

Filed: February 8, 2024

UNITED STATES PATENT AND TRADEMARK OFFICE  
BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

META PLATFORMS, INC.,  
Petitioner,

v.

EIGHT KHZ, LLC,  
Patent Owner.

---

IPR2023-01004  
U.S. Patent No. 9,282,196

---

**PETITIONERS' REQUEST FOR REHEARING  
UNDER 37 C.F.R. § 42.71(d)**

**TABLE OF CONTENTS**

|      |                                                                                                                               |    |
|------|-------------------------------------------------------------------------------------------------------------------------------|----|
| I.   | INTRODUCTION-----                                                                                                             | 1  |
| II.  | LEGAL STANDARD-----                                                                                                           | 3  |
| III. | ARGUMENT -----                                                                                                                | 4  |
| A.   | The Decision Misapprehended Li’s Teaching Of A Real-Time “Recording” and Overlooked Petitioner’s Evidence -----               | 4  |
| 1.   | Li Taught A “Voice Exchange” By Transmitting Real-Time “Recordings” During Conference Calls-----                              | 4  |
| 2.   | Li Taught Moving the SLP “During the Voice Exchange” By Positioning Audio Sources in Real-Time During a Conference Call ----- | 8  |
| B.   | The Decision Misapprehended Bedingfield’s Teaching of Binaural Sound and Overlooked Petitioner’s Evidence-----                | 10 |
| 1.   | The Decision Misapprehended Binaural Sound as Requiring Binaural Capture or HRTF Processing -----                             | 10 |
| 2.   | Bedingfield Taught “Providing Binaural Sound” by Delivering Different Sounds to Both Ears to Simulate Virtual Positions-----  | 12 |
| IV.  | CONCLUSION -----                                                                                                              | 15 |

**TABLE OF AUTHORITIES**

|                                                                                     | <b>Page(s)</b> |
|-------------------------------------------------------------------------------------|----------------|
| <b>Cases</b>                                                                        |                |
| <i>Ultratec, Inc. v. CaptionCall, LLC</i> ,<br>872 F.3d 1267 (Fed. Cir. 2017) ..... | 4              |
| <b>Other Authorities</b>                                                            |                |
| 37 C.F.R. § 42.71(c).....                                                           | 4              |
| 37 C.F.R. § 42.71(d) .....                                                          | 1, 3           |

## I. INTRODUCTION

Petitioner Meta Platforms, Inc. (“Petitioner” or “Meta”) respectfully requests rehearing pursuant to 37 C.F.R. § 42.71(d). The Petition sets forth grounds for invalidity based on two alternative primary references: Li (EX1006) and Bedingfield (EX1005). The Decision overlooked or misapprehended the teachings of both, and Petitioner’s evidence and explanation thereof.

*First*, the Decision misapprehended the use of the term “recording” in Li. The Decision found that the embodiments relied upon by the Petition, including the conference call embodiment, “all involve *recording* sound tracks and providing a GUI that allows the user to configure the sound that *will later be played back*.” Decision, 21 (emphasis added). But as the very citation upon which the Decision relies confirms, “The 3DSPA generates a configurable sound field on a graphical user interface using *recorded or pre-recorded* sound tracks.” *Id.*; Pet., 17, 23. Li distinguishes “recorded” (meaning real-time sound tracks) from “pre-recorded” (meaning sound tracks that are stored and are played back at a later time) and describes them in separate, alternative embodiments. However, the Decision conflated these concepts, misunderstanding Li’s use of “recording” to refer only to pre-recorded sound.

This misapprehension led the Board to incorrectly conclude that Li does not disclose moving a SLP of a voice of a computer program *during a voice exchange*. Decision, 22. But as the Petition explains, “Li discloses measuring HRTFs” to

spatialize the recorded sound at the SLP during the conference call “to mimic a real conference room environment.” Pet., 26-27. Importantly, this conference call is not pre-recorded, and the recorded sound is spatialized in real-time. Thus, Li teaches spatializing the sound *during the voice exchange* so that the conference call can proceed in real-time.

**Second**, PO misled the Board to conclude that Bedingfield must teach capturing binaural sound or expressly use the phrase “HRTFs” to disclose *providing binaural sound*. Decision, 31. This is incorrect. As Dr. Kyriakakis explained, “Binaural sound can be created in two ways. The first is to record binaural sound . . . The second is to synthesize binaural sound.” Kyriakakis Decl., ¶¶ 52-53. Bedingfield discloses the latter and is even entitled “Methods and Systems for Synthetic Audio Placement.” EX1005, (54).

The Decision overlooks the Petition’s evidence that Bedingfield discloses *providing binaural sound* using its multi-channel audio system that places participants in locations in the sound field. Dr. Kyriakakis further explained “‘binaural sound’ or binaural audio was commonly used to refer to ‘systems or techniques that capture or *simulate* ‘head-related’ signals containing natural acoustic differences between the two ears.” Kyriakakis Decl., ¶ 44; *see also* Pet., 8. Bedingfield taught simulating these natural acoustic differences between the two ears, specifically disclosing a system that “render[s] a sound field of different audio sources” that allows

“the listener of Bedingfield’s conferencing system [to] perceive[] the other participants as if they ‘were in the same location and speaking to each other directly.’” Pet., 12-13; Kyriakakis Decl., ¶¶ 79-81.

As Bedingfield explains, the sound field can be synthesized in various ways, such as by applying volume differences between the two ears to inform whether the audio source is coming from the left or right, relative volume changes to inform distance from the listener, or a “change in frequencies” to inform the audio source’s position as it moves. EX1005, 3:39-51. This processing results in the synthesis of binaural sound, as it provides different audio to both ears of a listener so that the audio appears to originate from a location within the sound field. Specifically, Bedingfield’s “synthetic audio placement allow[s] a listener to sense the location (within the sound field) of the speaker through the modified audio signal presented to the listener.” *Id.*, 4:9-12. Thus, Bedingfield teaches *providing binaural sound* by processing the audio to appear within a synthesized sound field and delivering audio to both ears. *See* Pet., 12-13, 53-57, 62; Kyriakakis Decl., ¶¶ 79-81, 198-209.

The Decision’s misapprehension of the teachings of Li and Bedingfield, and of the Petition, warrant rehearing and institution of all challenged claims.

## **II. LEGAL STANDARD**

“A party dissatisfied with a decision may file” a request for rehearing, “identify[ing] all matters the party believes the Board misapprehended or overlooked.” 37

C.F.R. § 42.71(d). The Board reviews its decision for abuse of discretion. 37 C.F.R. § 42.71(c). “The Board abuses its discretion if the decision: (1) is clearly unreasonable, arbitrary, or fanciful; (2) is based on an erroneous conclusion of law; (3) rests on clearly erroneous fact findings; or (4) involves a record that contains no evidence on which the Board could rationally base its decision.” *Ultratec, Inc. v. CaptionCall, LLC*, 872 F.3d 1267, 1272 (Fed. Cir. 2017).

### **III. ARGUMENT**

#### **A. The Decision Misapprehended Li’s Teaching Of A Real-Time “Recording” and Overlooked Petitioner’s Evidence**

##### **1. Li Taught A “Voice Exchange” By Transmitting Real-Time “Recordings” During Conference Calls**

The Decision overlooked evidence and adopted PO’s unsupported argument that Li does not disclose providing spatialized “recordings” in real-time. Decision, 20-21. The Decision erroneously conflates Li’s reference to “recording” with “pre-recording.” But Li treats each differently. In particular, Li describes a “recording” as sound captured and transmitted in real-time from a microphone array system, and a “pre-recording” as sound that has been stored and later played back. *See* Pet., 17, 23 (referring separately to a person’s voice and pre-recorded sounds).

For example, in column 5, Li discusses the spatialization of a “recording” and “pre-recording” as separate, alternative embodiments. As to a “recording,” Li explains:

The configurable sound field is configured to allow a configuration of positions and movements of the source sources. . . . In an embodiment, the 3D sound processing application dynamically processes the *recorded* sound tracks with the head related transfer functions (HRTFs) based on the acquired user selections to generate the configurable 3D binaural sound.

EX1006, 5:12-30 (emphasis added). In the subsequent paragraph, Li then addresses “pre-recording” in “another embodiment”:

In another embodiment, the method and the configurable 3D sound system disclosed herein also simultaneously generates configurable 3D sounds using sound tracks acquired from sound sources positioned in a 3D space *without using the microphone array system*. In this embodiment, the 3D sound processing application acquires the sound tracks from *pre-recorded* multiple sound tracks or *pre-recorded* stereo sound tracks.

*Id.*, 5:39-46 (emphasis added).

Unlike its pre-recorded sounds, which do not require a microphone array, Li uses microphones to “record” sounds that can be *transmitted*:

The sound sources may also comprise sources from where sound originates and *can be transmitted*. In an embodiment, the sound source is a microphone or a microphone element that records a sound track.

*Id.*, 4:64-67 (emphasis added). As Dr. Kyriakakis further explained, “Li describes the sound source as ‘similar or different sound generating devices or sound emitting devices’ including ‘a person’s voice’ and ‘pre-recorded multiple sound tracks.’”

Kyriakakis Decl., ¶ 102, *see also* Pet., 23. Accordingly, Li consistently

distinguished between the transmission of a person’s voice recording and a pre-recorded sound track.

Most critically, Li expressly confirms that its 3D sound processing application can be used “to compute 3D sound in real time.” EX1006, 35:7-8; Pet., 11-12, 16-17, 25. This disclosure directly conflicts with the statement in the Decision that “the ‘other embodiments’ [in Li] all involve *recording* sound tracks and providing a GUI that allows the user to configure the sound that will later be played back.” Decision, 21 (emphasis in original).

Indeed, Li’s express teaching to use its configurable 3D sound system during a conference call confirms that Li used “recording” to encompass real-time transmission. Kyriakakis Decl., ¶ 95; *see also* Pet., 17-18. Li states “the configurable 3D sound system 900 provides a method of communication among multiple people, for example, in a conference call, by placing different speaking users at different spots to *mimic a real conference room environment*.” EX1006, 22:1-5 (emphasis added). If the Decision were correct that Li can only spatialize pre-recorded sounds, then Li’s participants would not be able to interact in real-time. This is directly contradicted by Li’s express teachings of providing a “method of communication . . . in a conference call” that “mimic[s] a real conference environment.” *Id.* Dr. Kyriakakis explains that “[i]n modern working environments, teleconferences are a common form of group communication that allow professionals to communicate and

interact from home, in the office, and around the world.” Kyriakakis Decl., ¶ 111; *see also* Pet., 25-26. Notably, “[t]eleconferences often involve a mix of participants participating by phone as well as in person” and “can aid in communication to provide a physical representation of those participants participating by phone.” *Id.*

Thus, the plain language of Li discloses a *voice exchange*. Li does not disclose using its 3D sound system in a “pre-recorded” conference call, but in a conference call occurring in real-time. As Dr. Kyriakakis explained, “a ‘conference call’ . . . involves a *voice exchange between the processed voice of the speaker (computer program)*<sup>1</sup> and a user/person.” Kyriakakis Decl., ¶ 104; *see also* Pet., 23-24. By

---

<sup>1</sup> The Petition acknowledges that Li’s conference call embodiment does not expressly disclose a *voice of the computer program* but states that it would have been obvious to a POSA to modify the conference call embodiment to provide a computer-synthesized or processed voice of the speaker. Pet., 15-18, 23; Kyriakakis Decl., ¶¶ 88-96, 102-103. PO’s mischaracterization of the Petition misled the Board to conclude that the Petition proposes an alternative modification to include text-to-speech. The Petition does not. Rather, the Petition’s reference to text-to-speech as being generally known at the time was a motivation to modify Li. Pet., 15-18. Specifically, a POSA implementing a conference system would have recognized that “computerized voice synthesis would enable Li’s system to control for voice quality,

conflating Li’s “recording” with “pre-recording,” the Decision misapprehended Li and rests on a clearly erroneous factual finding.

**2. Li Taught Moving the SLP “During the Voice Exchange” By Positioning Audio Sources in Real-Time During a Conference Call**

Relying on its conflation of “recording” and “pre-recording,” PO misled the Board to conclude that “all examples and embodiments in Li are directed to recording, and then moving a 3D sound” and therefore, “in Li no moving of a sound can happen before first recording it.” Decision, 23 (citing POPR, 26-28). This is incorrect. In fact, this finding is contrary to one of the very problems that Li was trying to solve.

The configurable 3D sound system of Li expressly addressed the need to adjust the binaural sound location *during reproduction*:

In existing technology, . . . the recorded binaural sound cannot be modified or configured *during reproduction*. Although the existing technologies are able to achieve a few enhancements on the 3D spatial audio experience for a user, they do not provide an option for the user to adjust the source locations and directions of the recorded audio.

---

speed, and volume to improve clarity and minimize potential miscommunication.”

Kyriakakis Decl., ¶ 93; *see also* Pet., 15-17.

EX1006, 2:36-45. Dr. Kyriakakis explains that Li allows users “to provide user inputs to determine ‘movements of the sound sources’” by “generat[ing] a configurable sound field that provides an interface to give the user the freedom of configuring the positions and movements of multiple sound tracks, in order to render a customized 3D binaural sound.” Kyriakakis Decl., ¶¶ 77-78; *see also* Pet., 11-12.

By allowing users to move the sound sources during reproduction, Li confirmed that the 3D sound system could be used to improve multiple applications, from “use[] in the gaming industry to compute 3D sound *in real time*” to “enhanc[ing] the performance of radio dramas and *conference calls*.” EX1006, 35:7-8, 21:59-60 (emphasis added). “The configurable sound field is configured to allow a configuration of positions and movements of the sound sources.” *Id.*, 15:6-8. Indeed, “the user can place the sound sources in specific locations, *dynamically move the sound sources*, focus on or zoom in on one sound source and reduce others, etc., on the generated configurable sound field via the GUI.” *Id.*, 15:44-48 (emphasis added). In this manner, contrary to the Board’s conclusion, Li processes not just pre-recorded sound tracks, but also “dynamically processes 705 the recorded sound tracks using the acquired user selections to generate . . . a configurable 3D binaural sound.” *Id.*, 15:48-51.

Li allows users to move sound sources by dynamically processing the recorded sound in real-time. Thus, Li’s conference call embodiment allows users to

move the computer-synthesized voice of a speaker to a different location in real-time during a conference call. Pet., 11-12, 23-24. Dr. Kyriakakis explained “in its conferencing system embodiment Li ‘generate[s] configurable 3D sounds’, which involves processing the speaker’s voice to position ‘different speaking users at different spots.’” Kyriakakis Decl., ¶ 78; *see also* Pet., 11-12. “By allowing the user to provide commands that give the user the freedom to ‘move the sound source up and down, etc., on the configurable sound field,’ Li discloses moving the SLP of a sound based on the user’s command.” *Id.* Notably, “Li further explains that rather than merely establishing a SLP location before a voice exchange begins, the user (*user/person*) is able to actively move the SLP which would occur after an initial placement has occurred [and] the voice exchange has begun.” Kyriakakis Decl., ¶ 104; *see also* Pet., 23-24.

Thus, by teaching the user actively moving the SLP after the conference call has begun, Li taught moving the SLP during the voice exchange.

**B. The Decision Misapprehended Bedingfield’s Teaching of Binaural Sound and Overlooked Petitioner’s Evidence**

**1. The Decision Misapprehended Binaural Sound as Requiring Binaural Capture or HRTF Processing**

The Decision incorrectly requires Bedingfield to disclose capturing binaural sound or HRTFs to meet the *providing binaural sound* limitation. Decision, 31. But no such requirement exists. It is undisputed that *binaural sound* includes any

acoustical differences in the left and right ear to spatialize sound. As Dr. Kyriakakis explained, “It was well-known that ‘binaural sound’ or binaural audio was commonly used to refer to ‘systems or techniques that capture or simulate “head-related” signals containing natural acoustic differences between the two ears.’” Kyriakakis Decl., ¶ 44; *see also* Pet., 8-10. Dr. Kyriakakis further explains that such differences include interaural time differences (ITDs) and interaural level differences (ILDs), which are mechanisms known for “being responsible for source localization” and are “captured in the concept of [HRTFs].” Kyriakakis Decl., ¶¶ 46-51; *see also* Pet., 8-10. This is confirmed by PO’s own expert who admits that “the term ‘binaural sound’ is sometimes used as a proxy for 3D audio techniques” and that HRTFs “by definition will include interaural time and level differences.” EX2007, ¶¶ 22-23. Thus, *binaural sound* can reflect HRTFs, but also ITD, ILD, or other acoustical differences to localize sound. Kyriakakis Decl., ¶¶ 46-51; *see also* Pet., 8-10.

To *provide binaural sound*, Dr. Kyriakakis explained: “Binaural sound can be created in two ways. The first is to record binaural sound . . . The second is to synthesize binaural sound.” Kyriakakis Decl., ¶¶ 52-53; *see also* Pet., 8-10. With respect to the latter, Dr. Kyriakakis further explained binaural sound can be created “by processing non-binaural content with the left and right HRTFs of the desired placement angle and position.” Ultimately “[t]his processing produces a ‘virtual auditory space,’ with a two-channel synthetic binaural signal with the left channel

sent to the listener’s left ear and the right channel to the listener’s right ear (using headphones).” *Id.* ¶ 53; *see also* Pet., 8-10.

The Decision acknowledges this testimony from Dr. Kyriakakis but then goes on to agree with PO that Bedingfield does not teach synthesizing binaural sound because “[n]owhere does Bedingfield indicate the use of HRTF filtering, binaural processing, or interaural time delay.” POPR, 45. Incorrect. As discussed below, Bedingfield expressly uses binaural processing, including ILD, relative volume, and changes in frequency, to process the left and right audio to the listener, which meets the *providing binaural sound* limitation. EX1005, 3:35-61, 14:11-43, Figs. 14, 15; *see also* Pet., 65.

**2. Bedingfield Taught “Providing Binaural Sound”  
by Delivering Different Sounds to Both Ears to  
Simulate Virtual Positions**

The Decision concludes that “Petitioner does not explain in any detail *how* Bedingfield allegedly teaches or suggests ‘binaural sound’ provided to the listener who is communicating with other participants on a conference call.” Decision, 31. But the Decision overlooks the Petition’s evidence that Bedingfield discloses *providing binaural sound* using its multi-channel audio system that places participants in locations in the sound field. *See, e.g.*, Pet., 55 (“the multi-channel processing device 118 ‘creates the sound field 122 that is substantially similar . . . to an actual sound

field where the listener 102 is present between all of participants’”); *see also* Pet., 12-13, 53-57, 62; Kyriakakis Decl., ¶¶ 79-81, 198-209.

Bedingfield explains that its multi-channel audio system sends different sounds to the left and right ear of the listener to simulate the virtual positions of the participants in the sound field:

In an example, a sound field may include four participants configured in a circular configuration. With respect to a listener equipped with a multi-channel audio communication system, the listener may focus on a first speaker located at the 12 o’clock position relative to the listener and hear the speaker in stereo sound with both the right and left audio channels delivering substantially even volume levels. The listener may hear a second speaker, located at the 3 o’clock position relative to the listener, in stereo sound with an ***increased volume level in the right audio channel and a decreased volume level in the left audio channel***, as compared to the right and left volume levels received from the first speaker. Therefore, the multi-channel audio communication system may render single channel sound detected by a speaker’s microphone into a synthetic audio placement allowing a listener to sense the location (within the sound field) of the speaker through the modified audio signal presented to the listener.

EX1005, 3:62-4:12. By simulating the natural acoustic differences (*e.g.*, including volume level differences) between the left and right audio channels that are provided to both of the listener’s ears, Bedingfield is *providing binaural sound*. Kyriakakis Decl., ¶ 44; *see also* Pet., 8-10.

Indeed, Bedingfield expressly recognizes that “the human hearing mechanism . . . detects placement of a speaker (or source of sound) . . . by detecting the relative volume perceived by the listener's two ears.” EX1005, 3:35-39; *see also* 14:11-43, Figs. 14, 15. “Another way the placement of a sound source is determined is by relative volume. A signal of lower volume sounds farther away than a signal of higher volume.” *Id.* at 3:44-47. “A third way the human ear detects the position of a signal is a change in frequencies heard as a sound moves from the front of a listener's ear, where most frequencies are received with equal strength, to a position behind the listener's ear, where the higher frequencies are attenuated and the lower frequencies seem more apparent.” *Id.*, 3:49-54.

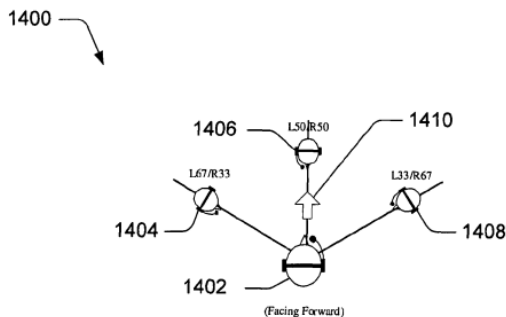


*Fig. 14*

EX1005 at Fig. 14

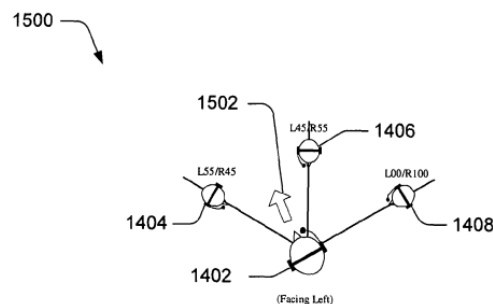


*Fig. 15*

EX1005 at Fig. 15

By adjusting the audio in both ears of the listener, “[e]ach participant is heard by the listener in multi-channel audio based on the participant’s virtual location in the sound field.” *Id.* at 3:32-34. Dr. Kyriakakis explained that “Bedingfield states

the audio signals provided to the multi-channel speaker system 120 ‘enable the listener 102 to hear the speaking participant in the sound field 122,’ which represents the perception for the listener 102 of the participants 104, 106, and 108 as if they ‘were in the same location and speaking to each other directly’ such that the listener ‘would hear sounds from the participants [] through ***both of the listener’s [] ears.***” Kyriakakis Decl., ¶ 206; *see also* Pet., 58. Dr. Kyriakakis further explains that because the sound field places the participants in specific virtual locations by “making use of both ears of a listener (*user*), Bedingfield discloses *providing binaural sound to the user.*” *Id.*

The Decision states that “Bedingfield does not use the term ‘binaural sound’ and Petitioner does not point to any other language in the reference that could be equated with, or suggest, providing binaural sound.” Decision, 30. This is incorrect. Again, the Petition repeatedly references the fact that Bedingfield’s multi-channel audio system provides audio to both ears of the listener to place participants in virtual locations within the sound field, thereby *providing binaural sound*. Pet., 12-13, 53-57, 62; Kyriakakis Decl., ¶¶ 79-81, 198-209.

#### **IV. CONCLUSION**

For the reasons above, Petitioner respectfully requests that the Board grant this request for rehearing and institute trial.

Respectfully submitted,

February 8, 2024

/Lisa K. Nguyen/  
Lisa K. Nguyen (Reg. No. 58,018)  
lisa.nguyen@allenoverly.com  
Allen & Overy LLP  
550 High Street  
Palo Alto, CA 94301  
Telephone: (650) 388-1724

**CERTIFICATE OF SERVICE**

The undersigned certifies that a complete copy of this Petitioner's Request for Rehearing Under 37 C.F.R. § 42.71(d) was served on the official correspondence address below for the patent owner via electronic mail, as agreed to by counsel, on February 8, 2024:

Kevin S. Tuttle  
Kyle L. Elliott  
Homayoon Rafatijo  
SPENCER FANE LLP  
1000 Walnut Street, Suite 1400  
Kansas City, MO 64106  
[ktuttle@spencerfane.com](mailto:ktuttle@spencerfane.com)  
[kelliott@spencerfane.com](mailto:kelliott@spencerfane.com)  
[hrafatijo@spencerfane.com](mailto:hrafatijo@spencerfane.com)

Erick S. Robinson  
Patrick M. Dunn  
Spencer Fane LLP  
3040 Post Oak Boulevard, Suite 1300  
Houston, TX 77056  
[erobinson@spencerfane.com](mailto:erobinson@spencerfane.com)  
[pdunn@spencerfane.com](mailto:pdunn@spencerfane.com)

Allen & Overy, LLP

February 8, 2024

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

Counsel for Petitioner,  
META PLATFORMS, INC.