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 IN SUPPORT OF PATENT OWNER'S RESPONSE

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 1

TABLE OF CONTENTS

I. INTRODUCTION 3
II. QUALIFICATIONS 3
III. BASIS OF OPINIONS 5
IV. RELEVANT LEVEL OF ORDINARY SKILL IN THE ART 5
V. The '179 Patent 6
VI. The '184 Patent 10
VII. Claim Construction 15
VIII. The '179 and '184 Patents Disclose Transitions Between Different Types of Environments 20
IX. The '179 and '184 Patents Disclose Alerting a User (1) When the User is Leaving a Physical Environment and Also (2) When a User is Leaving a 3D Virtual Environment 24
A. Events that precipitate an alert as described in the specifications. 25
- The '140 Application 29
B. Methods of creating alerts in the claims. 37
X. The Prior Art Does Not Invalidate the Claims of the '179 and '184 Patents 45
A. Sauk 7,876,903 45
B. Osman 9,037,468 47
C. '179 Patent Dependent Claims 48
D. Alerting/Warning by Reducing Volume of Binaural Sound 52
E. '184 Patent Dependent Claims 54
XI. CONCLUSION 58

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 2

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 these proceedings.

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 95 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. A true and correct copy of my CV is attached as Exhibit 2013 and incorporated by reference.

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 histories before the U.S. Patent and Trademark Office, the IPR Petitions concerning these patents, and exhibits and references cited therein in support of the Petitions. I have also reviewed the '140 Application. 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; a familiarity with basic principles of audio engineering; and a familiarity with relevant literature and means of implementing virtual ("3D") binaural audio processing, including those that were widely available at the time of disclosure.

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 5

V. The '179 Patent

7. Through various embodiments, the '179 Patent discloses providing binaural (3-D) sound to a user using a head mounted display (HMD), such as typically found in a virtual reality or augmented reality environment, and altering the provided cues to spatial hearing in response to the occurrence of an event, including when moving between different virtual and/or real-world environments, including when leaving a 3D virtual environment that is represented visually and/or aurally.

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;

<u>sensing, with one or more sensors in the HMD, when the user leaves the 3D virtual environment</u> 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 <u>alter audio cues to prevent the user from *experiencing external localization*</u> of the sounds and by <u>*lowering a volume of the sounds being provided to the user*.</u>

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.

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 7

9. The '179 Patent also contains dependent claims that include additional limitations.

10. Claims 3-6 depend from claim 1:

> 3. The method of claim 1, wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment.
>
> 4. The method of claim 1 further comprising:
>
> tracking, with head tracking in the HMD, head movements of the user; and
>
> switching the sounds from the binaural sound to stereo sound when a face of the user is directed away from the virtual image.
>
> 5. The method of claim 1 further comprising:
>
> switching the sounds from the binaural sound to stereo sound when the user activates a physical switch located in the HMD.
>
> 6. The method of claim 1 further comprising:
>
> switching the sounds from the binaural sound to stereo sound when the virtual image vanishes in the 3D virtual environment.

11. Claims 8-13 depend from claim 7:

> 8. The method of claim 7 further comprising:
>
> alerting the user that the user is leaving the environment by reducing a volume of the binaural sound being provided to the user.
>
> 9. The method of claim 7 further comprising:
>
> displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and
>
> switching a voice of the character from the binaural sound to stereo

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 8

sound when the character vanishes from the environment during the game.

10. The method of claim 7 further comprising:

alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.

11. The method of claim 7 further comprising:

switching the binaural sound to stereo sound upon activation of a switch on the WED.

12. The method of claim 7 further comprising:

wherein the WED includes electronic glasses, the environment is a physical room where the user is located, the virtual image is an augmented reality (AR) character that speaks to the user while the user plays a game, the WED includes head tracking, and a digital signal processor (DSP) in the WED reduces a volume of the binaural sound to alert the user when a face of the user is not directed to the virtual image in the environment.

13. The method of claim 7 further comprising:

sensing, with the WED, when a physical object interferes with a location of the virtual image in the environment; and

switching the binaural sound to stereo sound in response to sensing the physical object interferes with the location of the virtual image in the environment.

12. And claims 15-20 depend from claim 14:

15. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by degrading the binaural sound being provided to the user as the user is leaving the environment.

16. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by altering, by

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 9

the DSP, processing of the binaural sound so the user perceives less external localization.

17. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.

18. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by eliminating audio cues in the binaural sound being provided to the user.

19. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to stereo sound.

20. The method of claim 14 further comprising:

detecting, with a microphone in the WED, a voice of a person in the environment of the user; and

switching the binaural sound to stereo sound in response to detecting the voice.

VI. The '184 Patent

13. Through various embodiments, the '184 Patent discloses providing binaural (3-D) sound to a user using a head mounted display (HMD), such as typically found in a virtual reality environment, and means of altering the provided cues to spatial hearing in response to the occurrence of an event, including when moving between different virtual and/or real-world environments, including when leaving a 3D virtual environment that is represented visually and/or aurally.

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 10

14. The '184 Patent and the '179 Patent have the same specification and figures. Thus, like the '179 Patent, the '184 Patent discloses a transition between different virtual and/or real-world environments including when leaving a 3D virtual environment. that is represented visually and/or aurally.

15. 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
>
> <u>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*</u>.
>
> 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

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 11

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*.

16. The '184 Patent also contains dependent claims that include additional limitations.

17. Claims 3-6 depend from claim 1:

3. The method of claim 1, wherein the warning occurs when the user is leaving a virtual area where the virtual image is located while remaining in the 3D virtual environment.

4. The method of claim 1 further comprising:

tracking, with head tracking in the HMD, head movements of the user; and

switching the sound from the binaural sound to stereo sound when a face of the user is directed away from the virtual image such that the virtual image is no longer visible to the user.

5. The method of claim 1 further comprising:

switching the sound from the binaural sound to stereo sound when the user activates a button located on the HMD.

6. The method of claim 1 further comprising:

switching the sound from the binaural sound to stereo sound when the virtual image vanishes from the 3D virtual environment.

18. Claims 8-13 depend from claim 7:

8. The method of claim 7, wherein the spatial cue includes an interaural time difference (ITD).

9. The method of claim 7 further comprising:

displaying, by the WED, the virtual image while the user plays a game that includes a character as the virtual image; and

switching a voice of the character from the binaural sound to stereo sound when the WED no longer displays the virtual image during the game.

10. The method of claim 7 further comprising:

alerting the user that the user is leaving the environment by switching the binaural sound to stereo sound.

11. The method of claim 7 further comprising:

switching the binaural sound to stereo sound upon activation of a button on the WED.

12. The method of claim 7 further comprising:

wherein the WED includes electronic glasses, the environment is a physical room where the user is located, the virtual image is an augmented reality (AR) character that speaks to the user while the

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 13

user plays a game, the WED includes head tracking, and a digital signal processor (DSP) in the WED reduces a volume of the binaural sound to alert the user when a face of the user is not directed to the virtual image in the environment.

13. The method of claim 7 further comprising:

sensing, with the WED, when a physical object interferes with a location of the virtual image in the environment; and

switching the binaural sound to stereo sound in response to sensing the physical object interferes with the location of the virtual image in the environment.

19. And claims 15-20 depend from claim 14:

15. The method of claim 14, wherein the DSP alters processing of the binaural sound to alert the user by changing a spatial cue that includes at least one of an interaural time difference (ITD), an interaural level difference (ILD), and the HRTFs.

16. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by stopping, by the DSP, the processing of the sound with the HRTFs.

17. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by removing interaural time differences (ITDs) of the binaural sound being provided to the user.

18. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by eliminating one or more spatial cues in the binaural sound being provided to the user.

19. The method of claim 14 further comprising:

alerting the user that the user is leaving the environment by switching the binaural sound playing to the user to one of stereo sound and mono sound.

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 14

20. The method of claim 14 further comprising:

detecting, with a microphone in the WED, a voice of a person in the environment of the user; and

switching the binaural sound to stereo sound in response to detecting the voice.

VII. Claim Construction

20. I have applied the plain and ordinary meanings of the claim terms in the '179 and '184 Patents in my analysis, as a PHOSITA would have understood them in the context of the patents at the time of their priority dates.

21. The term "environment" as used in the '179 and '184 Patents is unambiguous and consistent with its modern definition ("the circumstances, objects, or conditions by which one is surrounded"- www.merriam-webster.com/) as well as with the etymology of its root word, "*environ*", derived from the ancient Greek word *perivallon*, ("that which surrounds us" www.proz.com/), and late 14th century French ("to surround, encircle, encompass," (www.etymonline.com). In a system as described in the patents, one can be surrounded by circumstances, objects or conditions that are either real, virtual, or both simultaneously.

22. In my opinion, a PHOSITA would have understood the term "environment" as used in claims 7 and 14 of the '179 and '184 Patents to

mean the "physical surroundings in which a user operates." This meaning is consistent with the description in the specification of the '179 and '184 Patents. The "environment" refers to the position of the user in physical space while they interact with virtual and real objects or components of the virtual rendering system.

23. In my opinion, a PHOSITA would have understood the term "3D virtual environment" to mean more specifically "virtual surroundings in which a user operates that incorporate 3D virtual images and/or 3D sound." This meaning is consistent with the description in the specification of the '179 and '184 Patents. This term appears, for example, in claim 1 of both the '179 Patent and the '184 Patent.

24. A PHOSITA would also understand that the phrase "3D virtual environment" refers to either a fully immersive or augmented reality (partially-immersive) virtual world that the user interacts within, such as a virtual environment that incorporates 3D virtual images and/or 3D sound. In this context, virtual sound sources are perceived to be externalized to the user at different sound localization points relative to the listener.

25. A PHOSITA would also recognize that the concept of "leaving the environment" to refer to the user moving from one physical environment condition to another physical environment condition. For example, a user

may move from outdoors in a telephone conversation to the inside of a car. When moving between physical environment conditions, the 179/184 Patent provides an auditory cue.

26. A PHOSITA would also recognize that the concept of leaving a "3D virtual environment", refers to an act of a transition of user perception of a change in location between virtual environments, or a change in the location or existence of SLPs. This recognition of a basic concept of virtual environments' capabilities is consistent with '179 and '184 specifications' disclosure of a "change in a virtual environment".

27. That the "3D virtual environment" includes 3D audio is consistent with the specification of the '179 and '184 Patents and the general understanding of a PHOSITA. The '179 and '184 Patents at column 2, lines 38-46 state "[b]y way of introduction, sound localization (i.e., the act of relating attributes of the sound being heard by the listener to the location of an auditory event) provides the listener with a three-dimensional (3D) soundscape or 3D sound environment where sounds can be localized to points around the listener. Binaural sound and some forms of stereo sound provide a listener with the ability to localize sound, though binaural sound generally provides a listener with a superior ability to localize sounds in the 3D environment."

28. Regarding the audio component of a 3D virtual environment, Begault, 2000 (p.1) wrote how the technology of 3D sound enables "malleable" transformations ("transmutations": changes in form, appearance or nature) between virtual acoustic environments, an effect that would be known to a PHOSITA that would aid in their understanding of the Patents' concepts regarding "leaving an environment":

> "Only imagination and the tools at hand pose limits to the spatial manipulations possible for a user to make with existing technology. Malleable virtual acoustic simulations, such as those described in the following examples, could be produced for a listener:
>
> 1) **Replication** of an existing, spatial auditory condition. Starting with a recording of a concert violinist made in a relatively reverberant-free environment, one could, with absolute control, create a virtual acoustic experience that gives rise to a percept of the violinist walking back and forth on the stage of a well-known auditorium like Carnegie Hall, from the perspective of any seat in the auditorium.
> 2) **Creation** of completely unknown spatial auditory experiences. For instance, shrinking the violinist to a sound source the size of a fly and then hearing it play while it zooms rapidly about your head in a marble bathroom the size of a football stadium.
> 3) **Transmutation** between one spatial experience and another. You're listening as if seated on the stage of Carnegie Hall. The violinist, playing a furious solo cadenza, moves rapidly toward you across a quiet, snow- covered field from 400 feet away. The music gets closer and closer, wider, more present, and more reverberant until, after a minute, right at the climax of the solo, the violinist arrives at the front of the stage, joined by the orchestra.
> 4) **Representation** of a virtual auditory world. Through headphones, a "slice of life" of the acoustic experience of a violinist could be experienced, as heard through the violinist's ears. The user of the 3-D sound system would use an interface to be transported at will from the limousine to the green room, then backstage behind the curtain before the concert, and then out a side stage exit back to

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 18

> the street. The 3-D sound system designer has in this case acoustically mapped the acoustic environment in advance, as well as the means for allowing the listener to control his or her interaction in the environment.

29. A PHOSITA would recognize that, prior to the advent of virtual reality, cueing a transition between visual and auditory environments through the use of sound is a familiar concept as old as the history of scene changes in dramatic stage arts, such as opera or theater. Transitions between virtual auditory environments has for many years been part of first monaural, then binaural audio, and then virtual 3D sound, in radio dramas, nature recordings and the like, as described in (Begault, 2000, p. 11-12):

> Compared to the synthesis of virtual visual experiences, the **transformation of experience** via acoustic storage and simulation of spatial audio is not only much easier to accomplish but is also usually more convincing. For example, consider the popular genre of environmental-nature recordings that are binaurally recorded expressly for the purpose of "transforming one's moods to a gentle northeastern farmland at dusk," etc. There is a rich history of radio drama production that attests to the power of combining monaurally delivered sound for narrative, dramatic purposes (e.g., Orson Wells' "War of the Worlds"). Intuitively it seems that audio is simply *easier* to use for transportation into a virtual world, especially if we keep our eyes closed and free from distractions. … Even the two-inch speaker in your television is capable of producing a non-interactive, virtual auditory environment; without looking at the screen, we can easily imagine the relationship of a live audience to a talk show host and her guests, an intimate dialogue or confrontation between two people, or the completely synthetic experience of a "talking head" such as a newscaster. The virtual auditory world of theater and opera was taken into radio and television dramas; hence, it is perfectly normal today in the virtual auditory world of TV to

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 19

hear a rock band wailing away in the same virtual auditory space of police cars screeching around corners in hot pursuit, or string orchestras in the same room as the unfaithful lover giving a confessional.

VIII. The '179 and '184 Patents Disclose Transitions Between Different Types of Environments

30. The types of environmental transitions addressed in the specification sections of the '179 and '184 Patents include both a transition in the physical environment of the user (either caused by movement of the user, or a change in the circumstances or conditions of things within the physical environment of the user), as well as transitions in the 3D virtual environment heard and seen by the user. Such transitions cause a change in spatial audio rendering from binaural to stereo or monaural sound to effect a cue to the user that the transition has occurred.

31. The following two tables give examples from the patent specifications for what a PHOSITA would recognize as physical environment or virtual environment transitions. Table I gives examples of transitions in physical environments, and auditory cues used to inform the user of the transition. Table II gives examples of transitions in 3D virtual environments, and auditory cues used to inform the user of the transition. The patents include other examples of transitions and auditory cues.

TABLE I. Examples from the '179/'184 Patents of transitions in physical environments

	CHANGE IN AUDIO RENDERING: **FROM…….**	**TO…….**
	MONO SOUND	BINAURAL SOUND
example	Consider an example in which a user listens to an audio drama that was recorded in binaural sound but is played in mono sound through car speakers while the user drives the car. Upon arriving at a destination, the person wants to continue listening to the audio drama, steps out of the car, and places headphones on his head. The system continues streaming the audio drama to the person uninterrupted by sending the stream to the headphones rather than to the car. At this time, the system knows the audio drama is a binaural signal and switches the audio drama to binaural sound as it transmits to and plays through the headphones of the person. (Ex. 1028, 109-110.) This example shows a transition of a user from one physical environment to another that cues a change in spatial audio rendering.	
	BINAURAL SOUND #1	BINAURAL SOUND #2
example	Consider an example in which an electronic device is set to provide a SLP of a voice of an intelligent personal assistant three feet in front of a listener. The electronic device includes a sensor (such as a camera or other type of sensor) to determine a distance from the electronic device and/or the listener to an object. When the object is within a	

	predetermined distance (such as being within three feet of the listener), then the electronic device takes an action with regard to the SLP, such as moving the SLP, removing the SLP, switching or changing to stereo or mono sound, etc. This action prevents the voice from appearing to originate or to emanate from the object when such is not the desire or intention of the listener. (Ex. 1028, 117-18.) This example shows a transition, or change, in the circumstances or conditions of things within the physical environment of the user.	
Transition type	BINAURAL SOUND	MONO SOUND
example	Consider the example above in which Alice wears earphones and localizes a voice of Bob in front of her during a phone call. The car has a sensor that determines Alice is on a binaural call and instructs an HPED of Alice to switch the call to mono. As another example, when Alice enters the car, a system in the car pairs with the HPED and automatically switches the call from binaural to mono. As another example, a GPS device or object recognition device (such as a camera with object recognition software) determine that Alice is entering or in the car and provides a signal to the HPED or other source of the call to switch the call from binaural to mono. (Ex. 1028, 120.) This example shows a transition of a user from one physical environment to another that cues a change in spatial audio rendering.	

TABLE II. Examples from the '179/'184 Patents of transitions in 3D virtual environments

	CHANGE IN AUDIO RENDERING: **FROM…….**	**TO…….**
	BINAURAL SOUND #1	BINAURAL SOUND #2
example	Consider an example in which a user wears or has an activity tracker or motion sensor (such as a device that monitors, tracks, and/or measures fitness-related metrics like distance walked, calories burned, rate of walking or running, etc.). The activity tracker or motion sensor detects when a person commences to walk quickly or run. When this event occurs, the computer system or electronic device changes or switches binaural sound. (Ex. 1028, 161). This example shows a transition in 3D virtual environments ("switches binaural sound") that cues a change in spatial audio rendering. In scenario 110, a user 120 wears electronic earphones 122 while simultaneously localizing a voice of an intelligent personal assistant (IPA) to a first sound localization point 124 and a voice of a friend to a second sound localization point126. As shown in box 128A, the user 120 localizes the voice of the IPA to his left and localizes the voice of his friend in front of himself and above his laptop computer that is situated on his desk. An image of the friend appears on a display of the laptop while the SLP 126 appears above the laptop. As shown in the transition from box 128A to box 128B, the user 120 stops externally localizing a voice of his friend, and the sound localization point 126 disappears. When the call ends, the system changes the sound localization point 124 of the IPA and automatically moves it to be in front of the user 120. A voice of the friend switches to mono or stereo	

	or gets localized internally to the user 120. (Ex. 1028, 102). This example shows a transition in 3D virtual environments (the location of a friend and the IPA) that cues a change in spatial audio rendering.	
Transition type	STEREO SOUND	BINAURAL SOUND
example	Consider an example in which Glen is on a phone call with Alice in which a voice of Alice appears to Glen as stereo sound through speakers in an HPED that he holds. Alice informs Glen that she wants to talk to him "face to face" and requests that they meet in a visually rendered chat room. Glen goes into a quiet area in his house and dons a heads-up display (HUD) that couples with his HPED and meets Alice in the chat room. This action of donning the heads-up-display automatically switches the voice of Alice from stereo to binaural. (Ex. 1028, 120-21.) This example shows a transition in 3D virtual environments (moving to a visually rendered chat room) that cues a change in spatial audio rendering.	

IX. The '179 and '184 Patents Disclose Alerting a User (1) When the User is Leaving a Physical Environment and Also (2) When a User is Leaving a 3D Virtual Environment

32. Claim 1 of the '179 and '184 Patents specify methods for alerting a listener (user) when the user is leaving a 3D virtual environment. Claims 7 and 14 of the '179 and the '184 Patents specify methods for alerting a listener (user) when the user is leaving a physical environment.

33. Specifically, when a user leaves a physical "environment" or a "3D virtual environment", the claims require a specific change in the mode of auditory

presentation to the listener that provides a cue or an "alert" as to the transition. The alert changes the spatial audio signal processing so as to affect the listener's perception of the location of one or more virtual sounds. This is explained in the first paragraph of the specification's "detailed description":

> When the event occurs, binaural sound changes, such as switching to stereo sound, switching mono sound, switching to altered binaural sound, removing or changing a sound localization point (SLP), moving a SLP (such as moving the SLP from being externally localized to being internally localized), or taking another action in accordance with an example embodiment. ('179/'184 Patent, column 2, lines 31-37.)

A. Events that precipitate an alert as described in the specifications.

34. Figure 1 of both '179 and '184 Patents (the figures are identical between the two) describe the examples of various environmental transitions that would *precipitate* such an alert. Figures 2-3 provide a description of the method for implementing the change in spatial audio rendering that *causes* the alert. In flow charts shown in Figures 4-16, the *change in environmental condition* (a change in the circumstances, objects, or conditions by which one is surrounded) that precipitates an_alert is described in the flow charts as "take action". These environmental changes are summarized below, and would be recognized as common design

challenges in the design of a virtual acoustic world view by a PHOSITA

35. "Object within area of SLP" (Figure 4); the event is an object is described in the patent as "a physical thing that occupies a space" ('179/'184 Patents at 7:67)in the real environment that is associated with an SLP. See '179/'184 Patents at 13:64-7. In Figure 1 this corresponds to "In scenario 110, the user 120 externally localizes sounds to emanate from objects on the desk, such as designating cup 127 as a SLP and designating stapler 129 as another SLP." '179/'184 Patents at 6:29-31.

36. "Event occurs" (Figure 5); this refers to a user who moves away from a "sweet spot" in the real environment, which is known to a PHOSITA as a location where either loudspeaker imagery of spatial sound is best maintained, or where communication with a sensor is optimized (see '179/'184 Patents at 15:38-16:12).

37. "Moving to a restricted area" (Figure 6). Restricted areas can include the user moving to a location in the real environment where fully immersive sound may be unsafe from a human factors perspective, such as to within a vehicle. '179/'184 Patents at 16:13-17:10.

38. "Two SLPs overlap" (Figure 7). When concurrent spatialization of two SLPs at the same location occurs in a virtual environment, the action is

taken so as to avoid a spatial conflict or "collision" of overlapping virtual sources. '179/'184 Patents at 17:11-18:19.

39. "Increase above threshold" (Figure 8); the action occurs when there is a significant degradation of 3D audio rendering due to data stream overloading ("packet loss") in the real environment. '179/'184 Patents at 18:20-51

40. "change request to the sound/SLP" (Figure 9); the action occurs when the SLP of a "intelligent virtual agent" is requested by the user to change its location in the virtual environment. '179/'184 Patents at 18:52-19:17.

41. "System change needed" (Figure 10); the action occurs under various scenarios of a user interacting in the real environment with communications having limited hardware capability (e.g., a land line telephone call) or with network-based home appliances that provide audio cues from headphones or loudspeakers (e.g., a home fan, or a clock). '179/'184 Patents at 19:18-21:12.

42. "Locations congruent" (Figure 11); the action occurs when a spatial compensation (offset) between perceived virtual location and actual location in the real world environment is required. '179/'184 Patents at 21:13-22:51.

43. "Permission granted" (Figure 12); the action occurs when a user grants

permission to switch to a personalized (optimized) HRTF set for superior virtual environment rendering, versus sub-optimal means of spatial audio rendering (e.g., stereo). '179/'184 Patents at 22:51-24:2.

44. "Threshold met" (Figure 13); the action occurs when there is a significant degradation of 3D audio rendering due to the hardware resources available in the real environment. '179/'184 Patents at 24:3-25:20.

45. "Provide an alert" (Figure 14); the action occurs when the user fails to notice an alert in the virtual environment that is triggered by an event in the real environment. '179/'184 Patents at 25:21-26:67.

46. "Threshold time has passed" (Figure 15); the action occurs when the user does not interact with sound in the virtual environment for a calculated time period. '179/'184 Patents at 27:1-61.

47. "Event detected" (Figure 16); this is a global description of the various changes that might occur in the real or virtual environment that would precipitate a change in spatial audio rendering: "Events are also triggered according to one or more audio cues detected in a user's or a remote user's physical or virtual environment such as cues indicating a location, position, or orientation, or a change in them, cues indicating a reference frame of a user or a remote user, a lateral or vertical motion, a change in distance, a change in a physical or a virtual environment such as its shape, acoustic

properties, noise level, or placement of objects or structures in the environment". '179/'184 Patents at 28:18-26.

- **The '140 Application**

48. I understand that the '140 Application (Exhibit 1028, pages 92-191) is the parent of both the '179 and '184 Patents and, as a result, the specification and figures of the '140 Application is the same as the specification and figures of both the '179 and '184 Patents. I understand that the claims are different in the '140 Application, the '179 Patent, and the '184 Patent. In the section below, I cite to the '140 Application but it should be understood that identical disclosure is in the '179 Patent and the '184 Patent.

49. I have been asked to provide an opinion on whether the '140 Application provides sufficient written description support for the subject matter claimed in the '179 Patent and the '184 Patent.

50. I understand that a claim has sufficient written description support when the disclosure as filed would have reasonably conveyed to a PHOSITA that the inventor was in possession of the claimed subject matter as of the filing date of the application.

51. The '140 Application discloses a *change in environmental condition* that precipitates an alert when a user is leaving an environment, including when

a user is leaving a 3D virtual environment (claim 1 of '179 and '184 Patents) or a physical environment (claims 7 and 14 of '179 and '184 Patents).

52. The '140 Application at page 119, line 18-page 120, line 27 describes FIG. 6 and in particular a method to determine a location of a person and to take an action when the person moves into a restricted area. An example is provided where "Alice wears earphones and localizes a voice of Bob in front of her during a phone call. The car has a sensor that determines Alice is on a binaural call and instructs an HPED of Alice to switch the call to mono. As another example, when Alice enters the car, a system in the car pairs with the HPED and automatically switches the call from binaural to mono." '140 Application, page 29, lines 20-24. In this example, Alice starts in one physical environment (outside her car) and one virtual environment (a 3D virtual environment where she experiences 3D sound) and moves into a different physical environment (inside her car) and a different virtual environment (a 1D virtual environment where she experiences mono sound).

53. The '140 Application at page 168, lines 15-16 states "Examples herein can take place in physical spaces, in computer rendered spaces (VR), in partially computer rendered spaces (AR), and in combinations thereof."

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 30

54. The '140 Application at page 113, lines 13-16 states "A person can receive binaural sound corresponding to one virtual or real or augmented space, and a switch occurs such that the person receives binaural sound from a second virtual or real or augmented space."

55. The disclosure regarding Figure 6, taken with the other disclosure in the '140 Application regarding the examples can take place in physical spaces, in computer rendered spaces (VR), in partially computer rendered spaces (AR), and in combinations demonstrates that the examples in Figure 6 regarding the restricted area take place in a combination of physical spaces and virtual spaces; that is, the "restricted area" corresponds to both a physical space and a virtual space where 3D audio is not permitted, as explained above.

56. In addition, the example of the user moving from a non-restricted area to a restricted area is an example of the user leaving a 3D virtual environment, as detected by a handheld portable electronic device (HPED) or wearable electronic device (WED) or a global positioning sensor (GPS) device. Ex. 1028, 120. The '140 Application expressly discloses that the WED includes a "GPS tracking system," location tracker, and sensors (Ex. 1028, 70) and that "one or more of [the components of the electronic devices] can be distributed or included in various electronic devices, such as some

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 31

components being included in an HPED" (Ex. 1028, 168). Thus, a PHOSITA would recognize that the '140 Application reasonably conveys that the "GPS tracking system" or location and movement tracker in the WED, for example, could be used determine the user's location as described in the example of the user leaving the non-restricted area.

57. In addition to the disclosure regarding the non-restricted area described in regard to Figure 6, the '140 Application provides more events that can precipitate altering binaural sound. At page 140, line 15-page 141, line 28, the '140 Application describes FIG. 16 and in particular a method to provide binaural sound to a person and to take action when an event occurs. Events "can be triggered by changes in a user's or a remote user's network conditions, systems resources, … a change in the ability to detect … position **in a physical or a virtual environment**, or a change in the ability to detect an environment…" '140 Application, page 140, line 31-page 141, line 4. The Patents also disclose "cues indicating … a *changing in a physical or virtual environment*," and "a degradation in the *visibility of a user's physical or virtual environment*" Ex. 1028, 140:31-141:10, 141:15-20.

58. Again, in view of the specific disclosure in the '140 Application that all of the disclosed examples "can take place in physical spaces, in computer

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 32

rendered spaces (VR), in partially computer rendered spaces (AR), and in combinations thereof" (Ex. 1028 168:15-16), a PHOSITA would recognize that a change in the ability to detect an environment reasonably corresponds to a change in the ability to detect a physical environment, a virtual environment, or both. In addition, a PHOSITA would recognize that the degradation in the visibility of a user's 3D virtual environment reasonably corresponds to when the user leaves the virtual or 3D virtual environment. As such, the '140 Application (and the '179 and '184 Patents) describe events as including changes to physical and 3D virtual environments and performing actions, such as adjusting binaural sound, in response to a detected event.

59. The adjustments in binaural sound, such as a change in sound format from binaural to mono or stereo or the reduction in volume as disclosed in the '140 Application, would implicitly alert or warn the user that the user is leaving the physical environment (claims 7 and 14 of the '179 and '184 patents) or 3D virtual environment (claim 1 of the '179 and '184 patents), as further discussed in the section below and the section entitled "Methods of creating alerts in the claims".

60. The '140 Application also discloses using **alerts and warnings**. The '140 Application states at page 115, line 31-page 116, line 6, "[i]n Figures 4-16,

example actions include, but are not limited to, one or more of switch the sound from binaural sound to stereo sound, switch the sound from binaural sound to mono sound … maintain binaural sound but alter the binaural sound, stop binaural sound, … lower a volume of the sound, provide a sound or audio alert, provide a visual alert…" At page 136, lines 3-10, the '140 Application states "For example, the alert is an audible alert and/or a visual alert to a person. By way of example, such alerts include, but are not limited to, one or more of a displaying a visual warning, providing an audible sound … displaying a visual or perceivable indication or warning … and providing another indication that notifies a user." At page 138 of the '140 Application, lines 24-27, it states "Consider an example in which a child runs behind a car as it is backing up. A camera at the back of the car provides a video and audio alert in stereo to the driver. The driver, however, does not see or hear the alert so the car switches the alert to a binaural sound that emanates a sound alert from a location of the child."

61. The passages from the '140 Application I have quoted in this declaration are intended to be examples only. My opinion is based on my review of the entire '140 Application.

62. A PHOSITA would recognize that the claimed subject matter of the '179 and '184 patents is reflected in the '140 Application as filed. In particular,

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 34

the '140 Application discloses the concept of alerting when the user is leaving a physical environment (see, for example, claims 7 and 14 of the '179 and '184 Patents) and the concept of warning when the user is leaving a 3D virtual environment (see, for example, claim 1 of the '179 and '184 Patents). The '140 Application also discloses the concept of using the HMD/WED to sense or determine when the user leaves the physical environment or the 3D virtual environment (see, for example, claims 1, 7, and 14 of the '179 Patent).

63. ***Dr. Roginska's declaration.*** I have reviewed Dr. Roginska's declaration submitted in support of the IPR petitions (Exhibit 1003), including her statement that the '140 Application "does not provide sufficient support to show the requisite possession of the claimed subject matter." Roginska Declaration, Exhibit 1003, at paragraph 223. Dr. Roginska also makes a blanket assertion that "the '140 application does not describe anywhere (including the claims) the triggering event or a user leaving the environment." She continues, "In other words, the '140 application does not give an example of using the user's location to determine that (and to warn or alert) the user that the user is leaving the virtual environment, as recited in the claims." Exhibit 1003, at paragraph 225. She also asserts that "there is no description of detecting when or altering the sound to warn

or alert the user that the user is leaving the environment." Exhibit 1003, at paragraph 228.

64. Based on my analysis of the specifications and claims of the applications and the patents presented in this declaration, it is clear that Dr. Roginska is incorrect for two fundamental reasons. First, contrary to Dr. Roginska's statement the '140 Application "does not support the requisite possession of the claimed subject matter," a PHOSITA would in fact have an understanding of how to implement the claimed spatial audio methods described in '179 and '184 Patents based on knowledge of spatial audio rendering, and from the examples given in the specifications common to the '140 Application and the patents at issue. A PHOSITA would have the understanding of the meaning of the word "environment" both in its everyday meaning, as such environments are presented in virtual and augmented reality, and of the physical real environment in which a user would operate a virtual reality system. Second, as shown above, the '140 Application contains multiple disclosures and examples of alerting the user when the user is leaving either a physical environment or a 3D virtual environment by altering the spatial audio rendering, including switching from binaural to monaural or stereo. In addition, based on the disclosures in the '140 Application, in my opinion a PHOSITA reading the '140

Application at the time of the '140 Application filing date would conclude that the '140 Application reasonably conveyed that the inventor was in possession of the claimed subject matter as of that date.

B. Methods of creating alerts in the claims.

65. The methods of altering sound spatialization to alert a user of a change or transition in environments are indicated in italics in paragraphs 9 and 16 that reproduce the independent claims. A change from head-tracked binaural sound to stereo, monaural, and/or a change in the filtering or volume of a sound, or a change that causes a sound to no longer be externalized, are forms of *degrading* a head-tracked binaural sound. These methods are summarized as follows, along with how a PHOSITA would understand how to implement them in signal processing.

a. Claims 1 and 7 of the '179 Patent: alter/degrade audio cues to prevent the user from experiencing external localization of the sounds

66. Claim 1 of the '179 Patent recites "…alter audio cues to prevent the user from experiencing external localization of the sounds …" ('179, claim 1). Claim 7 of the '179 Patent recites "… degrade audio cues in order to prevent the user from experiencing external localization of the binaural sound.: ('179, claim 7).

67. A PHOSITA would recognize that the "prevent[ing] the user from experiencing external localization of the binaural sound" means the user

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 37

hears "inside the head localization." That is, the opposite condition of sound source externalization is a sound heard inside of the head ("inside the head localization"). This occurs when the same signal is presented to both ears (monaural sound). This also occurs when the sound is presented in stereo, as a transition to stereo sound would allow hearing virtual images on the azimuthal plane but not in 3D, and with little or no externalization. (The use of HRTFs is required to achieve elevation and externalization of virtual images). Either type of audio transformation—from binaural to monaural sound or from binaural to stereo sound—would result in a noticeable "cue" or alert to a listener.

b. Claims 1 and 14 of the '179 Patent: lowering/reducing a volume of the sounds provided to the user

68. Claims 1 and 14 of the '179 Patent include "lowering a volume of the sounds being provided to the user" (claim 1) and "reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user" (claim 14).

69. A lowering of sound volume is equivalent to a reduction in signal gain, akin to a sound volume control. A significant change in the volume of a virtual sound will be noticeable to a listener; e.g., a PHOSITA would recognize that a 10 decibel reduction in sound level is equivalent to the

perception of a sound that is half as loud as a reference sound whose level is not reduced. A reduction in volume is also equivalent to a change in a spatial auditory distance cue. Generally, acoustic waveforms from sound sources in the real world propagate such that there is a 6 decibel reduction in level for every doubling of distance (a phenomenon described as the "inverse square law" and known to a PHOSITA). Ref. Begault, 2000 (p. 71, Figure 3.2).

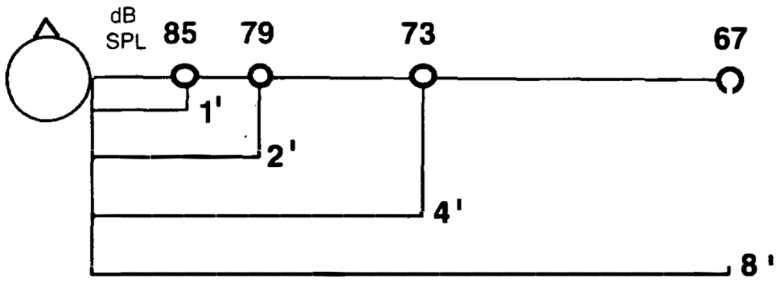


Figure 3.2. Reduction in intensity for virtual sources, using the assumptions of the inverse square law. Intensity falls by 6 dB for each doubling of distance.

70. Thus, lowering or reducing the volume of sound provided to a user would result in a noticeable "cue" or alert.

c. <u>Claim 1 of the '184 Patent: filtering, by the DSP, the binaural sound to alter a spectral cue and lower a volume of the sound playing to the user</u>

71. Claim 1 of the '184 Patent recites warning the user "…by filtering, by the DSP, the binaural sound to alter a spectral cue …" and lower a volume of

the sound playing to the user" ('184, claim 1). A PHOSITA would recognize that processes that involve altering a spectral cue and lower[ing] a volume of the sound are wholly consistent with well-known means for simulating an increase in the perceived distance of a sound.

72. As explained above with reference to '179 claims 1 and 14, which also include a recitation of lowering volume, a reduction in sound volume is equivalent to a change in a spatial auditory distance cue, based on the physical phenomenon of the inverse square law.

73. Furthermore, in the real world, spectral cues to auditory distance are such that, as distance to a sound source increases, higher frequencies of the sound are attenuated relative to lower frequencies. "The spectral content of a sound source relative to a receiver position can vary as a function of its distance. The effects include the influence of atmospheric conditions, molecular absorption of the air. and the curvature of the wavefront…" (e.g., Begault, 2000, p. 78, Figure 3.9, showing the effect of air absorption and humidity at 10 meters)…. All of these factors modify the spectral content of the sound source in an additive manner, along with the spectral cues of the HRTF (Begault, 2000, p. 76)…"

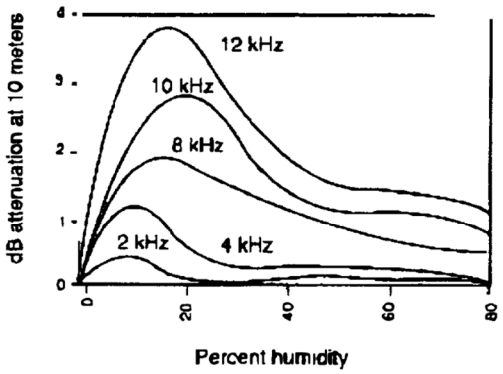


Figure 3.9. Absorption of sound in air as a function of relative humidity. Curves are shown for 2, 4, 6, 8, 10 and 12 kHz. After Harris (1966).

74. A spectral change caused by diminishing high frequencies as a cue to distance would also be known to a PHOSITA. "Butler, Levy, and Neff (1980) processed stimuli with filters of the low-pass (i.e., with high frequencies attenuated) and high-pass (with low frequencies attenuated) type at several "cut-off' frequencies. They found that the low-pass stimuli were consistently perceived to be farther away than the high-pass stimuli and suggest that "a lifetime of experience" with distant sources helps establish this cue. Perhaps, then, in the absence of other cues, a low-frequency emphasis applied to a stimulus would be interpreted as "more distant" compared to an untreated stimulus." (Begault, 2000, p. 78).

75. The Patents also describe the use of altering a spectral cue by dampening the high frequencies of a talker in a communication scenario who walks behind a wall. '179/'184 patent at 13:15-21. A PHOSITA would recognize

that high frequencies are attenuated when a sound source moves behind an object, such as a wall.

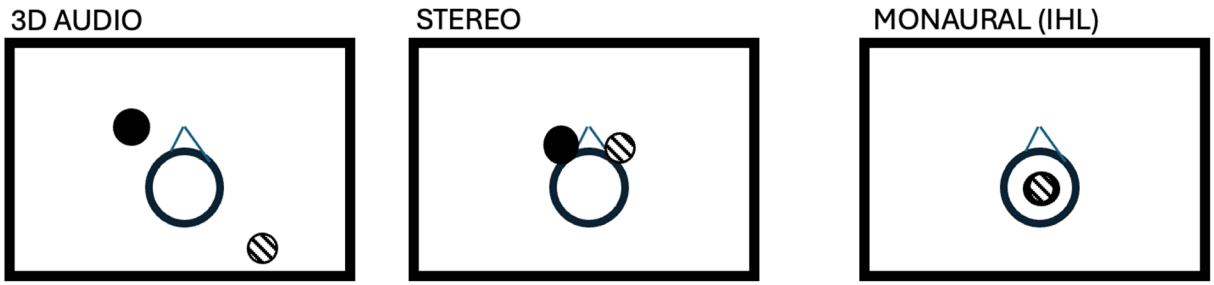
d. Claim 7 of the '184 Patent: degrade a spatial cue and to lower a volume of the sound playing to the user

76. Claim 7 of the '184 Patent recites "degrade a spatial cue and to lower a volume of the sound playing to the user." A PHOSITA would read "degrade a spatial cue and to lower a volume of the sound playing to the user" ('184, claim 7) to mean using spatial auditory means of processing other than HRTFs to process a target sound (such as stereo or monaural processing), in conjunction with lowering the gain of the sound as described for claims 1 and 14 of the '179 Patent above. A PHOSITA would recognize common and well-known methods of spatial processing that are relatively degraded compared to 3D sound techniques that use HRTF filtering.

77. For example, a transition in processing from 3D sound to stereo amplitude panning would maintain a left-right position of the positions assigned within a 3D audio display, but would have the cues to sound source elevation removed. The amplitude differences could be maintained to a "panned" position, but the spectral differences caused by the HRTF would be absent. (The word "pan" is an abbreviation of the term "panoramic

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 42

potentiometer," a processor found in professional stereo mixing consoles and familiar to most persons who have a stereo system as a "balance" knob or slider.) Furthermore, in stereo, externalization is minimized compared to HRTF processing. Another means of degrading spatial cues would be a transition to a sound heard inside of the head ("inside the head localization"), implemented by playing the same signal to both ears. See Figure 1.



78. FIGURE 1. Left: With 3D (binaural) audio, two sound sources (SLPs) can be located in front and rear of a listener and externalized. The two SLPs may also be heard at two different elevations. Middle: A degraded version of the spatial rendering of the same two SLPs would be via the use of stereo panning. The level differences are maintained, but the interaural time differences and spectral differences caused by binaural HRTF filtering are removed. The perceptual result for the user is that the externalization is degraded and front-back location differences of the two sounds are removed. Sounds are perceived on a two-dimensional axis to the left and

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 43

right of the user. Right: Another form of degradation occurs when the sound is made identical at the two ears (monaural processing). The perceptual result for the user is the sound appears inside of the head (IHL-inside the head localization) since interaural level, time and spectral differences are removed.

79. Begault, 2000 (p. 69) wrote how degraded spatial cues can still convey the impression of an environment. "Even with nonspatial cues, one can tell something about the distance of different sounds and their context; for instance, you can experience distance listening with one ear, and you can tell sometimes in a telephone conversation where a person is calling from, based on the type of background noise. But the inclusion of 3-D sound techniques and spatial reverberation cues can greatly increase the immersiveness of a simulation, as well as the overall quality and the available nuance in representing different environmental contexts".

e. Claim 14 of the '184 Patent: 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

80. Claim 14 of the '184 Patent recites "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." A PHOSITA would understand

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 44

that 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 ('184, claim 14) to mean changing some aspect of the position of the SLP (for example from binaural to stereo, or from binaural to monaural) in conjunction with a reduction in amplitude (volume), so as to effect an alert to the listener (user) regarding a transition.

X. The Prior Art Does Not Invalidate the Claims of the '179 and '184 Patents

81. I previously provided a declaration in support of Patent Owner's Preliminary Response in which I opined that the prior art relied upon by the Petitioner does not invalidate the independent claims of the '179 and '184 Patents. I incorporate the opinions expressed in my prior declaration into this declaration.

A. Sauk 7,876,903

82. Sauk (Ex. 1007) relates to techniques for "producing, combining, and customizing virtual sound environments to provide the user with understandable information regarding their surroundings." Ex. 1007, 2:45-49. Sauk discloses several applications of its system in the context of military reconnaissance missions. Id., 10:44-15:23.

83. Sauk generally describes means of prescribing spatial attributes to specific types of sounds, referred to as "enunciated messages", that are comprised of communications, pre-recorded speech messages, or synthetic tones

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 45

("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).

84. Sauk describes three modes of operation: Selective Relevance, Selective Filtering, and Alerts and Warnings. In both Selective Filtering and Alerts and Warnings modes, head-tracked binaural audio is used. Selective Filtering is used to convey the position of soldiers by spatially processing their voice communications. In Alerts and Warnings mode, spatial processing of a warning signal (earcon) is used to convey the position of a hazard. In Selective Relevance mode, user-selected voice communication channels can be presented in either 3D audio, stereo, or monaural sound.

85. In only Selective Relevance Mode, 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. "Relevancy" or "importance" is a function of the user's selective 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.

86. Sauk does not use the change in spatial audio rendering in any of the three modes to function as an alert. Also, Sauk does not use a change in the type of spatial audio rendering in his "Alerts and Warnings" mode; rather, the warning sound is always processed as head-tracked binaural audio. This makes sense since stereo or monaural sound would not enable sufficient spatial awareness to notify the user of the location of a real-world hazard. This contrasts with the '184 and '179 Patents claims, which require changing from head-tracked binaural audio to stereo, monaural, or other methods of degrading binaural sound. Finally, the change in spatial audio rendering in Selective Relevance Mode does not depend on the user's position in a physical environment or a 3D virtual environment; instead, the change in spatial audio rendering is selected by the user based on the user's preference.

87. Sauk's disclosure of an idea that allows the users to pre-select the type of spatial audio rendering for voice communications in Selective Relevance Mode is thus a completely different concept than the binaurally-processed warning sounds used in his Alerts and Warnings mode.

B. Osman 9,037,468

88. Osman relates to techniques for simulating the location of a sound source. Ex. 1006, abstract. To provide sound localization, Osman discloses HRTF processing of sound to provide cues to auditory localization *Id.*, 4:48-67.

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 47

Osman teaches that "a good HRTF model" is more important to "improve the accuracy of sound localization for objects outside the field of display" than visible objects, because "the brain uses the visual information to identify the sound origin" and "when the sound originates outside the field of view, the extra visual cue for sound localization is lost." *Id.*, 7:31-42. Osman utilizes HRTF processing to achieve 3D audio presentation, but does not discuss the use of stereo processing, monaural processing, or a reduction in sound volume as cues to environmental transition.

C. '179 Patent Dependent Claims

89. The dependent claims of the '179 Patent are reproduced above. The dependent claims in the '179 Patent can be broken in to four general categories:

1) claims that require alerting/warning the user by removing/degrading the binaural sound received by the user (claims 10, 15-19 of the '179 Patent);
2) claims that require switching the binaural sound to stereo sound based on the virtual image (claims 4, 6, 9 of the '179 Patent);
3) claim 3 of the '179 Patent, which requires warning when the user is leaving a virtual area where the virtual image is located while remaining in the 3D virtual environment; and
4) claim 8 of the '179 Patent, which requires alerting the user by

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 48

reducing the volume of the binaural sound provided to the user, similar to the lowering the volume limitations in independent claims 1 and 14.

90. Neither Sauk nor Sauk in combination with Osman and Clement disclose a method of alerting/warning the user by removing/degrading the binaural sound received by the user (claims 10, 15-19 of the '179 Patent). Specifically, in claim 17, a method of degrading the binaural sound involves removing the interaural time difference (ITD) of the binaural HRTF. Normally, a binaural HRTF includes an ITD based on the dimension of the head and the delay of a signal from one ear to another as a function of the location of the virtual source. If one manipulates the ITD by removing the ITD, the binaural sound received by a user will be degraded. A PHOSITA would recognize that manipulating ITD would likely create a significant change in a listener's perception of the spatial location of a sound.

91. Claims 10, 15, 16, 18, and 19 of the '179 Patent each represent similar degradations of the binaural sound by removing spatial auditory cues normally present in a binaural HRTF.

92. Neither Sauk nor Sauk in combination with Osman and Clement disclose a method of switching the binaural sound to stereo sound based on the

virtual image (claims 4, 6, 9 of the '179 Patent). Regarding claims 4 and 9, Petitioner asserts that, because Osman discloses head tracking, when a face is directed away from an object, its audio becomes less relevant under Sauk. Petitioner further asserts that because the audio becomes "less relevant," it would be switched to stereo. Petitioner's assertions mischaracterize teachings of both Osman and Sauk. The word "stereo" is not present anywhere in Osman's patent. Sauk never mentions assigning audio to stereo based on the direction of the face. Petitioner's asserted combination of Osman and Sauk is incorrect because neither Osman nor Sauk teach or even suggest the recited features of claim 4.

93. In fact, Osman explicitly discloses that providing spatial awareness is more important when objects are not in view by using HRTF processing for binaural sound rendering, and not stereo or monaural or any degraded version of binaural sound. Specifically:

> [A] good HRTF model will greatly improve the accuracy of sound localization for objects outside the field of display. This is because inaccuracies in virtual surround systems are ignored by the brain in the presence of visual feedback. However, when the sound originates outside the field of view, the extra

visual cue for sound localization is lost. In this case,

> ***a good HRTF enhances sound localization for objects outside the field of view***.

Osman, 7:31-42.

94. Petitioner does not rely on Clement, Osman, or Sauk to teach warning the user "when the user is leaving a virtual room while remaining in the 3D virtual environment," as required by claim 3 of the '179 Patent. Jang (Ex. 1015) describes a method whereby virtual space is divided into a focus area and non-focus areas. He cues a transition from a focus area to a non-focus area by gradually decreasing or increasing the volume of sounds. However, Petitioner does not read claim 3 in context with claim 1. That is, claim 1[e] describes a method of "warning the user … by filtering, by the DSP, the binaural sound to alter audio cues to prevent the user from experiencing external localization of the sounds" and claim 3 recites "wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment." As such, the warning method recited in claim 3 is accomplished by methods that are not considered by Jang: "filtering, by the DSP, the binaural sound to prevent the user from experiencing external localization of the sounds."

95. Claim 8 of the '179 Patent depends from claim 7 and further requires

Eight khz, LLC
Exhibit 2012
IPR2023-01019
Page 51

"alerting the user that the user is leaving the environment by reducing a volume of the binaural sound being provided to the user." This limitation is similar to the limitations in claims 1 and 14 of the '179 Patent and is addressed below.

D. Alerting/Warning by Reducing Volume of Binaural Sound

96. Neither Clement nor Clement in combination with Sauk disclose a method of alerting/warning the user by reducing a volume of the binaural sound (claims 1, 8, 14 of '179 Patent; claims 1, 7, and 14 of '184 Patent). A mere "fade out" of virtual audio, as described in Clement, does not inherently involve or suggest

1. "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 of '179)

2. "altering processing of the binaural sound to degrade audio cues to prevent the user from experiencing external localization of the binaural sound" (claim 7 of '179) and "reducing the volume" (claim 8 of '179)

3. "reducing a volume of the binaural sound that localizes from the virtual image and that is being provided to the user" (claim 14 of '179)

4. "filtering, by the DSP, the binaural sound to alter a spectral cue and lower a volume of the sound playing to the user" (claim 1 of '184)

5. "altering processing of the sound by the DSP to degrade a spatial cue and to lower the volume of the sound playing to the user" (claim 7 of '184)

6. "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" (claim 14 of '184).

97. While Clement discloses a "fade-out" of volume, it is only based on the *x*-*y* location of the user in a physical space, and does not disclose binaural audio or a fade-out of specific sound localization points because Clement does not track the position of the head in pitch, yaw, roll. Furthermore, Clement fades the volume of the virtual simulation until it is acoustically eliminated, and replaces it with real-world ambient sound, a pre-set tone, or silence. The specification of the '179/'184 Patents, however, shows that the binaural sound being provided to the user can fade between two environments, or abruptly change its level. See, e.g., '179/'184 Patent at 16:10-12 ("As another example, when a telephone rings, this event causes the home music system to lower the music volume and switch the sound to

mono"). Sauk discloses using binaural audio in his Alerts and Warnings mode to warn users of the location of a hazard, but does not teach or suggest lowering the volume of the warning (in fact, lowering the volume of an alert would be counter-productive to alerting a user to danger). Petitioner's expert, Dr. Roginska, agrees that Sauk does not disclose to reduce the volume of Sauk's warning.

98. Furthermore, Sauk's disclosure that in "Selective Relevance Mode," the "decibel level of the 1-D, 2-D or 3-D audio can be adjusted to whatever the listener desires for that dimension" is clearly limited to the user's pre-selected preferences for audio and does not teach or suggest reducing the "decibel level" as an alert or warning.

E. '184 Patent Dependent Claims

99. The dependent claims of the '184 Patent are reproduced above. The dependent claims in the '184 Patent can be broken into three general categories:

1) claims that require alerting/warning the user by removing/degrading the binaural sound received by the user (claims 8, 10, 15-19 of the '184 Patent);
2) claims that require switching the binaural sound to stereo sound based on the virtual image (claims 4, 6, 9 of the '184 Patent); and

3) claim 3 of the '184 Patent, which requires warning when the user is leaving a virtual area where the virtual image is located while remaining in the 3D virtual environment.

100. Neither Sauk nor Sauk in combination with Osman and Clement disclose a method of alerting/warning the user by removing/degrading the binaural sound received by the user (claims 8, 10, 15-19 of the '184 Patent). Specifically, in claim 8, a method of degrading the binaural sound involves altering the interaural time difference (ITD) of the binaural HRTF. Normally, a binaural HRTF includes an ITD based on the dimension of the head and the delay of a signal from one ear to another as a function of the location of the virtual source. If one manipulates the ITD by increasing or decreasing this time delay, the binaural sound received by a user will be degraded. A PHOSITA would recognize that manipulating ITD would likely create a significant change in a listener's perception of the spatial location of a sound.

101. Specifically, in claim 15, a method of degrading the binaural sound involves altering the interaural time difference (ITD), interaural level difference (ILD), or the HRTFs used for a spatial audio simulation. As explained above, if one manipulates the ITD by increasing or decreasing this time delay, the binaural sound received by a user will be degraded.

Normally, a binaural HRTF includes an ILD caused by the shadowing of frequencies from one ear to the other, whose wavelengths are smaller than the dimension of the head. A PHOSITA would recognize that manipulating ILD and/or ITD would likely create a significant change in a listener's perception of the spatial location of a sound. Furthermore, changing the HRTFs used for effecting a SLP would also create a significant change in a listener's perception of the spatial location of a sound.

102. Claims 10, 17, 18, and 19 of the '184 Patent each represent similar degradations of the binaural sound by removing spatial auditory cues normally present in a binaural HRTF.

103. Neither Sauk nor Sauk in combination with Osman and Clement disclose a method of switching the binaural sound to stereo sound based on the virtual image (claims 4, 6, 9 of the '184 Patent). Regarding claims 4 and 9, Petitioner asserts that, because Osman discloses head tracking, when a face is directed away from an object, its audio becomes less relevant under Sauk. Petitioner further asserts that because the audio becomes "less relevant," it would be switched to stereo. Petitioner's assertions mischaracterize teachings of both Osman and Sauk. The word "stereo" is not present anywhere in Osman's patent. Sauk never mentions assigning

audio to stereo based on the direction of the face. Petitioner's asserted combination of Osman and Sauk is incorrect because neither Osman nor Sauk teach or even suggest the recited features of claim 4.

104. In fact, Osman explicitly discloses that providing spatial awareness is more important when objects are not in view by using HRTF processing for binaural sound rendering, and not stereo or monaural or any degraded version of binaural sound. Specifically:

> [A] good HRTF model will greatly improve the accuracy of sound localization for objects outside the field of display. This is because inaccuracies in virtual surround systems are ignored by the brain in the presence of visual feedback. However, when the sound originates outside the field of view, the extra visual cue for sound localization is lost. In this case, ***a good HRTF enhances sound localization for objects outside the field of view***.

Osman, 7:31-42.

105. Petitioner does not rely on Clement, Osman, or Sauk to teach warning the user "when the user is leaving a virtual room while remaining in the 3D virtual environment," as required by claim 3 of the '184 Patent. Jang

(Ex. 1015) describes a method whereby virtual space is divided into a focus area and non-focus areas. He cues a transition from a focus area to a non-focus area by gradually decreasing or increasing the volume of sounds. However, Petitioner does not read claim 3 in context with claim 1. That is, claim 1[e] describes a method of "warning the user … by filtering, by the DSP, the binaural sound to alter a spectral cue" and claim 3 recites "wherein the warning occurs when the user leaves a virtual room while remaining in the 3D virtual environment." As such, the warning method recited in claim 3 is accomplished by methods that are not considered by Jang: "filtering, by the DSP, the binaural sound to alter a spectral cue."

XI. CONCLUSION

106. 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: April 2, 2024

By:

Durand R. Begault Ph.D.