

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

META PLATFORMS, INC.,
Petitioner,

v.

EIGHT KHZ, LLC,
Patent Owner.

Case IPR2023-01019
Patent 10,368,179 B1

Case IPR2023-01020
Patent 10,448,184 B1

**DECLARATION OF DR. DURAND R. BEGAULT REGARDING
U.S. PATENT NOS. 10,368,179 AND 10,448,184**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. QUALIFICATIONS	1
III. BASIS OF OPINIONS	3
IV. RELEVANT LEVEL OF ORDINARY SKILL IN THE ART	3
V. The ‘179 Patent	4
VI. The ‘184 Patent	5
VII. GROUND 1 OF THE ‘179 PETITION AND THE ‘184 PETITION: CLEMENT AND OSMAN AND SAUK	13
A. Clement	13
B. Sauk	17
1. Combination of Clement and Sauk	19
VIII. CONCLUSION	23

I, Durand R. Begault, declare as follows:

I. INTRODUCTION

1. I am currently employed as Director, Audio Forensic Center, a wholly-owned subsidiary of Charles M. Salter Associates, an acoustical consulting firm located in San Francisco, California. I am also currently employed as a research scientist in the Human Systems Integration Division of NASA Ames Research Center, Moffett Field, California.
2. I have been retained as an independent expert consultant in the *inter partes* review (“IPR”) proceeding involving the 10,368,179 Patent and the 10,448,184 Patent by Eight kHz LLC (“8kHz”). I will refer to these patents using common shorthand. I have been asked to consider the arguments raised by the Petition regarding the validity of the ‘179 Patent and the ‘184 Patent in this proceeding.

II. QUALIFICATIONS

3. I am the director for Audio Forensic Center, where I perform investigations with regards to acoustics; video and audio evidence recordings; audio patent infringement analysis; acoustic measurement; speech intelligibility and warning signal audibility; human factors and psychoacoustics; musicology and copyright infringement; voice comparison; and recorded gunshot analysis. I also serve as a research scientist for NASA Ames

Research Center specializing in psychoacoustics, human factors, multimodal systems, and communications engineering research, applied to aerospace safety, virtual reality, and speech communications. I obtained my Ph.D. in 1987 from the University of California, San Diego with a focus in Computer Audio, and was made a fellow of the Audio Engineering Society in 2002 “for contributions to understanding of spatial hearing”.

4. For over 30 years I have conducted analyses and testified in cases for attorneys, law enforcement agencies, and corporate internal investigations. I have qualified as an expert in numerous State and U.S. Federal Courts; the United States International Trade Commission (USITC); and courts in Canada and international arbitration panels. I have over 100 publications in the area of acoustics, psychoacoustics and forensic audio, and have testified over 85 times in court or deposition, including in the area of digital audio recording authentication and metadata (e.g., the International Centre for Settlement of Investment Disputes, The World Bank, Washington, D.C.; California Superior Courts; United States District Court). I have served as an adjunct faculty member of the Sound Recording Department at McGill University in Montreal, Canada, and have received a Fellowship Award from the Audio Engineering Society. I have also served on numerous journal review boards, technical committees and Ph.D.

dissertation committees in the area of acoustics and audio engineering.

Attached hereto and incorporated by reference as Exhibit “2007” is a true and correct copy of my CV.

III. BASIS OF OPINIONS

5. My opinions in this Declaration are based on my review of the ‘179 Patent and the ‘184 Patent, their prosecution history before the U.S. Patent and Trademark Office, the IPR Petitions concerning these patents, and references cited therein in support of the Petitions. I draw upon my experience and knowledge of acoustical engineering, psychoacoustics, signal processing, human-machine interaction, and virtual reality systems, including my book “3-D Sound for Virtual Reality and Multimedia” (Begault, 2000).

IV. RELEVANT LEVEL OF ORDINARY SKILL IN THE ART

6. In my opinion, as relates to the ‘179 Patent and the ‘184 Patent, a person having ordinary skill in the relevant art (a PHOSITA) 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.

V. The ‘179 Patent

7. Through various embodiments, the ‘179 Patent discloses providing binaural sound to a user using a head mounted display (HMD), such as in a virtual reality environment, and altering the binaural sound in response to the occurrence of an event, including when moving between different virtual and real-world environments.

8. The ‘179 Patent contains three independent claims:

1. A method comprising:

storing head-related transfer functions (HRTFs) in memory of a head mounted display (HMD) worn on a head of a user;

displaying, with a display of the HMD worn on the head of the user, a three-dimensional (3D) virtual environment that includes a virtual image;

processing, by a digital signal processor (DSP) in the HMD, sounds with the HRTFs to generate binaural sound that is provided to the user and that externally localizes in empty space to the virtual image in the 3D virtual environment;

sensing, with one or more sensors in the HMD, when the user leaves the 3D virtual environment that includes the virtual image; and

warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter audio cues to prevent the user from experiencing external localization of the sounds and by lowering a volume of the sounds being provided to the user.

7. A method executed by one or more electronic devices, the method comprising:

storing head-related transfer functions (HRTFs) in memory;

displaying, by a display of a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

processing, by a digital signal processor (DSP), sounds with the HRTFs to generate binaural sound in empty space that is provided to the user and that externally localizes to the empty space and to the virtual image in the environment; and

determining, with the WED, when the user is leaving the environment; and

alerting the user that the user is leaving the environment by altering processing of the binaural sound to degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound.

14. A method executed by one or more electronic devices, the method comprising:

storing head-related transfer functions (HRTFs) in memory;

displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes in empty space from the virtual image in the environment;

determining, with the WED, when the user is leaving the environment; and

alerting the user that the user is leaving the environment by reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user.

VI. The '184 Patent

9. Through various embodiments, the '184 Patent discloses providing binaural sound to a user using a head mounted display (HMD), such as in

a virtual reality environment, and altering the binaural sound in response to the occurrence of an event, including when moving between different virtual and real-world environments.

10. The '184 Patent and the '179 Patent have the same specification and figures. Thus, like the '179 Patent, the '184 Patent discloses a transition from the virtual reality environment to the real-world environment.

11. The '184 Patent contains three independent claims:

1. A method comprising:

storing head-related transfer functions (HRTFs);

displaying, with a display of a head mounted display (HMD) worn on a head of a user, a three-dimensional (3D) virtual environment that includes a virtual image;

processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that plays to the user and that externally localizes in empty space at the virtual image in the 3D virtual environment; and

warning the user that the user is leaving the 3D virtual environment by filtering, by the DSP, the binaural sound to alter a spectral cue and lower a volume of the sound playing to the user.

7. A method executed by one or more electronic devices, the method comprising:

storing head-related transfer functions (HRTFs);

displaying, by a display of a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that plays to the user and that externally localizes at the virtual image in the environment; and

alerting the user that the user is leaving the environment by altering processing of the sound by the DSP to degrade a spatial cue and to lower a volume of the sound playing to the user.

14. A method executed by one or more electronic devices, the method comprising:

storing head-related transfer functions (HRTFs);

displaying, with a wearable electronic device (WED) worn on a head of a user, a virtual image in an environment;

processing, by a digital signal processor (DSP), sound with the HRTFs to generate binaural sound that externally localizes to the user at the virtual image in the environment; and

alerting the user that the user is leaving the virtual image in the environment by altering processing of the binaural sound by the DSP and reducing a volume of the binaural sound that localizes to the user at the virtual image.

12. The independent claims of the '184 Patent and the '179 Patent described in paragraphs 8 and 11 above specify methods for altering the virtual location of an auditory signal in response to an external condition or action precipitated by the listener. Specifically, when a user leaves an "environment", a "3D virtual environment", or a "virtual image in the environment", the claims require a specific change in the mode of auditory presentation to the listener. By altering the method of sound spatialization,

a consequent change in the perception of the location of virtual sounds occurs.

13. In my opinion, a PHOSITA would recognize that the stated “environment”, a “3D virtual environment”, or a “virtual image in the environment” in the subject claims all to refer to the same type of immersive virtual world context that the user interacts within. In this context, virtual sound sources are perceived to be externalized to the user at different sound localization points relative to the listener. *See* Figure 1 below, left (the violin and bell sound sources are heard outside of the head).
14. A PHOSITA would also recognize that the concept of “leaving” the stated “environment”, a “3D virtual environment”, or a “virtual image in the environment” all to refer to an act of a significant transition of user perception as a result of a user’s movement or a user’s change in location relative to the respective environment.
15. A PHOSITA would also recognize the use of binaural processing as the best method to produce sound imagery that is externalized in 3D dimensional space about a user, and that stereo or monaural sound cannot accomplish equivalent sound imagery. “Binaural processing” is equivalent to “3-D audio processing”, in that directional cues caused from filtering by the pinnae (outer ears) are included along with directional cues caused by

interaural time and level differences (Begault, 2000, page x)¹ (Begault, 2000, page 41).² These cues are implemented by means of filters capturing the binaural “Head Related Transfer Function” (HRTF).

16. Sound “externalization” specifically refers to auditory distance perception such that sounds are perceived at locations outside of the user’s head.

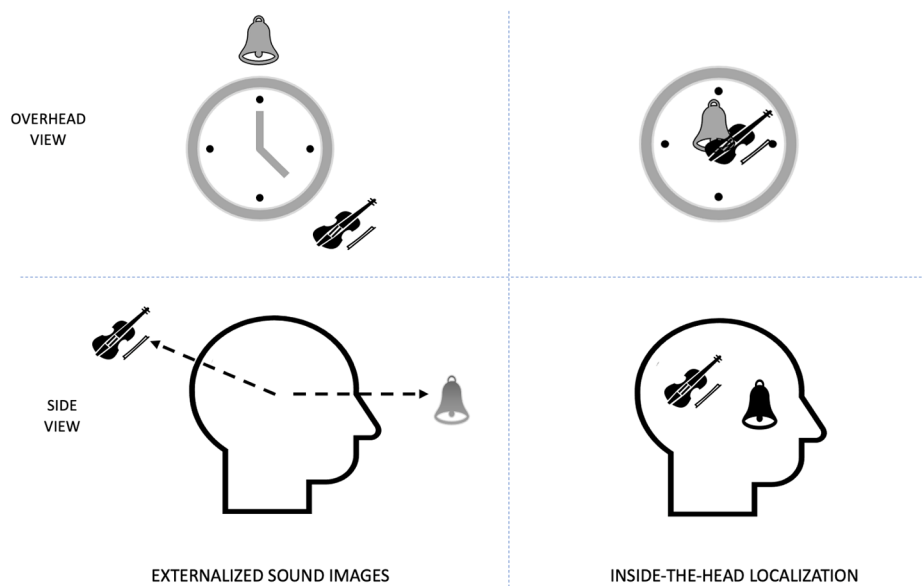
¹ From Begault (2000), page x: “While this book mainly uses the phrase “3-D sound,” equivalent designations include **virtual acoustics**, **binaural audio**, and **spatialized sound**. It’s certain that as new products appear on the market, further variants on these phrases will appear. Fundamentally, all of these refer to techniques where the outer ears (the **pinnae**) are either directly implemented or modeled as digital filters. By filtering a digitized sound source with these filters, one can *potentially* place sounds anywhere in the virtual space about a headphone listener.”

² From Begault (2000), page 41: “The use of HRTF spectral shaping is typically featured as the key component of a 3D sound system, from either direct measurement or modeling. This is based on the theory that the most accurate means to produce a spatial sound cue is to transform the spectrum of a sound source at the eardrum as closely as possible to the way it would be transformed under normal spatial hearing conditions”

When sounds are not externalized, the result is referred to as inside-the-head localization (IHL) or lateralization of sound within the head. See Figure 1, right.

17. Figure 1 shows: Left, top: overhead view of externalized sounds (a bell and a violin). The “clock” (representing directions relative to the head) shows a bell at 12 o’clock (in front of the listener) and a violin at at 5 o’clock (to the rear of a listener) at distances outside the head. Left, bottom: same, viewed from the side. Right, top: overhead view of the same sounds but heard inside the head. Right, bottom: same, viewed from the side.

FIGURE 1.



18. When the spectral cues provided by binaural sound are degraded, or when stereo (2 channel sound without directional pinnae cues) or monaural sound (1 channel sound) are heard over headphones, inside-the-head localization occurs. Begault (2000, p. 90: emphases added) addresses the well-known problem of sound source externalization with degraded binaural sound cues:

Durlach and Colburn (1978) have mentioned that the externalization of a sound source is difficult to predict with precision, but "clearly, however, it increases as the stimulation approximates more closely a stimulation that is natural." The goal of eliminating IHL effects arose in the 1970s with the desire to make improved binaural (dummy head) recordings. Many who heard these recordings were disturbed by the fact that the sound remained inside the head, as with lateralization. Ensuring that the sound was filtered by **accurate replicas of the human pinnae** and head was found to be an important consideration.

19. The '179 patent outlines three methods of altering binaural audio to signal to a listener that they are leaving a virtual environment: "**filtering**, by the DSP, **the binaural sound** to alter audio cues to prevent the user from experiencing external localization of the sounds and by lowering a volume

of the sounds being provided to the user.” (claim 1); “**altering processing of the binaural sound** to degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound.” (claim 7); and “**reducing a volume of the binaural sound** that localizes from the virtual image and that is being provided to the user. (claim 14). In each case, a consequent change in the perception of the location of virtual sounds occurs.

20. The ‘184 patent outlines two methods of altering binaural audio to signal to a listener that they are leaving a virtual environment: “**alter a spectral cue** and lower a volume of a sound playing to a user” (claim 1); “**degrade a spatial cue** and to lower a volume of a sound playing to a user (claim 7); and one method of altering binaural audio to signal to a listener that they are leaving a virtual image in an environment: “**altering processing of the binaural sound** by the DSP and reducing a volume of the binaural sound” (Claim 14). In each case, a consequent change in the perception of the location of virtual sounds occurs.

21. First, a PHOSITA reading these claims would recognize that “lowering the volume of a sound” is a salient cue to a change in the character of a sound. It becomes an even more salient cue when in combination with manipulation of sound source externalization. Furthermore, in an

augmented reality context, it would allow increased intelligibility of real-world sound.

22. Second, a PHOSITA would recognize that the most effective methods to “prevent the user from external localization of sounds” – in other words, to cause sound to be heard inside of the head—would be to “alter spectral cues” and to “degrade a spatial cue” provided by an HRTF filter; or to use stereophonic or monaural sound.

VII. GROUND 1 OF THE ‘179 PETITION AND THE ‘184 PETITION: CLEMENT AND OSMAN AND SAUK

23. I understand that Petitioner relies on U.S. Patent No. 9,690,374 (“Clement”), U.S. Patent No. 9,037,468 (“Osman”), and U.S. Patent No. 7,876,903 (“Sauk”) to allege that these references teach or suggest every limitation of the independent claims in the ‘179 and ‘184 patents. I disagree, and for the reasons discussed below, I am of the opinion that Petitioner mischaracterized teachings of Clement and Sauk.

A. Clement

24. Clement refers to sound only in terms of a “fade” process, and as an “alert” for boundary detection. There are no references to any spatial audio aspect, either in terms of user perception or in terms of signal processing.

25. Clement discloses fading out “virtual audio content” heard by a user when transitioning out of a virtual environment and into the real world. Clement also discloses that the fading out of the audio heard by a user may be “accompanied by a fade in of ambient sound.” Clement at 9:22-23.
26. The technique Clement discloses is analogous to an audio mixing console, which “mixes” different sound inputs at different levels to an output. An audio mixing console analogous to Clement would have two inputs: (1) virtual audio; and (2) real world audio. Its output would be a proportional mix of these two inputs to its output to a listener. In the virtual world, input 1 (virtual audio) is set to 100% level at the output and input 2 (real world audio) is 0% level at the outset. As one transitions from the virtual world to the real world, audio input 1 (virtual audio) is proportionally reduced from 100% to 0% while audio input 2 (real world audio) is increased from 0% to 100%. Clement refers to this transition as a “fade in” and “fade out” between the real and virtual worlds.
27. Clement does not disclose any specific form of sound processing specific to creating virtual sound imagery; in particular, binaural sound is not mentioned. Clement also does not disclose transitioning or converting binaural sound to monophonic or stereophonic sound.

28. Clement does not refer to any perceptual aspect of virtual audio sources that would occur through the described fading technique, nor to any perceptual aspect of spatial hearing. Clement only refers ambiguously to “audio content associated with the virtual world” and “ambient sound”. A PHOSITA would have no guidance on how “audio content” or “ambient sound” would be implemented or processed, other than this fading process.
29. Finally, Clement does not disclose a method of degrading spatial audio cues present in binaural sound for the purpose of creating inside-the-head localization. Clement does not indicate a method of switching between binaural and degraded binaural, stereo, or monaural audio cues when transitioning between the real world and a virtual world.
30. Clement discloses a specific application of sound as a “warning indicator” to a listener to signal initiation or a transition process; to avoid an obstacle or hazard by playback of, for example, a “buzz”. Clement 7:28-8:35. This warning indicator is not shown to be spatialized to indicate directionality of the obstacle or hazard; it simply draws attention to the fact that a transition or a hazard is present.
31. In summary, a PHOSITA would only understand Clement to show two well-known concepts: producing a warning audio signal related to a user’s condition; and fading between two audio scenes, as with a mixing console.

Case IPR2023-01019
Patent 10,368,179 B1

Case IPR2023-01020
Patent 10,448,184 B1

B. Sauk

32. Sauk generally describes means of prescribing spatial attributes to specific types of sounds, referred to as “enunciated messages”, that are comprised of either pre-recorded speech messages or synthetic tones (“earcons”). Each type of sound (“Type 1 information”) is associated with metadata (“Type 2 information”) that instructs an audio processor how to render the sound in virtual space for a particular perceived location, and in some cases an intended group of listeners (receivers). The rendering can be “3-D binaural”, “stereophonic” or “monophonic”, and can include reverberation cues (“binaural audio environments”).
33. The defined connection between a specific sound (the “enunciated message”) and its location and processing method (its metadata) means that the spatial position of the specific sound is not dynamically changing between different states in response to user position in an environment (e.g., the virtual world). Sauk, 2:24-59. The metadata assigns different sounds to specific positions, analogous to the way in which elementary school students are assigned to a specific desk in a classroom in front a teacher.

34. Sauk discloses the possibility for a user to select the means of spatial reproduction from binaural to monophonic or stereophonic, based on relevancy or importance of the audio to the user. Sauk, 2:61-64; 11:26-41; 12:14-51. “[A] user could change the less relevant audio [from binaural (3-D)] to a monophonic (1-D) or stereophonic (2-D) dimension.” Sauk 11:25-40. “Relevancy” or “importance” is a function of the user’s decision regarding multiple streams of audio information they are monitoring, as opposed to a transition of the user between the real-world and the virtual world.
35. Sauk does not disclose a transition from an immersive virtual world environment to the real world. Petitioner’s characterization suggests Sauk discloses dynamic “changing [of] an audio format.” Sauk does not indicate a method of switching between binaural and degraded binaural, stereo, or monaural audio cues to indicate a transition between the real world and a virtual world. Instead, as mentioned above, Sauk generates new reproductions of enunciation data based on the user switching between binaural, stereophonic, and monophonic audio reproduction, only under conditions of deterministic user interaction. Sauk, 2:61-64; 11:26-41; 12:14-51.

Sauk and alerts/warnings

36. Sauk also discloses using a binaurally-processed signal to alert users to the direction and location of hazards in the real world, for example, a chemical toxin. There is no disclosure of changing the method of spatial rendering from binaural to stereo or degraded binaural as part of this warning system. Such hazard alerts are only heard binaurally. Sauk, 12: 39-51. This makes sense since stereophonic or monophonic sound would not enable sufficient spatial awareness to notify the user of the location of a real-world hazard.

1. Combination of Clement and Sauk

The types of alerts disclosed in Clement and Sauk teach away from claimed invention

37. Petitioner argues that “POSA implementing Clement would have looked to Sauk for additional teachings regarding binaural sound, including signaling the importance of things or events to provide flexibility and additional options for different conditions in, e.g., virtual gameplay, thus ensuring a suitable warning regarding the transition (e.g., leaving a virtual room while remaining in the virtual environment vs. completely leaving the environment)”. ‘184 Pet. 14. Sauk cites military tactical communication examples applicable to soldiers in a real-world situation

involving GPS tracking and radio frequency (RF) signals. Sauk's disclosure is therefore pertinent to an augmenting real-world visual and audio stimuli with informational auditory messages and signals that are overlaid on the real world. However, neither Sauk nor Clement contemplates or uses the word "game" or "game play" in a virtual environment. A PHOSITA would not understand how Sauk's tactical system would provide "flexibility and additional options for different conditions" in immersive virtual environments, as asserted in the Petition. In particular, Sauk teaches the use of predetermined spatial information contained in metadata that fixes spatial reproduction in advance. The only spatial "flexibility" mentioned by Sauk is how a user can deterministically change the means of spatial reproduction from binaural to monophonic or stereophonic, based on the user's decision as to their relevancy or importance of the transmitted information.

38. Petitioner argues that "a POSA would have known that other options [besides a buzzer] are available and maybe preferable. For example, Sauk teaches changing from binaural to stereo or monophonic sound to signal "relevancy or importance," such as warning of a "hazard." Sauk 11:26-40, 12:33-51 (warning of a hazard); Roginska ¶¶ 111-113." '184 Pet. 14. Petitioner here conflates the new concept of "other options [besides a

buzzer]”, which refers to a type of alert sound and not sound spatialization, with Sauk’s spatialization prioritization scheme, in which a user deterministically changes the means of spatial reproduction from binaural to monophonic or stereophonic, based on relevancy or importance of the audio as decided by the user. “Relevancy” or “importance” is a function of the user’s decision regarding multiple streams of audio information they are monitoring, contrasting the ‘179 and ‘184 patents’ description of automated spatial sound transitions imposed by the system and not by the user, based on the user’s location within the real and/or virtual worlds.

39. If a PHOSITA were to look to the “other options” to alert/warn, “such as those taught by Sauk,” as asserted in the Petition, the PHOSITA would find that Sauk teaches adding additional binaural alerts when required to avoid hazards to an existing display of informational messages. Sauk, 12: 39-51. This is very different from the ‘179 and ‘184 teachings of transitioning away from binaural audio (e.g., by degrading spatial audio cues) in response to location. Thus, rather than supposed altering of binaural audio down to stereo or mono to create inside-the-head localization, the alerts taught by Sauk actually teach the opposite, to add in new externalized binaural alerts to indicate the location of a hazard. Sauk’s alerts are additive to other informational messages in the display,

not alterations of existing sound messages and/or degradative processing resulting in a change in perceived audio position.

40. I do not agree with Petitioner's argument that a PHOSITA would combine Clement and Sauk. Clement's "fade" processing and alert signal applies to virtual worlds and a transition to an unspecific "ambient" condition, without specification of any sort of spatial audio processing or change in spatial audio processing. Sauk's message system applies to a spatial message and hazard alert augmentation to the real world. Although "alerting" is a common concept in Clement and Sauk, the applicable context of each system, the means of interaction of the user with the system, and the means of transition from one audio spatialization state to another are not comparable nor obvious to combine.

41. In addition, it would not be obvious to a PHOSITA to equate Clement's "fade" processing functionality with the additional alerts envisioned by Sauk. Petitioner argues that "Clement teaches degrading the sound by fading it out (reducing its volume) as discussed above with respect to Element 14[e]." '179 Pet. 25-26; '184 Pet. 29. A mere "fade out" of virtual audio, as described in Clement, does not inherently involve or suggest "degrading the sound" or "degrading a spatial cue," as a mere lowering of

volume of virtual audio does not “alter[] processing of the sound by the DSP to degrade a spatial cue”.

42. When a user transitions from a virtual environment with binaural sound to a real environment with stereo or monaural sound, this constitutes an “event” that is signaled by a significant change in the perceived location of sound, as described in the ‘179 and ‘184 patents. For example, in the specification of the ‘184 patent, “Events can be triggered by a change in a user's ability to associate visual cues or images with the associated audio cues, such as a visually rendered character vanishing from an augmented or virtual environment, the presence or absence of a physical object, a failure of a visual display system, a degradation in the visibility of a user's physical or virtual environment, or the impairment or failure of a user's physical eyesight”. These types of events are described in several examples in the specification of both patents, with the event causing a degradation of spatial imagery effected by binaural sound.

VIII. CONCLUSION

43. 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 14, 2023

By: 
Durand R. Begault Ph.D.