I declare that all statements made herein of my knowledge are true, that all statements made on information and belief are believed to be true, and

that these statements were made with knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code. I reserve the right to modify my opinions as additional information becomes available to me.

Dated: September 13, 2023

By: _____