

Case IPR2023-01003
U.S. Patent No. 9,226,090

META PLATFORMS, INC.

V.

EIGHT KHZ, LLC

October 29, 2024

Roadmap

1. Overview of Technical Concepts
2. Claim 15
3. Bedingfield/Andrea
4. Claim 1
5. Bedingfield/Andrea
6. Claim 8
7. Bedingfield/Ballas
8. Grounds Based on Li

Color Codes

1. **Gray Boxes & Highlights = Board/Fed. Cir.**
2. **Green Boxes & Highlights = PO/PO's Expert**
3. **Red Boxes & Highlights = Petitioner/Petitioner's Experts**
4. **Other Colors = Limitation-by-Limitation Analysis or Focusing on Specific Portions of Testimony or Specification**

Overview of Technical Concepts

Petitioner Uses This Definition of Binaural

sound.” POR 53. Incorrect. As PO represented in district court: “The word [binaural] has two parts: ‘bi’ meaning ‘two,’ and ‘aural’ meaning ‘of or relating to the ear or to the sense of hearing.’” EX1027, 18. And Bedingfield plainly teaches providing two different sounds to each ear.

Reply, 16



sound,” as recited in limitations 1[a] and 1[d].³ Dec. 28–31. Petitioner argued in the Petition that “[b]y rendering a sound field and providing synthetic audio placement using both ears, Bedingfield provid[es] binaural sound to the user/person.” *Id.* at 28 (quoting Pet. 58). We found that

Sur-Reply 16 (citing *IPR2023-01004*, Paper 14, 7-8)

Petitioner Misleads the Board: Parties Agreed to a Construction

As such, no construction is needed for “binaural voice,” but if the Court disagrees, it should construe the term as “representation of vocal sound waves as they would be received by each of the left and right ears of a listener, which a listener can localize” or alternatively, a “binaural sound.”

Ex.1027, 21

VIII. Joint Proposed Construction for “Binaural Voice”

After 8K submitted its responsive claim construction brief, the parties agreed that the term “binaural voice” should be construed as “vocal sound waves (including representations of same) as they would be received by each of the left and right ears of a listener, which a listener can localize.” This resolved the parties’ disputes for this term.

Ex.2024, 22

Under Petitioner's Construction Every Sound is Binaural

Q Do you agree that a person without hearing impairment -- I apologize. Strike that.

Do you agree that for a person without hearing impairment -- impairment, both stereo and binaural sounds can be heard with two ears?

ATTORNEY TENNANT: Objection. Form.

A All sounds can be heard with two ears, so -- including binaural and stereo.

Ex.2031, 57:13-18

“A Listener Can Localize” the Binaural Sound

As such, no construction is needed for “binaural voice,” but if the Court disagrees, it should construe the term as “representation of vocal sound waves as they would be received by each of the left and right ears of a listener, which a listener can localize” or alternatively, a “binaural sound.”

Ex.1027, 21

VIII. Joint Proposed Construction for “Binaural Voice”

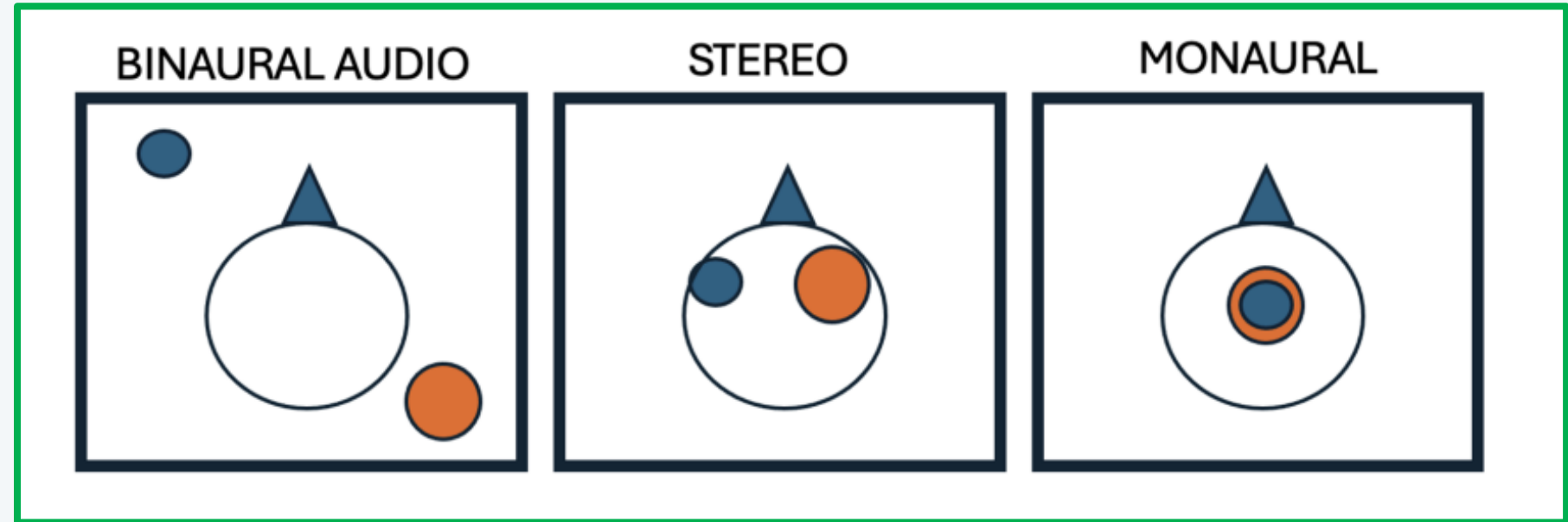
After 8K submitted its responsive claim construction brief, the parties agreed that the term “binaural voice” should be construed as “vocal sound waves (including representations of same) as they would be received by each of the left and right ears of a listener, which a listener can localize.” This resolved the parties’ disputes for this term.

Ex.2024, 22

“A Listener Can Localize” the Binaural Sound



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center



Ex.2011, ¶10

“[Binaural sound] is the only true way for producing externalization.”

POR, 43 (citing Ex.2011, ¶22)

POR, 46 n.13 (citing Ex.2011, ¶10);
POR, 42 (citing Ex.2011, ¶¶9- 22);
POR, 43 (citing Ex.2011, ¶22)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Petitioner's Expert

47. A related, but different type of sound is called stereophonic. Unlike binaural sound which is provided over wearable telephone receivers, or headphones, stereophonic sound is intended for playback over loudspeakers. Olson defines stereophonic sound as: “A stereophonic sound reproducing system is a field type sound reproducing system in which two or more microphones, used to pick up the original sound, are each coupled to a corresponding number of independent transducing channels which in turn are each coupled to a corresponding number of loudspeakers arranged, in substantial geometrical correspondence to that of the microphones, as in Fig. 4.” Ex. 1034, 1, Fig. 4 (reproduced below); *see also* Ex. 1013, ¶¶4, 16.

Ex.1003, ¶47

Petitioner's Expert's Opinion

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

Petitioner's Expert Opinion

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

‘Head-Related Signals’ Indispensable for Externalization

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1. Ex.1003, ¶46

humans are able to localize sounds positioned in the front versus the back. This is due to other factors that also impact the manner in which an ear receives sound from a point in space. Such factors include reverberation, reflection, refraction, attenuation, and even the shape of the listener’s ears, head, shoulders, and torso, which as the POSA would have readily recognized, all affect the sound that reaches a listener’s ears, depending on the environment. See, e.g., Ex. 1009 (describing a Ex.1003, ¶52

‘Head-Related Signals’ Indispensable for Externalization

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

53. These factors are captured in the concept of head-related transfer functions (HRTFs). As I wrote in 1998:
[...] Spatial information is thus encoded onto the signals received by the eardrums through a combination of direction-dependent and direction-independent filters [18], [19]. The magnitude and phase of these head-related

Ex.1003, ¶53

“Differences Between the Two Ears”

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

50. **ITD**: The ITD is the “perceived difference[] in the time of arrival of sound at the two ears.” *Id.* For example, and as the patent recognizes, when a sound

Ex.1003, ¶50

51. **ILD**: Another well-known cue for sound source localization is the ILD. For a sound source positioned closer to one ear than the other, the sound level between the two ears also differs, especially in the higher frequencies in the range

Ex.1003, ¶51

ILD and ITD “Contained” in HRTFs but not Vice Versa

In the general definition of the HRTF all linear properties of the sound transmission are included. All proposed descriptors of localization cues, such as inter-aural differences in arrival-time, ITD, in phase, IPD, in level/intensity, ILD/IID as well as monaural cues, are contained in the HRTFs. They can thus be derived from the HRTFs, whereas the opposite is not generally the case.

POR, 13 (citing Ex.2019, 240)

Petitioner's Expert Agrees

In the general definition of the HRTF all linear properties of the sound transmission are included. All proposed descriptors of localization cues, such as inter-aural differences in arrival-time, ITD, in phase, IPD, in level/intensity, **ILD/IID as well as monaural cues, are contained in the HRTFs. They can thus be derived from the HRTFs, whereas the opposite is not generally the case.**

POR, 13 (citing Ex.2019, 240)

Q Do you agree with that paragraph?

A I think it paraphrases what I said before, ITD and ILD, where ILD is the same as IID are included in the HRTF. So yeah, I agree with that.

POR, 13 (citing Ex.2016, 248:6-249:6)

Q Okay. So what does it mean to you when it says, the very last sentence: Thus it can be derived from the HRTFs where the opposite is generally not the case. What does that mean?

A If you have the HRTFs, that means you have ITD, ILD, and spectral modifications. And you can mathematically operate on the HRTFs to find what the ITD and ILD are.

But if you just had the -- if you -- what they mean is if you just had -- if you were just given ILD and ITD information, you would not be able to reconstruct the HRTFs because you don't have the spectral modifications.

The Board Agrees

In the general definition of the HRTF all linear properties of the sound transmission are included. All proposed descriptors of localization cues, such as inter-aural differences in arrival-time, ITD, in phase, IPD, in level/intensity, ILD/IID as well as monaural cues, are contained in the HRTFs. They can thus be derived from the HRTFs, whereas the opposite is not generally the case.

POR, 13 (citing Ex.2019, 240)



Ex. 1003 ¶ 53. Further, as Patent Owner correctly points out, “[s]imply disclosing some of the sound characteristics modified by the use of HRTFs does not necessarily mean that HRTFs were used to modify them.” *See* Opp. 14–15.

Sur-Reply (citing *IPR2023-01004*, Paper 14, 9)

There are TWO Ways of Creating Binaural Sound

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

54. Binaural sound can be created in two ways.

Ex.1003, ¶54

Binaural Recording: “Capture Head-Related Signals”

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

54. Binaural sound can be created in two ways. The first is to record binaural sound using two microphones that are spaced at the same distance as the width of a human head (*i.e.*, a binaural recording). This is often achieved by placing microphones on the sides of a mannequin (or dummy) head. When the left

Ex.1003, ¶54

Binaural Synthesis: Processing with HRTFs

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

55. The **second** is to **synthesize** binaural sound by processing non-binaural content with the left and right HRTFs of the desired placement angle and position.

Ex.1003, ¶55

Sur-Reply, 5-6 (citing Ex.1003, ¶46);
POR, 42 (citing Ex.2011, ¶¶9- 22);
POR, 36 (citing Ex.1003, ¶55)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

PHOSITA: Mixing Binaural Synthesis and Recording



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

25. There are two pertinent omissions in Dr. Kyriakakis's declaration: (1) the ability to combine binaural recording and binaural synthesis signals to overlay the captured sound with binaural synthesis; and (2) the ability to

Ex.2011, ¶25

A common way of making recordings in studios is to record each instrument or voice in its own channel, and then afterwards mix these down to the intended stereo perspective, using time and intensity differences. With a binaural mixing console such multi-channel recordings are transferred into binaural signals in the mixing phase. The binaural mixing console can thus be regarded as an *electronic artificial head*. If an [actual] artificial head is used for the recording, then the physical set-up in the recording studio defines the position of the sound source for the playback situation.

Ex.2011, ¶26 (quoting Ex.2019, 246)

Petitioner's Expert Provides Examples Involving HRTFs

Q Oh, a direction of the sound, where you want the sound to come from?

A Yes. Where you want it from, and it might be a different direction than where the binaural recording was made from, and that is why I was objecting to that.

Based on what we just talked about, I can take a binaural recording, treat each channel as monaural, and process it through HRTFs from a different angle than where the binaural originally came from. And, therefore, I manipulated it into being a different angle. Ex.2016, 138:1-12

Q Okay. Can you explain it again?

A It is – I'm not breaking it up. I'm just taking a two-channel signal that may have been recorded with a binaural head, putting it into my magic HRTF box, and getting out a new binaural signal. Ex.2016, 139:14-19

Q Okay. You're processing each part of the binaural signal differently?

A I'm processing it through a different set of HRTFs that would correspond to a slight difference in angle corresponding to the width of your head. Ex.2016, 140:13-18

So let's say I have a binaural recording from a certain angle. If I wanted to make it appear that it's from a different angle, I would synthesize the left channel of that binaural recording, let's say, with a 30-degree HRTF and the right channel with a 28-degree HRTF because of the separation between them. And then generate two binaural signals, again, one from each side, and combine them together to make one binaural signal at whatever that new direction is, 29 degrees or something, so – or 35 or whatever other angle. Ex.2016, 143:12-144:8

Claim 15

'090 Patent, Claim 15

15. A computer system, comprising:
a first wearable electronic device that a first person wears;
a second wearable electronic device that a second person wears and that captures a binaural voice of the second person during an electronic phone call with the first wearable electronic device that the first person wears;
a memory that stores instructions; and
a processor that executes the instructions to adjust the binaural voice of the second person during the electronic phone call in order to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person, wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.

Ex.1001, 44:58-45:4

Petitioner: Claim 15 Concerns Binaural Recording



“binaural voice” from the second person. *Id.* at 33. In its analysis, Petitioner refers to many of its earlier arguments regarding similar limitations 1[a] (capturing binaural sound), 1[c] (adjusting the binaural sound), and 1[d] (providing the binaural sound), which we address above.

Decision, 24

Andrea discloses earbuds (*electronic earphone*) worn by a user at left and right ears that “create a binaural recording.” EX1007, 14:55-58; *see also id.*, 11:66–12:3, 3:28-31, 1:44-51, 14:55-61. Thus, Andrea’s earbuds *captur[e] binaural sound*. Kyriakakis ¶¶123-124.

Petition, 22

PHOSITA: Claim 15 Requires Synthesis

15. A computer system, comprising:
a first wearable electronic device that a first person wears;
a second **wearable electronic device** that a second person wears and that **captures a binaural voice** of the second person during an electronic phone call with the first wearable electronic device that the first person wears;
a memory that stores instructions; and
a processor that executes the instructions to **adjust the binaural voice** of the second person during the electronic phone call in order **to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person, wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.**

Ex.1001, 44:58-45:4



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

59.Claim 15 of the '090 refers to an “wearable electronic device”. While this independent claim does not specify the location of **the microphones** other than to state that **they are “binaural”**, the specification states “Examples of **the wearable electronic device with microphones include, but are not limited to, earpieces, headphones, earphones, electronic glasses, hearing aids, and electronic devices that fit into or near the ear”**. Ex. 1001, 25:5-9. Claim 15, like claim 1, therefore refers either to a two-channel headphone with **two microphones mounted near the head to capture the binaural speech of a talker.**

Ex.2011, ¶59

PHOSITA: Claim 15 Requires Synthesis

15. A computer system, comprising:
a first wearable electronic device that a first person wears;
a second **wearable electronic device** that a second person wears and that **captures a binaural voice** of the second person during an electronic phone call with the first wearable electronic device that the first person wears;
a memory that stores instructions; and
a processor that executes the instructions to **adjust the binaural voice** of the second person during the electronic phone call in order **to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person, wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.**

Ex.1001, 44:58-45:4



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

POR, 39-40 (citing Ex.2011, ¶60);
Sur-Reply, 12 n.5; POR, 48-49,
Sur-Reply, 9-10

60. An electronic earphone or a wearable electronic device that is equipped with **a binaural microphone captures the spatial perspective of the talker and transmits the same to the listener.** Note that the talker's captured voice would only represent how the talker would hear their speech binaurally; it would not provide binaural cues for how a listener would hear the talker. Therefore, for use of an equivalent system for an electronic call as described in the '090 patent, binaural synthesis would be necessary to process the microphone signals. **Without binaural synthesis and only providing the unprocessed signal from the binaural microphones, a listener would only hear the talker's speech from the spatial perspective of the talker, which is either inside their head, or from their mouth.** However, for

PHOSITA: Claim 15 Requires Synthesis

15. A computer system, comprising:
a first wearable electronic device that a first person wears;
a second wearable electronic device that a second person wears and that captures a binaural voice of the second person during an electronic phone call with the first wearable electronic device that the first person wears;
a memory that stores instructions; and
a processor that executes the instructions to adjust the binaural voice of the second person during the electronic phone call in order to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person, wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.

Ex.1001, 44:58-45:4



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

86. In order to “adjust the binaural voice” of the second person during the electronic phone call in order to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person,” one would necessarily require binaural synthesis and a processor to process the voice of the talker with HRTFs.

87. Accordingly, Claim 15 is directed at binaural synthesis for the purpose of externalizing the voice of the talker for a listener. Unlike Claim 1, Claim 15 does not address the sound of the environment, and thus Claim 15 has nothing to do with binaural recording.

Ex.2011, ¶¶86-87

Petitioner's Expert Agrees

15. A computer system, comprising:
a first wearable electronic device that a first person wears;
a second **wearable electronic device** that a second person wears and that **captures a binaural voice** of the second person during an electronic phone call with the first wearable electronic device that the first person wears;
a memory that stores instructions; and
a processor that executes the instructions to **adjust the binaural voice** of the second person during the electronic phone call in order **to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person, wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.**

Ex.1001, 44:58-45:4

Q **Do the two mics capture your perspective of how you hear your voice?**

ATTORNEY TENNANT: Objection. Form.

A **They do.** Ex.2031, 35:5-8

Q So if this recording is played back for another person, how would they hear the voice in the recording?

A **You're referring to the recording made with the microphones in my ears?**

Q Yes. Thank you.

A **If they were to listen to it without any other processing, just over headphones, they would perceive me as being somewhere in the middle in front of their face.** Ex.2031, 35:19-36:6

Petitioner's Expert: Claim 15 Requires Synthesis

15. A computer system, comprising:
a first wearable electronic device that a first person wears;
a second **wearable electronic device** that a second person wears and that **captures a binaural voice** of the second person during an electronic phone call with the first wearable electronic device that the first person wears;
a memory that stores instructions; and
a processor that executes the instructions to **adjust the binaural voice** of the second person during the electronic phone call in order **to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person, wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.**

Ex.1001, 44:58-45:4

Q Okay. Well, can you look at the entire Claim 15 and tell me if any part of Claim 15 implies binaural synthesis?

ATTORNEY TENNANT: Same -- objection. Form.

A **I would say the last limitation that starts with, "A processor that executes the instructions to adjust the binaural voice," is talking about synthesis in the sense that it is intended to move to a different sound localization point.**

Ex.2031 39:7-17

Petitioner's District Court Expert Agrees

Q And virtually, then, you and I are -- I'm virtually across from you in your room, where you're physically present, and the SLP's going to be -- for me, my voice is going to be on the other side of the transparent whiteboard.

Would you use HRTFs to transform that sound to give the real impression, the perceptual plausibility, that there is a reverberation object in between the two of us?

MR. HENRY: Objection to form. Incomplete hypothetical. Ambiguous.

THE WITNESS: This hypothetical question has too many variables to actually answer. One is that it's not standard to capture the binaural voice of a person who's speaking with two microphones in their own ears.

So I wouldn't because I can't -- you can only use -- or it's most prudent to use HRTFs when you know that the two sound sources contain the same information so that when you do all these algorithms you know the source. You know that the difference that you're creating between the two ears started off with the same sound. Ex.2015, 108:11-18

15. A computer system, comprising:
a first wearable electronic device that a first person wears;
a second wearable electronic device that a second person wears and that captures a binaural voice of the second person during an electronic phone call with the first wearable electronic device that the first person wears;
a memory that stores instructions; and
a processor that executes the instructions to adjust the binaural voice of the second person during the electronic phone call in order to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person, wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.

BY MR. ELLIOTT:

Q I appreciate --

A -- the sound source itself would not be something that I would want to convolve with HRTFs because it already has some degree of bias between -- I can't readily assume that the HRTF -- sorry -- that the sounds that were captured at both of the microphones on the headset, that they have the same content.

Ex.2015, 108:1-109:9

Binaural Synthesis Requires HRTF-Processing

Q Oh, a direction of the sound, where you want the sound to come from?

A Yes. Where you want it from, and it might be a different direction than where the binaural recording was made from, and that is why I was objecting to that.

Based on what we just talked about, I can take a binaural recording, treat each channel as monaural, and process it through HRTFs from a different angle than where the binaural originally came from. And, therefore, I manipulated it into being a different angle. Ex.2016, 138:1-12

Q Okay. Can you explain it again?

A It is – I'm not breaking it up. I'm just taking a two-channel signal that may have been recorded with a binaural head, putting it into my magic HRTF box, and getting out a new binaural signal. Ex.2016, 139:14-19

Q Okay. You're processing each part of the binaural signal differently?

A I'm processing it through a different set of HRTFs that would correspond to a slight difference in angle corresponding to the width of your head. Ex.2016, 140:13-18

So let's say I have a binaural recording from a certain angle. If I wanted to make it appear that it's from a different angle, I would synthesize the left channel of that binaural recording, let's say, with a 30-degree HRTF and the right channel with a 28-degree HRTF because of the separation between them. And then generate two binaural signals, again, one from each side, and combine them together to make one binaural signal at whatever that new direction is, 29 degrees or something, so – or 35 or whatever other angle. Ex.2016, 143:12-144:8

Petitioner's Colorable Claim Differentiation Argument

Second, the dependent claims confirm that the term must encompass methods other than HRTFs. It is a basic tenet that dependent claims cannot be broader than the claim from which they depend. *Littlefuse, Inc. v. Mersen USA EP Corp.*, 29 F.4th 1376, 1380 (Fed. Cir. 2022). Here, the independent claims must be broad enough to cover at least ITD and interaural level difference (“ILD”) because the dependent claims recite adjustments to the binaural sound can include adjustments to ITD or ILD. *See, e.g.*, EX1001, claims 2, 5, 17. Thus, PO’s interpretation must be rejected because it fails to cover “explicitly claimed embodiments” in the dependent claims. *Baxalta Inc. v. Genentech, Inc.*, 972 F.3d 1341, 1346 (Fed. Cir. 2020).

Reply, 4

Petitioner's Argument Supports PO's Position

15. A computer system, comprising:
a first wearable electronic device that a first person wears;
a second wearable electronic device that a second person wears and that captures a binaural voice of the second person during an electronic phone call with the first wearable electronic device that the first person wears;
a memory that stores instructions; and
a processor that executes the instructions to adjust the binaural voice of the second person during the electronic phone call in order to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person, wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.

Q Okay. Well, can you look at the entire Claim 15 and tell me if any part of Claim 15 implies binaural synthesis?

ATTORNEY TENNANT: Same -- objection. Form.

A I would say the last limitation that starts with, "A processor that executes the instructions to adjust the binaural voice," is talking about synthesis in the sense that it is intended to move to a different sound localization point.

Ex.2031, 39:7-17

17. The computer system of claim **15**, wherein the processor further executes the instructions to change an interaural time difference (ITD), an interaural level difference (ILD), and a head-related transfer function (HRTF) in order to position the sound localization point of the binaural voice to the location that is away from and proximate to the first person.

Bedingfield/Andrea

Petitioner's Ground 1 Theory

It would have been obvious to use a WED that captures *binaural* sound, as taught in Andrea, as the communication device for Bedingfield's conference participants. Kyriakakis ¶¶96-106.

Pet. 16

more life-like conversation. Kyriakakis ¶102. Thus, a POSA would have recognized that applying Andrea's capturing *binaural* sound to Bedingfield would have enhanced the participants' experience on the electronic call. Kyriakakis ¶¶103-104.

Pet. 17

Modifying Bedingfield would have been as simple as adding a second microphone to Bedingfield's WED for capturing *binaural* voice. Indeed, Andrea's

Pet. 18

Petitioner's Ground 1 Theory

A POSA would have had a reasonable expectation of success, because the proposed modification involves nothing more than a combination of known prior art elements (microphones) according to known methods and known techniques (using multiple microphones to capture binaural voice) to yield predictable results (capture binaural sound), and involves use of a known technique to improve a similar device (WED) in the same way (adding a microphone). Kyriakakis ¶112.

Pet. 19

Binaural Recording: “Capture Head-Related Signals”

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

54. Binaural sound can be created in two ways. The first is to record binaural sound using two microphones that are spaced at the same distance as the width of a human head (*i.e.*, a binaural recording). This is often achieved by placing microphones on the sides of a mannequin (or dummy) head. When the left

Ex.1003, ¶54

Sur-Reply, 6 (citing Ex.1003, ¶46);
POR, 42 (citing Ex.2011, ¶¶9- 22);
POR, 36 (citing Ex.1003, ¶54)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Petitioner's Ground 1 Theory



during a telephone call with the participants. *Id.* Petitioner acknowledges that “Bedingfield does not expressly teach capturing a binaural sound,” but argues that doing so would have been obvious based on Andrea, which teaches “earbuds . . . worn by a user at left and right ears that ‘create a binaural recording.’” *Id.* at 22 (quoting Ex. 1007, col. 14, ll. 55–58) (emphasis omitted). According to Petitioner, a person of ordinary skill in the art would have been motivated to modify Bedingfield’s system based on Andrea because, among other reasons, capturing binaural sound would have “made it easier for the listener to distinguish between signal and noise,” “improved call quality,” “allow[ed] for a more life-like conversation,” and “enhanced the participants’ experience on the electronic call.” *Id.* at 16–18

Decision, 19

Claim 15: More than Just Adding a Second Mic

It Requires Binaural Synthesis

15. A computer system, comprising:
a first wearable electronic device that a first person wears;
a second **wearable electronic device** that a second person wears and that **captures a binaural voice** of the second person during an electronic phone call with the first wearable electronic device that the first person wears;
a memory that stores instructions; and
a processor that executes the instructions to **adjust the binaural voice** of the second person during the electronic phone call in order **to position a sound localization point of the binaural voice to a location that is away from and proximate to the first person, wherein the sound localization point represents an origin of the binaural voice of the second person to the first person.**

Ex.1001, 44:58-45:4

Sur-Reply 9-10 (citing
Ex.2031 39:7-17); POR, 52

Q Okay. Well, can you look at the entire Claim 15 and tell me if any part of Claim 15 implies binaural synthesis?

ATTORNEY TENNANT: Same -- objection. Form.

A **I would say the last limitation that starts with, "A processor that executes the instructions to adjust the binaural voice," is talking about synthesis in the sense that it is intended to move to a different sound localization point.**

Ex.2031 39:7-17

PHOSITA: Andrea is NOT about Binaural Synthesis



(Andrea 11:18-22)). The single instance in Andrea where the word “binaural” is used to refer to binaural recording, and not capturing binaural voice of a talker for the specific purpose of processing it with a binaural synthesis device to *provide* a SLP at an arbitrary location to a listener.

Ex.2011, ¶74

Petitioner's Expert Agrees

Q Okay. So you agree that Andrea discloses capturing binaural sound and voice, correct?

A **It has two microphones next to the ears, so it captures binaural everything.**

Q So is this binaural recording the binaural recording part that we talked about with creating binaural sound?

A **Assuming that what was captured is listened to by – over headphones, then yes.**

Q **So does Andrea disclose the second prong, binaural synthesis using HRTF?**

A **I did not see that in Andrea nor did I use that.**

Ex.2016, 259:2-14

Petitioner's New Theory at Reply

binaural sound, adjusts that sound, then provides it to the listener. As Dr. Kyriakakis explained, a POSITA would have modified the Bedingfield system by adding both “an additional microphone to Bedingfield’s *electronic earphone* and *a software module to process the binaural sound* as taught by Andrea.” Kyriakakis, ¶125; Pet.

Reply, 15

Petitioner's Theory at Reply

binaural sound, adjusts that sound, then provides it to the listener. As Dr. Kyriakakis explained, a POSITA would have modified the Bedingfield system by adding both “an additional microphone to Bedingfield’s *electronic earphone* and *a software module to process the binaural sound* as taught by Andrea.” Kyriakakis, ¶125; Pet.

Reply, 15

by Andrea.” Reply, 15. However, Andrea never mentions any software module—this is another baseless claim by Dr. Kyriakakis. Even so, Dr. Kyriakakis admitted that Andrea does not involve binaural synthesis. Ex.2016, 259:2-14. Furthermore, Petitioner provides nothing describing this “software module.” Thus, the rationale

Sur-Reply, 15

PHOSITA: Bedingfield is NOT about Binaural Synthesis



43. In summary: Bedingfield (1) never mentions the term HRTF or discusses HRTF spectra, which includes time differences as part of the spectral transfer function in terms of its phase components; (2) never mentions altering frequencies to correspond to the HRTF spectrum effected by a listener's ears, and in fact is at odds with the characteristics of HRTF spectra; (3) never mentions sound source externalization. In fact, Bedingfield only describes a basic stereo technique where sound sources are only separated by amplitude differences ("This stereo effect is used advantageously in nearly all modern music reproduction systems.").

Ex.2011, ¶43

PHOSITA: Bedingfield is NOT about Binaural Synthesis



stereo system.” *Id.* All that Bedingfield is discussing is a *relative change* in *overall* high-frequency content (essentially the same as turning the treble control down on a stereo loudspeaker tone control) to help distinguish “back” for “front.” Such a technique would be immediately recognized by a POSITA as insufficient and uncharacteristic of the HRTFs used in binaural synthesis, and more similar to stereo audio playback.

Ex.2011, ¶41

PHOSITA: Bedingfield is NOT about Binaural Synthesis



50. A POSITA would therefore not look to Bedingfield as even remotely similar to the technology of binaural synthesis that is based on HRTFs. Rather, it suggests stereophonic or (in Figure 16) multi-channel loudspeaker placement. Furthermore, Bedingfield does not disclose ITD or altering ITD to move in response to head orientation of the listener.

Ex.2011, ¶50

The Board: Bedingfield does NOT Teach Binaural Sound Creation



Ex.2017, 31

Ex. 1003 ¶¶ 52–53 (citations and emphasis omitted). We agree with Patent Owner that Bedingfield does not teach either method. See Prelim. Resp. 45–47. In any case, 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.⁴

binaural sound; (2) “Bedingfield does not use the term ‘binaural sound’ and Petitioner [did] not point to any other language in the reference that could be equated with, or suggest, providing binaural sound”; and (3) Petitioner had not shown that Bedingfield teaches either of the two ways for creating binaural sound described by Dr. Kyriakakis, and did not “explain in any detail *how* Bedingfield allegedly teaches or suggests ‘binaural sound’

IPR2023-01004, Paper 14, 7-8

POR, 11 (citing Ex.2017, 31);
Sur-Reply, 16 (citing IPR2023-
01004, Paper 14, 7-8)

Petitioner: Bedingfield is a Binaural Sound Reference

sound.” POR 53. Incorrect. As PO represented in district court: “The word [binaural] has two parts: ‘bi’ meaning ‘two,’ and ‘aural’ meaning ‘of or relating to the ear or to the sense of hearing.’” EX1027, 18. And Bedingfield plainly teaches providing two different sounds to each ear.

Reply, 16

hearing.” Ex. 1027 at 18. I agree, and Bedingfield discloses providing binaural sound because Bedingfield teaches providing different sounds to each ear.

Ex.1055, ¶44

Petitioner Misleads the Board: Parties Agreed to a Construction

As such, no construction is needed for “binaural voice,” but if the Court disagrees, it should construe the term as “representation of vocal sound waves as they would be received by each of the left and right ears of a listener, which a listener can localize” or alternatively, a “binaural sound.”

Ex.1027, 21

VIII. Joint Proposed Construction for “Binaural Voice”

After 8K submitted its responsive claim construction brief, the parties agreed that the term “binaural voice” should be construed as “vocal sound waves (including representations of same) as they would be received by each of the left and right ears of a listener, which a listener can localize.” This resolved the parties’ disputes for this term.

Ex.2024, 22

Under Petitioner's Construction Every Sound is Binaural

Q Do you agree that a person without hearing impairment -- I apologize. Strike that.

Do you agree that for a person without hearing impairment -- impairment, both stereo and binaural sounds can be heard with two ears?

ATTORNEY TENNANT: Objection. Form.

A All sounds can be heard with two ears, so -- including binaural and stereo.

Ex.2031, 57:13-18

Petitioner: Bedingfield is a Binaural Sound Reference

Bedingfield explains that its multi-channel audio system sends different sounds to the listener's left and right ears to simulate virtual positions of the participants in the sound field. For example, in a sound field that includes four participants in a circular configuration, a "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." EX1005, 3:62-4:12, Figs. 14, 15, 19a.

Reply, 17

Petitioner Misleads the Board: Bedingfield Sends STEREO Sounds

Bedingfield explains that its multi-channel audio system sends different sounds to the listener's left and right ears to simulate virtual positions of the participants in the sound field. For example, in a sound field that includes four participants in a circular configuration, a "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." EX1005, 3:62-4:12, Figs. 14, 15, 19a.

Reply, 17

Petitioner's Expert: Binaural Synthesis Results in a Two-Channel Binaural Signal

55. The **second** is to **synthesize** binaural sound by processing non-binaural content with the left and right HRTFs of the desired placement angle and position. For example, a monaural sound source, such as a person's voice, is processed (via convolution) with a left ear HRTF and a right ear HRTF, where each HRTF is designated at the desired placement angle and position, such as 30° to the front left of the listener and at a specified distance from the listener. Ex. 1025, 5; *see also* Ex. 1013, ¶¶11-12, 17, 20, 71, 75-77. This 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).

Ex.1003, ¶55

Petitioner: Bedingfield is a Binaural Sound Reference

externalization.” Begault, ¶ 43. Although Bedingfield does not use the terms HRTF or externalization, it is my opinion that both are disclosed.

Ex.1055, ¶43

46. This is seen in Figures 14 and 15, where Bedingfield shows different volume levels in the left channel and right channel depending where the sound appears to originate in relation to the listener.

Ex.1055, ¶46

49. Bedingfield teaches that the volume change will localize the sound behind the listener. Ex. 1005, 13:30-43 (in FIG. 15, the participant 1408 that is now speaking will sound as if the speaker were *located behind* the listener 1402.”). A

Ex.1055, ¶49

9:50-57. Notably, Bedingfield’s teaching of volume differences to localize sound is reflected in the ’090 Patent. EX1001, 12:27-30.

Reply, 18

ILD and ITD “Contained” in HRTFs but not Vice Versa

In the general definition of the HRTF all linear properties of the sound transmission are included. All proposed descriptors of localization cues, such as inter-aural differences in arrival-time, ITD, in phase, IPD, in level/intensity, **ILD/IID** as well as monaural cues, are contained in the HRTFs. They can thus be derived from the HRTFs, whereas the opposite is not generally the case.

POR, 13 (citing Ex.2019, 240)

Petitioner's Expert Agrees

In the general definition of the HRTF all linear properties of the sound transmission are included. All proposed descriptors of localization cues, such as inter-aural differences in arrival-time, ITD, in phase, IPD, in level/intensity, **ILD/IID as well as monaural cues, are contained in the HRTFs. They can thus be derived from the HRTFs, whereas the opposite is not generally the case.**

POR, 13 (citing Ex.2019, 240)

Q Do you agree with that paragraph?

A I think it paraphrases what I said before, ITD and ILD, where ILD is the same as IID are included in the HRTF. So yeah, I agree with that.

POR, 13 (citing Ex.2016, 248:6-249:6)

Q Okay. So what does it mean to you when it says, the very last sentence: Thus it can be derived from the HRTFs where the opposite is generally not the case. What does that mean?

A If you have the HRTFs, that means you have ITD, ILD, and spectral modifications. And you can mathematically operate on the HRTFs to find what the ITD and ILD are.

But if you just had the -- if you -- what they mean is if you just had -- if you were just given ILD and ITD information, you would not be able to reconstruct the HRTFs because you don't have the spectral modifications.

The Board Agrees

In the general definition of the HRTF all linear properties of the sound transmission are included. All proposed descriptors of localization cues, such as inter-aural differences in arrival-time, ITD, in phase, IPD, in level/intensity, ILD/IID as well as monaural cues, are contained in the HRTFs. They can thus be derived from the HRTFs, whereas the opposite is not generally the case.

POR, 13 (citing Ex.2019, 240)



Ex. 1003 ¶ 53. Further, as Patent Owner correctly points out, “[s]imply disclosing some of the sound characteristics modified by the use of HRTFs does not necessarily mean that HRTFs were used to modify them.” *See* Opp. 14–15.

Sur-Reply (citing IPR2023-01004, Paper 14, 9)

“Head-Related Signals” Enable External Localization

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

humans are able to localize sounds positioned in the front versus the back. This is due to other factors that also impact the manner in which an ear receives sound from a point in space. Such factors include reverberation, reflection, refraction, attenuation, and even the shape of the listener’s ears, head, shoulders, and torso, which as the POSA would have readily recognized, all affect the sound that reaches a listener’s ears, depending on the environment. *See, e.g.*, Ex. 1009 (describing a

Ex.1003, ¶52

“Head-Related Signals” Enable External Localization

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

53. These factors are captured in the concept of head-related transfer functions (HRTFs). As I wrote in 1998:
[...] Spatial information is thus encoded onto the signals received by the eardrums through a combination of direction-dependent and direction-independent filters [18], [19]. The magnitude and phase of these head-related

Ex.1003, ¶53

Claim 1

Claim 1: Scope and Meaning

1. A method, comprising:
capturing, with an electronic earphone located at a head of a talking person, binaural sound that will be provided to a listening person during a telephone call;
designating, with a computer system, a sound localization point in empty space that is away from and proximate to the listening person such that the sound localization point simulates an origin of the binaural sound at the empty space that the listening person hears during the telephone call;
adjusting, with the computer system, the binaural sound captured at the earphone of the talking person so the binaural sound originates during the telephone call from the sound localization point in empty space that is away from and proximate to the listening person; and
providing, with an electronic earphone located at a head of the listening person, the binaural sound to the listening person during the telephone call such that the origin of the binaural sound for the listening person occurs at the sound localization point in empty space that is away from and proximate to the listening person.

Ex.1001, 42:35-55

Claim 1: Scope and Meaning



Independent claim 15 is somewhat similar to claim 1, but recites a computer system rather than a method. Petitioner argues that Bedingfield

“binaural voice” from the second person. *Id.* at 33. In its analysis, Petitioner refers to many of its earlier arguments regarding similar limitations 1[a] (capturing binaural sound), 1[c] (adjusting the binaural sound), and 1[d] (providing the binaural sound), which we address above.

Decision, 24

Claim 1's "Sound" Includes the "Talking Person" Voice

Limitation 1[A]

1. A method, comprising:
capturing, with an electronic earphone located at a head of
a talking person, binaural sound that will be provided to
a listening person during a telephone call;

⁸ The Petition applies PO's interpretation that *telephone call* is "voice communication over an electronic network between at least two parties." See EX1027, 4.

Pet. 19 n.8

Q And what I asked you was whether or not this included -- would be live voice exchange.

A So there is live voice exchange during a telephone call, but I was trying to say before we ended that other stuff like background noise and things could also go through there. So it's not the only thing is what I'm trying to say.

Ex.2016, 176:21-177:6

Claim 1's "Sound" = Vocal + Non-Vocal Components

EX1051, 41:3-5. Thus, a voice is "sound," as is a door opening, footsteps, or a bird chirping. Importantly, each of these sounds by themselves would meet the "sound" limitation of claim 1, but not the "voice" limitation of claims 8 and 15.

Reply, 11-12

Limitation 1[A]

1. A method, comprising:
capturing, with an electronic earphone located at a head of
a talking person, binaural sound that will be provided to
a listening person during a telephone call;

Q And what I asked you was whether or not this included -- would be live voice exchange.

A So there is live voice exchange during a telephone call, but I was trying to say before we ended that other stuff like background noise and things could also go through there. So it's not the only thing is what I'm trying to say.

Ex.2016, 176:21-177:6

Reply, 11-12, POR, 21 (citing Ex.2016, 176:10-177:6); Sur-Reply, 11-13

Claim 1: Mix of Binaural Recording and Synthesis



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

68. Claim 1 also indicates how both binaural synthesis and binaural recording might occur simultaneously during a telephone call. The specification gives an example of a binaural recording situation where a sound source is later transformed using binaural synthesis “later in a call”:

Ex.2011, ¶68

Claim 1: Mix of Binaural Recording and Synthesis

Consider an example in which a listener and a talking person wear or use an electronic device that captures, transmits, and/or provides binaural sound during a telephone call. In the first part of the call, a VMP at the location of the speaker is designated to be at the point of the actual physical microphones being worn by the speaker (e.g., being at the ears of the speaker). As such, various sounds in the environment of the speaker are perceived by the listener to emanate from locations relative to the ears of the listener, and the SLP of the speaker's voice is perceived by the listener as approximately at the point of the mouth of the listener. Later in the call it is

determined that the vocal part of the sound being sent by the speaker should be moved to a SLP off to the right side of the face of the listener. By way of example, the vocal part of the sound captured by the device worn by the speaker can be identified by the intersection of the set of frequencies associated with voice and with amplitudes and sound pressures matching those that likely emanated from the oral cavity a few inches away from the microphones, sounds matching a pre-defined voiceprint of the speaker, sounds associated with

Ex. 1001 38:57-39:9

POR, 41-42 (citing Ex. 1001
38:57-39:9); POR, 41-42
(citing Ex.2011, ¶68)

Claim 1: Mix of Binaural Recording and Synthesis



Ex.2011, ¶69

69.As this embodiment explains, the sound being captured necessarily contains binaural content (sound of the environment) and non-binaural content (voice of the speaker, i.e., the vocal part of the sound). However, the sound that is provided to the listener is binaural because the listener moves the SLP of the vocal part of the sound through limitations [b] and [c]. This is similar to the “binaural mixing console” discussed in Paragraph 26 above. In fact, the ‘090 specification states that the invention can distinguish between the voice of the speaker and the sound of the environment present in the speaker’s room, and the binaural synthesis only applies to the voice of the speaker. *See* Ex. 1001 38:57-39:24 (“[T]he vocal

PHOSITA: Combining Binaural Synthesis & Recording



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

25. There are two pertinent omissions in Dr. Kyriakakis's declaration: (1) the ability to combine binaural recording and binaural synthesis signals to overlay the captured sound with binaural synthesis; and (2) the ability to

Ex.2011, ¶25

A common way of making recordings in studios is to record each instrument or voice in its own channel, and then afterwards mix these down to the intended stereo perspective, using time and intensity differences. With a binaural mixing console such multi-channel recordings are transferred into binaural signals in the mixing phase. The binaural mixing console can thus be regarded as an *electronic artificial head*. If an [actual] artificial head is used for the recording, then the physical set-up in the recording studio defines the position of the sound source for the playback situation.

Ex.2011, ¶26 (quoting Ex.2019, 246)

Claim 1: Mix of Binaural Recording and Synthesis



Ex.2011, ¶70

70. Dr. Kyriakakis never acknowledges or recognizes the use of binaural synthesis in Claim 1. However, because the application of Claim 1 is to a telephone call, it only makes sense to use binaural synthesis for the voice of a talker to be externalized at a sound localization point away from and proximate to the listener. Otherwise, by only using binaural recording from microphones located on the talker's earphones, the listener would have the unusual experience of hearing the talkers' voice inside of their head or in front of their mouth (See discussion in Section V.E. above). The talker's voice would not originate for the listener from a "...sound localization point in empty space that is away from and proximate to the listening person" as specified in Claim 1.

Petitioner's Expert Agrees

Q Do the two mics capture your perspective of how you hear your voice?

ATTORNEY TENNANT: Objection. Form.

A **They do.**

Ex.2031, 35:5-8

Q So if this recording is played back for another person, how would they hear the voice in the recording?

A You're referring to the recording made with the microphones in my ears?

Q Yes. Thank you.

A If they were to listen to it **without any other processing**, just over headphones, they would perceive me as being **somewhere in the middle in front of their face.**

Ex.2031, 35:19-36:6

Claim 1: Mix of Binaural Recording and Synthesis



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

Ex.2011, ¶71

71. Furthermore, if Claim 1 was solely directed at binaural recording as Dr. Kyriakakis characterizes, then the Claim would make no sense to a POSITA, because a POSITA would know that the SLP of a sound created by binaural recording is fixed and cannot be arbitrarily transformed to new locations, as I explained above in Section V.B. Therefore, from a technical perspective, limitation 1[b] and 1[c] would make no sense to a POSITA considering Claim 1 to be exclusively about binaural recording.

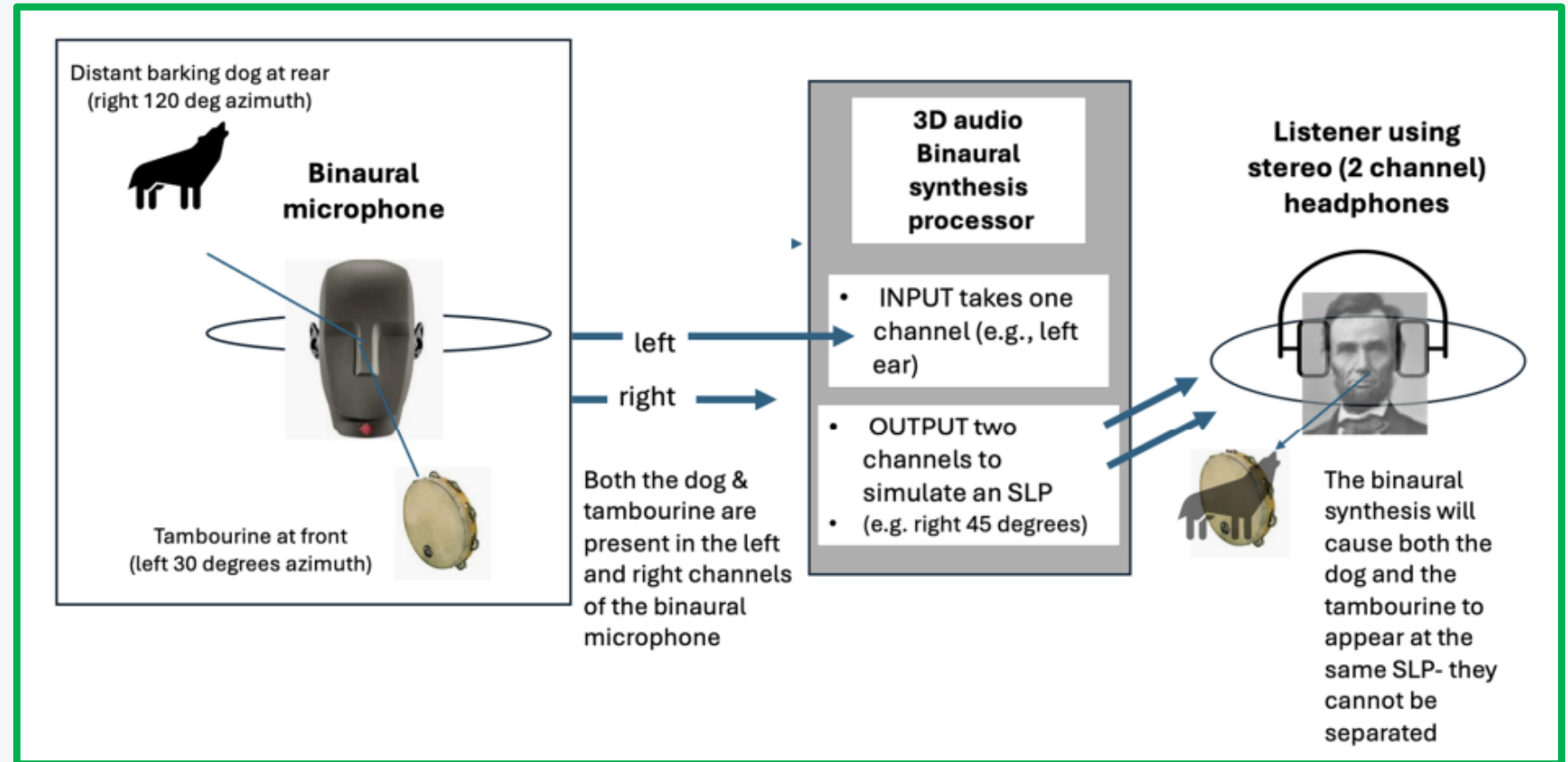
Limitation 1[C]

adjusting, with the computer system, the binaural sound captured at the earphone of the talking person so the binaural sound originates during the telephone call from the sound localization point in empty space that is away from and proximate to the listening person; and

The SLP of a Binaural Recording is Fixed



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center



“A binaural recording would capture the spatial relationship between the dog barks and the tambourine; their distances and angle of incidence to the dummy head are distinct.”

“[S]umming the two channels or taking only one channel of the binaural recording would include both the tambourine and the dog barks, and one would end up placing both the dog and the tambourine at one synthesized position.”

Ex.2011, ¶29

Petitioner Demonstrates Misunderstanding of the Technology

Critically, the claim language as recited requires “*the* binaural sound captured” to originate from “*the* sound localization point.” Thus, the claim language does not make sense under PO’s interpretation, which would require all binaural sounds to be captured and then adjusted to originate from the *same* sound localization point. For example, if the binaural sounds of the mother’s voice and a bird at different locations were captured, under PO’s interpretation, the claim would require both to be localized at the same point—“capturing each and every sound wave at the talker’s system, adjusting the same, and then providing the adjusted sound to the listener.” POR 1. But these two sources of binaural sound do not share the same sound localization point.

Reply, 10-11

Sound Provided to the Listener from a Single SLP

1. A method, comprising:

capturing, with an electronic earphone located at a head of a talking person, binaural sound that will be provided to a listening person during a telephone call;

designating, with a computer system, a sound localization point in empty space that is away from and proximate to the listening person such that the sound localization point simulates an origin of the binaural sound at the empty space that the listening person hears during the telephone call;

adjusting, with the computer system, the binaural sound captured at the earphone of the talking person so the binaural sound originates during the telephone call from the sound localization point in empty space that is away from and proximate to the listening person; and

providing, with an electronic earphone located at a head of the listening person, the binaural sound to the listening person during the telephone call such that the origin of the binaural sound for the listening person occurs at the sound localization point in empty space that is away from and proximate to the listening person.

“PO never argued multiple SLPs for the sound provided to the listener; both vocal and non-vocal components are delivered to the listener at a single designated SLP.”

Sur-Reply, 12

Ex.1001, 42:35-55

Petitioner's Colorable Claim Differentiation Argument

Second, the dependent claims confirm that the term must encompass methods other than HRTFs. It is a basic tenet that dependent claims cannot be broader than the claim from which they depend. *Littlefuse, Inc. v. Mersen USA EP Corp.*, 29 F.4th 1376, 1380 (Fed. Cir. 2022). Here, the independent claims must be broad enough to cover at least ITD and interaural level difference (“ILD”) because the dependent claims recite adjustments to the binaural sound can include adjustments to ITD or ILD. *See, e.g.*, EX1001, claims 2, 5, 17. Thus, PO’s interpretation must be rejected because it fails to cover “explicitly claimed embodiments” in the dependent claims. *Baxalta Inc. v. Genentech, Inc.*, 972 F.3d 1341, 1346 (Fed. Cir. 2020).

Reply, 4

Petitioner's Argument Supports PO's Position

1. A method, comprising:

capturing, with an electronic earphone located at a head of a talking person, binaural sound that will be provided to a listening person during a telephone call;

designating, with a computer system, a sound localization point in empty space that is away from and proximate to the listening person such that the sound localization point simulates an origin of the binaural sound at the empty space that the listening person hears during the telephone call;

adjusting, with the computer system, the binaural sound captured at the earphone of the talking person so the binaural sound originates during the telephone call from the sound localization point in empty space that is away from and proximate to the listening person; and

providing, with an electronic earphone located at a head of the listening person, the binaural sound to the listening person during the telephone call such that the origin of the binaural sound for the listening person occurs at the sound localization point in empty space that is away from and proximate to the listening person.

2. The method of claim 1 further comprising:

tracking a head orientation of the listening person with respect to the sound localization point;

adjusting an interaural time difference of the sound that the listening person hears in response to the listening person changing the head orientation with respect to the sound localization point.

Bedingfield/Andrea

Bedingfield/Andrea Fails for Neither Teaches Binaural Synthesis



Q Okay. So you agree that Andrea discloses capturing binaural sound and voice, correct?

A **It has two microphones next to the ears, so it captures binaural everything.**

Q So is this binaural recording the binaural recording part that we talked about with creating binaural sound?

A **Assuming that what was captured is listened to by – over headphones, then yes.**

Q **So does Andrea disclose the second prong, binaural synthesis using HRTF?**

A **I did not see that in Andrea nor did I use that.**

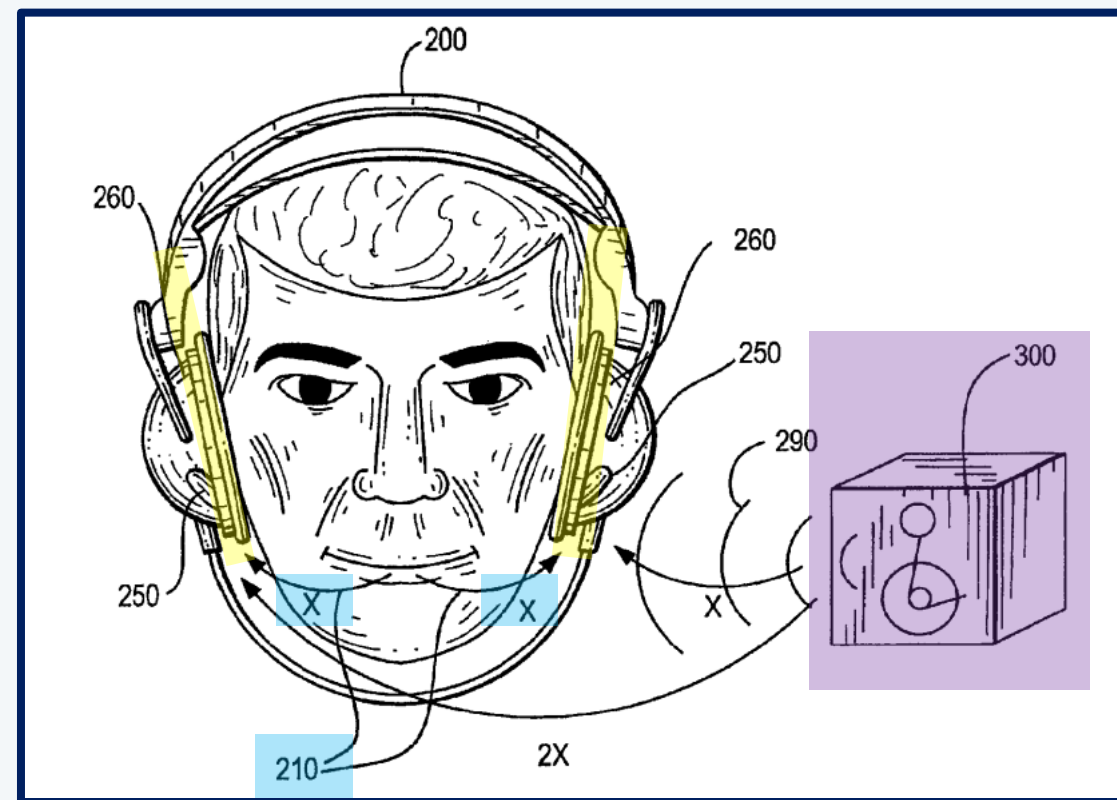
Ex.2016, 259:2-14

Ex. 1003 ¶¶ 52–53 (citations and emphasis omitted). We agree with Patent Owner that Bedingfield does not teach either method. See Prelim. Resp. 45–47. In any case, 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.⁴

Ex.2017, 31

Additionally, Andrea's Second Mic is for Noise Cancellation

the microphones. From the centered position of the mouth, the voice signal 210 will look like a plain wave and arrive in-phase at same time with equal amplitude at both the microphones, while noise from the sides will arrive at each microphone in different phase/time and be cancelled by the adaptive processing of the algorithm. Illustration of such an instance is clearly shown in FIG. 10, for example, where noise coming from a speaker 300 on one side of the user is cancelled due to varying distances (X , $2X$) of the sound waves 290 from either microphone 250. However, the voice signal 210 travels an equidistant (Y) to both microphones 250, thus providing for a high fidelity far field noise canceling microphone that possesses good background noise cancellation and that may be used in any type of noisy environment, especially in environments where a lot of music and speech may be present as background noise (as in a game arena or internet cafe).



Ex.1007, 11:6-24

PHOSITA: Andrea's Second Mic is for Noise Cancellation



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

76. Andrea uses two microphones as part of a beamforming system, such that the talker's voice is emphasized by the directionality of the microphone beam, and surrounding sound is de-emphasized as a form of noise cancellation. See Figure 10 of Andrea where an external loudspeaker to the side of the talker generates a noise signal. Andrea's two microphones work in tandem to cancel this noise while providing a directional beam for the talker's voice. Therefore, if one were to add Andrea's noise cancellation microphone system to Bedingfield's headphones, one would cancel or minimize surrounding sound. This is the opposite of the goal of Claim 1, which is to capture the voice of the talker *and* surrounding sounds in the environment.

Ex.2011, ¶76

Petitioner's Expert Agrees

Bedingfield's stated goal to improve sound quality. Specifically, an objective of Bedingfield is to "allow a person to hear the speaker better" during an electronic call, like the exemplified conference call. *See* Ex. 1005, 1:44-46. A person wearing a wearable electronic communication device during an electronic call may be at a location with a lot of ambient noise. Ambient noise is the background or environmental sound that is present in a recording or listening situation. When a microphone on a wearable device picks up the person's voice, the background or environmental sound is also captured by the microphone. This can often interfere with the intelligibility of the intended sound source and reduce the quality and clarity of a person's voice.

Ex.1003, ¶98

Petitioner's Expert Agrees

100. During the relevant period (and well before) a known solution to this problem was noise cancelation as taught in Andrea.

101. Andrea discloses a noise canceling solution that may be employed for electronic communications, including for “directional listening.” Ex. 1007, 1:44-51.

Andrea teaches that the noise canceling solution requires two or more microphones that capture binaural sound. Ex. 1007, 6:13-14 (“In the simplest embodiment the array comprises two microphones”), 11:25-27 (“The two elements or microphones

Ex.1003, ¶¶100-01

PHOSITA: Andrea's Second Mic is NOT Used for Processing Vocal Component of Sound



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

77. In summary: Andrea does not teach capturing binaural voice of a talker for the specific purpose of processing it using binaural synthesis to *provide* a SLP at an arbitrary location to a listener.

Ex.2011, ¶77

Claim 8

Claim 8: Scope and Meaning

8. A non-transitory computer readable storage medium storing instructions that cause a computer system to execute a method, comprising:

commence an electronic call between a first individual wearing an electronic earphone and a second individual that are remote from each other;

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

execute the electronic call such that the origin of the voice of the second individual appears to the first individual to originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and

maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first individual hears in response to the first individual moving the head orientation with respect to the sound localization point.

Ex.1001, 43:39-61

Claim 8: Voice Presented to Listener in 8[C] then Altered in 8[D]

Limitation 8[C]

execute the electronic call such that the origin of the voice of the second individual appears to the first individual to originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first individual hears in response to the first individual moving the head orientation with respect to the sound localization point.

Limitation 8[D]

Q So I want you to look at the claim and I want you to tell me at what limitation in Claim 8, the listener, first individual, first hears the voice of the speaker, the second individual.

ATTORNEY TENNANT: Objection to form. Lack of foundation.

A I'm not sure I understand what you mean by "first." Are you saying when the call starts or what – I'm not understanding the question.

Q Yes.

A In this language, perhaps it means it's the third one, where it says, "Execute the electronic call." It's not really clear.

Q Okay. So that would be -- you're talking about the limitation that starts with "Execute the electronic call," correct?

A Correct.

Ex.2031, 25:14-26:8

Claim 8 Simulates HRTFs—Though Less Accurately

8. A non-transitory computer readable storage medium storing instructions that cause a computer system to execute a method, comprising:

commence an electronic call between a first individual wearing an electronic earphone and a second individual that are remote from each other;

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

execute the electronic call such that the origin of the voice of the second individual appears to the first individual to originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and

maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first individual hears in response to the first individual moving the head orientation with respect to the sound localization point.

10. The non-transitory computer readable storage medium storing instructions of claim 8 further to cause the computer system to execute the method comprising:

monitor movements of the head orientation of the first individual in response to the first individual localizing the voice of the second individual during the electronic call;

adjust head-related transfer functions (HRTFs) of the first individual to more accurately simulate real HRTFs of the first individual in response to the movements of the head orientation of the first individual localizing the voice of the second individual during the electronic call.

Binaural Synthesis: “Simulate Head-Related Signals”

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

55. The **second** is to **synthesize** binaural sound by **processing non-binaural content with the left and right HRTFs of the desired placement angle and position.**

Ex.1003, ¶55

Sur-Reply, 6 (citing Ex.1003, ¶46);
POR, 42 (citing Ex.2011, ¶¶9- 22);
POR, 36 (citing Ex.1003, ¶55)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

PHOSITA Agrees: Claim 8 Simulates HRTFs



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

orientation....” A POSITA reading both Claims 8 and 10 would understand that, Claim 8’s reference to altering ITD is also “simulate real HRTFs of the first individual in response to the movements of the head orientation.” The only difference is that ITD-based simulations are less accurate comparing to those based on “adjust[ing] head-related transfer functions.”

Ex.2011, ¶79

PHOSITA Agrees: Claim 8 Simulates HRTFs



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

80. In sum, both Claims 8 and 10 use HRTFs to externalize the voice of the talker with respect to the listener at a location away from and proximate to the listener, thereby claiming binaural synthesis. The only meaningful difference between the two claims is that Claim 8 alters ITD in response to head orientation of the listener to maintain the SLP of the voice, whereas Claim 10 adjusts the entire HRTFs of the first individual to accomplish the same. The latter approach (i.e., Claim 10) improves accuracy but is also computationally more expensive than the former (i.e., Claim 8).

Ex.2011, ¶80

The Voice Presented to the Listener is External to the Listener

Limitation 8[B]

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

Ex.2016, 193:5-19

Q It is external, right?

MS. WANG: Asked and answered.

Q It is external, correct?

MS. WANG: Asked and answered.

A According to Eight KHz, that is the definition I used in my -- in my report.

Q And the sound originates externally from the listener, correct?

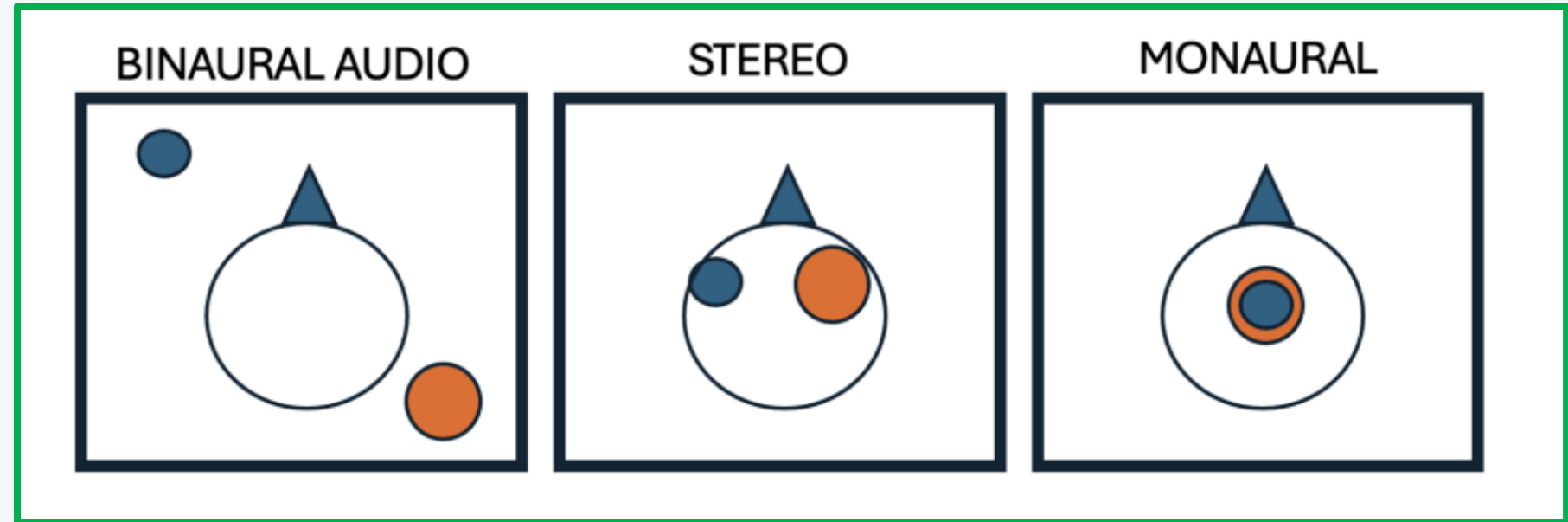
A The sound appears to the listener. There is no actual sound, right? It is in that room. The sound is synthesized. It appears to the listener to be away from and proximate.

Q Which is external to the user, correct?

MS. WANG: Asked and answered.

A Using that definition from Eight KHz, yes.

“A Listener Can Localize” the Binaural Sound



Ex.2011, ¶10

“[Binaural sound] is the only true way for producing externalization.”

POR, 43 (citing Ex.2011, ¶22)

POR, 46 n.13 (citing Ex.2011, ¶10);
POR, 42 (citing Ex.2011, ¶¶9- 22);
POR, 43 (citing Ex.2011, ¶22)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Petitioner Expert: Externalization & Binaural Sound

46. In the 1980s, Frederic Wightman and Doris Kistler at the University of Wisconsin published a set of papers that described the cues for locating external sound sources and methods to accurately simulate the same set of cues over headphones. Specifically, the researchers demonstrated how they were able to accurately simulate external sound sources at an SLP in an empty space (described as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Petitioner Evidence Ties Together Externalization & HRTFs

SIMULATING EXTERNAL SOURCES OVER HEADPHONES

If one really understood all the cues responsible for locating external sound sources, then one should be able to simulate that same set of cues using headphones. This simple, but surprisingly daunting, challenge is the essential premise of the research program initiated by Wightman & Kistler at the University of Wisconsin. In their papers (Wightman & Kistler 1989a,b; Wightman et al 1987), they outline a program of research to compare sound localization in the free field and under headphone listening. The first step in this process is to measure directional transfer functions for multiple sound source locations around the listener. Six loudspeakers face the listener on a circular arc that is rotated about the subject in 15° steps of azimuth. Speakers are spaced from -36° to $+54^\circ$ elevation, in 18° steps. A broadband stimulus

...

location in degrees of elevation and azimuth. The general conclusion is that "The data . . . from the headphone condition are nearly identical to the data from the free-field condition" (Wightman & Kistler 1989b). Thus, to within the error of measurement associated with these experiments, the transfer functions contain all the information needed to simulate the localization of sound sources in the free field.

Ex.1026, 154

Petitioner Evidence Ties Together Externalization & HRTFs

An important practical application of HRTFs is to convolve them with another sound, such as an anechoic recording of a person talking, and then to play the result over headphones to a listener. If done carefully (with suitable equalization for the effects of the microphones and the headphones, and assuming the listener is the same as the person used for measuring the HRTFs), then the listener will perceive the talker to be at the distance and direction of the HRTF recording. This *virtual auditory space* percept will be externalized: the sound will appear to be outside the listener. The effect stands in contrast to the “within-the-head”, lateralized percept that is otherwise typical with headphone-presented sounds. The illusion can be compelling: Hartmann and Wittenberg (1996) showed that the virtual source could be indistinguishable from a real one. Unfortunately, as noted above in relation to ILDs, the details of the HRTFs differ considerably from one listener to another because of individual differences in their head and outer ears (e.g., Wight-

Ex.1025, 5

Petitioner: Claim 8 Requires ITD not HRTF

8. A non-transitory computer readable storage medium storing instructions that cause a computer system to execute a method, comprising:

commence an electronic call between a first individual wearing an electronic earphone and a second individual that are remote from each other;

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

execute the electronic call such that the origin of the voice of the second individual appears to the first individual to originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first individual hears in response to the first individual moving the head orientation with respect to the sound localization point.

listener.]” Notably, claim 8 separately requires something other than HRTFs and 6DOF tracking, reciting in limitation 8[d] “alter[ing] an interaural time difference [ITD]” when the head orientation of the first listener moves. *Id.*, 43:57-61. As Dr. Begault confirmed, ITD is not an HRTF. EX1051, 23:6-7, 23:20-23. Thus, the

Reply, 4

Petitioner Ignores: ITD CANNOT Create Externalization

8. A non-transitory computer readable storage medium storing instructions that cause a computer system to execute a method, comprising:

commence an electronic call between a first individual wearing an electronic earphone and a second individual that are remote from each other;

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

execute the electronic call such that the origin of the voice of the second individual appears to the first individual to originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first individual hears in response to the first individual moving the head orientation with respect to the sound localization point.

HRTFs), then the listener will perceive the talker to be at the distance and direction of the HRTF recording. This *virtual auditory space* percept will be externalized: the sound will appear to be outside the listener. The effect stands in contrast to the “within-the-head”, lateralized percept that is otherwise typical with headphone-presented sounds. The illusion can be compel-

Ex.1025, 5

source by our hearing. A very strong localization cue is the arrival-time difference between left and right ear signals, which positions the auditory event laterally” (p. 240). Note that these characteristics must be imposed

Ex.2011, ¶35



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

POR 18-20 (citing Ex.2011, ¶35); Sur-
Reply 6-7 (citing Ex.1025, 5)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Petitioner's Expert: Claim 8 Sound is Synthesized

Limitation 8[B]

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

Q It is external, right?

MS. WANG: Asked and answered.

Q It is external, correct?

MS. WANG: Asked and answered.

A According to Eight KHz, that is the definition I used in my -- in my report.

Q And the sound originates externally from the listener, correct?

A The sound appears to the listener. There is no actual sound, right? It is in that room. The sound is synthesized. It appears to the listener to be away from and proximate.

Q Which is external to the user, correct?

MS. WANG: Asked and answered.

A Using that definition from Eight KHz, yes.

Ex.2016, 193:5-19

POR, 43-44 (citing Ex.2016, 193:5-19); POR, 45

Binaural Synthesis: “Simulate Head-Related Signals”

as a free field) using headphones. Ex. 1026, 154. 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.” Ex. 1015, 1.

Ex.1003, ¶46

55. The **second** is to **synthesize** binaural sound by **processing non-binaural content with the left and right HRTFs of the desired placement angle and position.**

Ex.1003, ¶55

“Origin of the Voice” in Claim 8 is Located Externally

Petitioner Expert: It Can Be Binaural

Limitation 8[B]

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

Q And we've also established that that is external to the user, correct?

A Based on the -- based on the Eight KHz definition, yes.

Q Could it also be binaural sound?

A Can it be binaural sound? That is -- that is one way to achieve that, but as I'm saying, there are other ways as well. That is why I'm not concluding that this is talking about binaural sound only.

Ex.2016, 194:20-195:19

“Origin of the Voice” in Claim 8 is Located Externally Petitioner Expert INITIALLY: It CANNOT Be Stereo

Limitation 8[B]

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

Q Can that be stereo sound?

A It could not be stereo – it could be other things. There are other – other processes that we could follow to externalize sound that don't involve HRT and don't involve binaural. It could not be stereophonic because at least in this claim, we are talking about headphones, not loudspeakers, and stereophonic requires loudspeakers.

Ex.2016, 195:20-196:17

PHOSITA: It CANNOT Be Stereo



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

15. Breebaart and Faller (2007, p. 9-10) (Ex. 2026) addressed the problem of externalization from *stereo audio playback* over headphones (ref. Figure 1, center). “Stereo audio signals are mostly produced in an optimized way for loudspeaker playback...one single stereo signal can be used for either loudspeaker or headphone playback. A major difference is that headphone listening with such stereo signals is limited to in-head localization, i.e., the width of the auditory spatial image is limited to being inside of the head of the listener...” Stereo audio is based on techniques for loudspeaker playback as opposed to headphones, and does not include the localization cues used in binaural synthesis (HRTFs, 6DOF tracking). Typically, only

Ex.2011, ¶15

POR, 42-44 (citing Ex.2011, ¶9- 22);
POR 46 n.13 (citing Ex.2026, 9-10)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

“Origin of the Voice” in Claim 8 is Located Externally Petitioner Expert CANNOT Name “Other Things”

Limitation 8[B]

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

Q Okay. So it is not stereophonic sound, but it could be binaural sound?

A It could be binaural or other things as well.

Q I'm sorry, did you say interaural?

A No, it could be binaural or it could be other things as well.

Q What other things?

A I don't know if they have a name. I know there are processes that you do -- you can process signals, so that when they arrive at your ears, externalize the sound without being binaural.

Ex.2016, 196:19-197:10

Petitioner Expert at Reply: It Can Be Stereo too

8. Outside of ILDs, ITDs, and HRTFs, there were also many other known ways to provide a sound localization point that is “away from and proximate to” the listener such that the sound localization point simulates an origin of the sound source. For example, Asahi (1979) discussed out-of-head externalization using stereophonic sound reproduction over headphones and without using HRTFs, but instead a dip filter. Ex. 1052. Similarly, Bauer (1961), which was also cited in Dr. Begault’s book (Ex. 1026) at page 69), discussed substituting earphones for loudspeakers without substantial loss of space perspective using filters and delays, not HRTFs. Ex. 1053.

Petitioner Expert is Inconsistent

Limitation 8[B]

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

Q Can that be stereo sound?

A It could not be stereo – it could be other things. There are other – other processes that we could follow to externalize sound that don't involve HRT and don't involve binaural. It could not be stereophonic because at least in this claim, we are talking about headphones, not loudspeakers, and stereophonic requires loudspeakers.

Ex.2016, 195:20-196:17

PHOSITA: Externalization Requires Binaural Synthesis and HRTF-Processing



3. Dr. Begault: A consistent expert

Petitioner mischaracterizes Dr. Begault's testimony. Reply, 7-9. First, Dr. Begault repeatedly stated that "[i]f the sound wasn't externalized, it couldn't be away from and proximate to," and "you cannot have sound externalization without HRTF processing," which "[o]ver headphones, that's absolutely true." Ex.1051, 67:24-70:10.

Sur-Reply, 10

PHOSITA: No Externalization with Stereo



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

Ex.2011, ¶45

sound from where the virtual sound source originates. In stereo audio using loudspeakers, sound images are restricted to an approximate placement on a horizontal axis within the “stereo field” between two loudspeakers or at the loudspeakers, for a listener sitting in the “sweet spot” directly in front of and between the loudspeakers. This is the result of volume “panning” as explained above in paragraph 15. This imagery is distorted once a listener moves away from the “sweet spot”. Such an SLP is incapable of simulating the origin of the sound such that the sound appears to the listener as it originates from the SLP. And over headphones, the stereo field is heard inside of the head as Breebaart and Faller (2007, p. 9-10) explain, and as I discussed in paragraphs 15-20 above.

POR, 11-17 (citing Ex.2011, ¶¶41-50);
Sur-Reply, 20-21 (Ex.2011, ¶45)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

PHOSITA: Externalization Requires Binaural Synthesis and HRTF-Processing



- b. The phrase “away from” and “proximate to” in limitation 8b. indicates a mode of binaural sound that would require externalization of the sound image. I have already established that the means of externalization for a listener requires both binaural sound with HRTFs and 6DOF tracking of the listener relative to the SLP that represents the talker.

Ex.1003, ¶79

Specification Makes Clear Claim 8's Voice is Binaural

Consider a first scenario in which John and Paul are talking as they sit twelve feet apart in an apartment room. During the conversation, John stands from his chair, walks to the refrigerator, opens a can of soda, and returns to his chair. Paul remains in his chair as John walks to the refrigerator. Paul hears John since both individuals are located in the room together.

In this second scenario, the notebook computer provides Paul's voice in stereo sound. John does not hear or perceive Paul's voice in the same manner as the first scenario since the notebook computer provides Paul's voice in stereo sound.

Specification Makes Clear Claim 8's Voice is Binaural

Consider a third scenario in which Paul travels to another country and initiates a call with John. John and Paul both wear electronic earpieces that record and provide binaural sound. When the call initiates, the voice of Paul is localized to his empty chair in John's apartment while John sits in his chair. John and Paul have the same conversation as in the first scenario, and John performs the same actions (i.e., during the conversation, John stands from his chair, walks to the refrigerator, opens a can of soda, and returns to his chair). In this third scenario, John hears and perceives Paul's voice in a same manner as in the first scenario. In the first and third scenarios, Paul's voice had a same sound localization point for John even as John moved around the room. In the first and third scenarios, the sound intensity, the interaural time differences, the interaural level differences, and output from the HRTFs changed in real time as John stood from his chair, walked to the refrigerator, opened the can of soda, and returned to his chair. From John's point-of-view, the first and third scenarios were audibly the same since the sound that John heard in the third scenario copied or emulated the sound that John heard in the first scenario.

Consider a fifth scenario that is similar to the third scenario. In the fifth scenario, however, the sound localization point of Paul changes or moves during the conversation. Initially, Paul's voice is localized to his chair. When John stands up and begins to walk toward the refrigerator, the sound localization point of Paul moves and follows John as if Paul were present in the apartment and moving with the sound localization point.

Bedingfield/Ballas

Petitioner Relies Solely on Bedingfield for 8[B]

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

Limitation 8[B]

3. 8[b]: Designate SLP

Bedingfield discloses this limitation. Kyriakakis ¶¶212-213.

As explained for 1[b], Bedingfield describes its *computer system* as *designat[ing] ...an SLP ...to the listener 102 (first individual)*. See Section VIII.C.3 (Ground 1, 1[b]); Kyriakakis ¶213. In Bedingfield, that *designat[ion]* is *for the first individual* because the sound is delivered to the listener 102 (*first individual*) through the headphones (*electronic earphone*). *Id.* The SLP simulates a location...from

Pet. 50-51

“Origin of the Voice” in Claim 8 is Located Externally

Petitioner Expert INITIALLY: It CANNOT Be Stereo

Limitation 8[B]

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

Q Can that be stereo sound?

A It could not be stereo – it could be other things. There are other – other processes that we could follow to externalize sound that don't involve HRT and don't involve binaural. It could not be stereophonic because at least in this claim, we are talking about headphones, not loudspeakers, and stereophonic requires loudspeakers.

Ex.2016, 195:20-196:17

Bedingfield is NOT a Binaural Reference



Ex.2017, 31

IPR2023-01004, Paper 14, 7-8

POR 11-12 (citing Ex.2017, 31);
Sur-Reply, 16 (citing IPR2023-
01004, Paper 14, 7-8)

Ex. 1003 ¶¶ 52–53 (citations and emphasis omitted). We agree with Patent Owner that Bedingfield does not teach either method. See Prelim. Resp. 45–47. In any case, 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.⁴

binaural sound; (2) “Bedingfield does not use the term ‘binaural sound’ and Petitioner [did] not point to any other language in the reference that could be equated with, or suggest, providing binaural sound”; and (3) Petitioner had not shown that Bedingfield teaches either of the two ways for creating binaural sound described by Dr. Kyriakakis, and did not “explain in any detail *how* Bedingfield allegedly teaches or suggests ‘binaural sound’

PHOSITA: Bedingfield is NOT about Binaural Synthesis



43. In summary: Bedingfield (1) never mentions the term HRTF or discusses HRTF spectra, which includes time differences as part of the spectral transfer function in terms of its phase components; (2) never mentions altering frequencies to correspond to the HRTF spectrum effected by a listener's ears, and in fact is at odds with the characteristics of HRTF spectra; (3) never mentions sound source externalization. In fact, Bedingfield only describes a basic stereo technique where sound sources are only separated by amplitude differences (“This stereo effect is used advantageously in nearly all modern music reproduction systems.”).

Ex.2011, ¶43

Petitioner Focuses on “Perception” of Externalization

Second, PO contradicts itself, admitting that Bedingfield (and specifically Figures 19a-c) does “allow[] the listener to perceive the stereo sound as externalized” but alleges this is based on the “imagination of the listener.” POR 57. PO’s sole support for this argument is Dr. Begault’s say-so. Dr. Begault cites no objective evidence supporting his assertion that Bedingfield’s adjustments to volume, echo, delay, balance, attenuation, and other aspects of sound “externalize a sound based solely on *imagination* as opposed to a manipulation of auditory cues.” EX2011, ¶17.

Reply, 20

PHOSITA: Cognitive Cues Can Create Perception of Externalization even with Stereo



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

17. Cognitive cues sometimes allow a listener to use higher level perception (based on memory and familiarity) to externalize a sound based solely on *imagination* as opposed to a manipulation of auditory cues; for example, a baseball game heard on an old-fashioned monaural AM radio with a single earpiece can be imagined to be outside of the head, *independent* of a conflict with the normal spatial auditory cues a person who was actually present at the baseball game would have. However, the concern of virtual

Ex.2011, ¶17

PHOSITA: Cognitive Cues Can Create Perception of Externalization even with Stereo



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

Ex.2011, ¶46

46. As with monaural sound, a stereo-based audio system like Bedingfield's can only create the perception of an externalized sound outside the head for a headphone listener by the listener's association of the sound with memory and association; that is, higher level "cognitive cues" (ref. par. 17). Despite this possible circumstance, stereo audio cannot provide the localization cues of an SLP designated by the listener, including externalized sound. Dr. Kyriakakis specifically addresses that many systems similar to Bedingfield use "visual representations" to guide the perception of externalization. As I will discuss below, Claims of the '090 Patent does not

POR, 11-17 (citing Ex.2011, ¶¶41-50); Sur-Reply, 18 (citing Ex.2011, ¶46)

Dr. Begault Points to a Scientific Fact Known Also by Petitioner Expert

claims. *Id.*, 199:16-200:5. Dr. Kyriakakis also agrees with Petitioner's other IPR expert that "if you show me a picture of a dog and I hear barking, I can perceive that sound as coming from that location whether or not the sound accurately represents that." Ex.2031, 51:11-52:3. He admits that visual-cue-based systems can "fool some people" to "attribute the sound from your speakers coming from the [dog's picture]," but "the sound does not actually originate with the dog picture." *Id.*, 43:21-44:9.

Sur-Reply, 18

Claim 8 Requires that the Sound Originates from the SLP

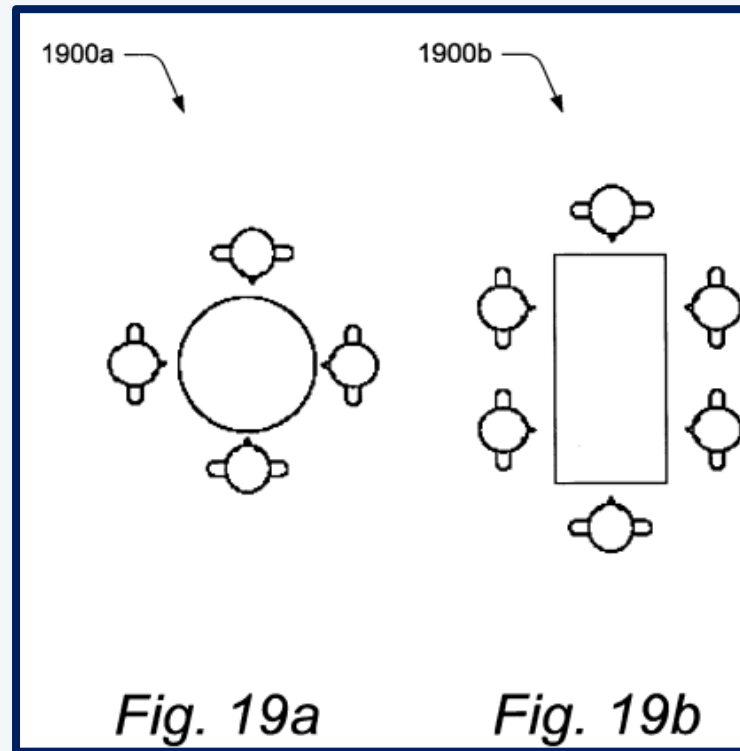
Limitation 8[B]

designate, for the first individual, a sound localization point that simulates a location in empty space from where an origin of a voice of the second individual occurs for the first individual through the electronic earphone such that the sound localization point in empty space is located away from and proximate to the first individual;

Limitation 8[C]

execute the electronic call such that the origin of the voice of the second individual appears to the first individual to originate from a fixed location that is the sound localization point in empty space while the first individual moves and changes a location and a head orientation with respect to the sound localization point; and

Bedingfield Uses Visual Cues to Create a Perception of Externalization



FIGS. 19a-19c include schematics of sample default configurations for a listener and participants within a sound field in accordance with the disclosure. FIG. 19a includes a circular sound field 1900a with participants arranged in a circular configuration. FIG. 19b includes a polygon sound field 1900b with participants arranged around the perimeter of the polygon. FIG. 19c includes a presentation sound field 1900c with participants arranged in an audience configuration orientated towards a speaker. FIGS. 19a-19c are demonstrative of exemplary default configurations of methods and systems of synthetic audio placement. One should appreciate that other default configurations may be included and still remain within the spirit and scope of the disclosure.

Ex.1005, 16:36-48

POR, 56-57 (citing Ex.1005, Figs. 19a-c); Pet. 37 (citing Ex.1005, 16:36-48);
Sur-Reply 18

Petitioner : Bedingfield Deals with Listener's Perception

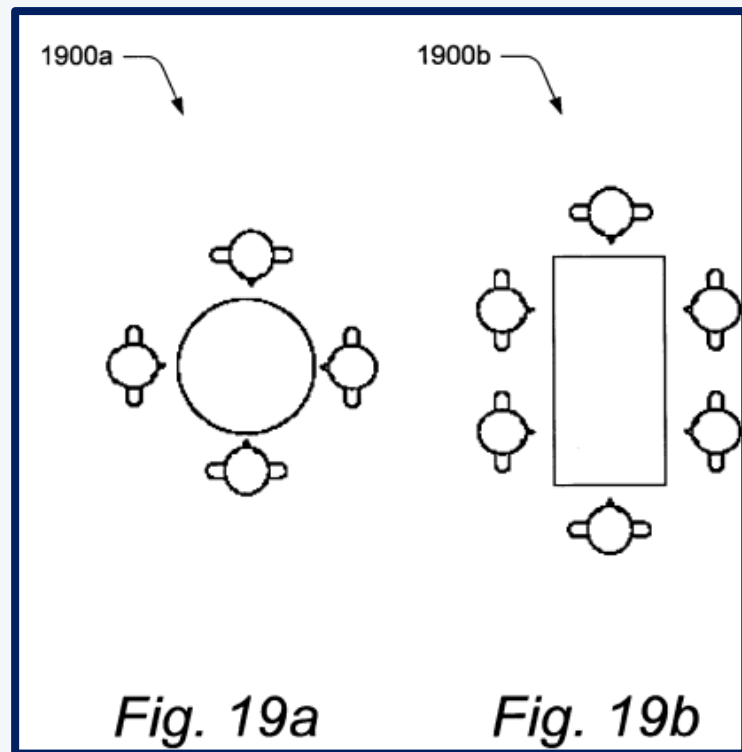
EX1005, 14:30-43 (cited by EX2011, ¶82). A POSITA reading Bedingfield would have understood that **perceiving the speaker behind the listener is externalized** and not inside the head. Kyriakakis Reply, ¶¶48-49. Indeed, to make his argument, Dr.

Reply, 19

EX1005, 4:66-5:3. As Bedingfield further describes, the *computer system* simulates the sound field 122 so that **the listener's perception of the call is as if the listener and participants "were in the same location and speaking to each other directly."** *See*

Pet. 22

Petitioner Expert Describes Systems Using Visual Cues for Creating Perception of Externalization



Ex.1005, Figs. 19a-c

POR, 56-57 (citing Ex.1005, Figs. 19a-c); Sur-Reply 18-19 (Ex.1003, ¶133)

133. Similar to Vertesi, Torrejon involves a hybrid conference environment with participants (students in this case) communicating in person and remotely. Ex. 1038, 1, 5. Torrejon introduces an added element: augmented reality. *Id.*, 1. “Augmented Reality . . . allowed students to interact with objects fellow students and the lecturer over great distances.” *Id.* Those participating by phone or other remote communication means select a seat that is not occupied by a physically present student (i.e. an empty chair). *Id.*, 5, 6. The phone participant’s image is then projected in the location of the selected seat. *Id.*, 6, 7. During the call, voices of remote participants are produced such that they appear to originate from the projected image, which is located in the physically empty chair. *Id.*, 6-8.

Claim 8's Voice *Originates* from the SLP—NOT Perceived at the SLP Like in Bedingfield



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

listener. Ex. 1003 ¶¶127-136. However, Dr. Kyriakakis fails to recognize that an SLP in the '090 Claims must simulate the origin of the binaural sound from where the virtual sound source originates. In stereo audio using

Ex.2011, ¶45

an empty chair.” Ex.1003, ¶133. This highlights a key point: the voices are perceived to originate there, *not actually originating* there. Yet, the claims require that the sound *originates* from the SLP.

Sur-Reply, 18-19

Petitioner District Court Expert: Sound Perception is a Different Field

Q There's nothing in just a standard accepted practice and distance in audiology that, you know, you observe when people typically feel the need to raise their voice?

A I would say that I'm not aware of any literature because, again, I -- my work is not in voice and voice perception. That's actually a different field. We deal with the creation, you know, of audio.

So if there is an audio perception piece of literature, I'm not aware of it because it's not something that we standardly use in this area of work to figure out -- because this is an audio perception question that you're asking, not an audio signal processing question.

In addition, remember, sound perception is a subjective experience. Just because I rendered a sound using all of these fancy techniques that we've talked about today, that doesn't mean that someone actually hears the sound as coming from outside of their head. Like, there are a good section of

Ex.2015, 190:20-25

Claims are NOT Directed at Perception Despite Petitioner Expert's Attempt to Show Otherwise

factors into consideration, a person of ordinary skill in the art at the time of the claimed priority date of the '090 patent would have had at least a bachelor's degree in computer science or engineering with a focus on audio signal processing, with at least two years of relevant experience. It would have taken approximately at least

Ex.1003, ¶34

Claims are NOT Directed at Perception Despite Petitioner Expert's Attempt to Show Otherwise

Q Give me your overall field and then we might break it down.

A The way I describe my field is, my expertise is at the intersection of acoustics, auto signal processing, and psychoacoustics. And so algorithms that are informed by what happens to sort of the -- that's the -- algorithms is the signal processing part, and that are informed by what sound -- what happens to sound in various environments, that's the acoustics part, and how it is perceived by listeners, and that's the psychoacoustics part.

Ex.2031, 53:2-13

Q So does the '090 patent, the one that is at issue and we're talking about today, does that concern psychoacoustics?

A One aspect of it, the -- the ability to detect -- or to perceive sound from certain directions. There are many aspects of psychoacoustics, but that's one of them.

Q So is it your opinion that to understand this patent, the '090 patent, that a person having ordinary skill in the art should be knowledgeable about psychoacoustics?

A I don't know if they have to be knowledgeable about every single aspect of psychoacoustics, but the part that has to do with sound localization is relevant to this patent.

Definition of PHOSITA Has NO Reference to Psychoacoustics

factors into consideration, a person of ordinary skill in the art at the time of the claimed priority date of the '090 patent would have had at least a bachelor's degree in computer science or engineering with a focus on audio signal processing, with at least two years of relevant experience. It would have taken approximately at least

Ex.1003, ¶34

Petitioner Expert Conflates “Providing” with “Perceiving”

Q Where in that claim is it describing perception of sound? And I'm sorry, excuse me. I'm talking about the same one, in the '090 patent. Excuse me.

ATTORNEY TENNANT: Objection. Form.

A One part is in the last limitation, where it says – well, I'll just read it: **“Providing with an electronic earphone located at the head of the listening person the binaural sound to the listening person during the telephone call, such that the origin of the binaural sound for the listening person occurs at the sound localization point in empty space that is away from and proximate to the listening person.”**

So that implies that the person is hearing it in empty space at a certain location, and away from and proximate, **which is perception.**

(Reporter clarification.)

A At a certain location and empty space that is away from and proximate.

Q That limitation itself is talking about providing the sound, correct?

ATTORNEY TENNANT: Objection. Mischaracterizes testimony.

Q And I apologize. When I say "that limitation," I'm talking about the limitation that starts with, "Providing with an electronic earphone located at the head of the listening person," that last limitation of Claim 1. That is talking about providing a sound, correct?

A **It's talking about both, providing and perception by the person receiving it.**

Ex.2031, 55:7-56:22

Petitioner: Bedingfield Teaches Altering ITD

As explained for 2[b] and claims 17 and 18, Bedingfield discloses *alter[ing] an ITD of the voice that the first individual hears*. See Sections VIII.D.2 and VIII.F.2 (Ground 1, 2[b] and 17) and Section IX.C (Ground 2, 18); Kyriakakis ¶218. A POSA would have appreciated it was commonplace to alter ITDs to maintain the origin of a sound while the user moves head orientation. Kyriakakis ¶218.

Pet. 52

2. Claim 17: Change ITD, ILD, HRTF

Bedingfield renders obvious this limitation. Kyriakakis ¶¶170-174; see Sections VIII.E.5-7 (Ground 1, 15[d]-15[d2]).

Bedingfield discloses *ITD* is a cue that provides directionality and places a sound source in the sound field. See Section VIII.D.1 (Ground 1, 2). Another such

Pet. 38

Petitioner: Bedingfield Teaches Altering ITD

b) 2[b]: Adjust ITD

Bedingfield discloses this limitation. Kyriakakis ¶¶145-147.

...

Bedingfield also discloses *adjusting an ITD of the sound that the listening person hears* in order to orient the listener within the sound field 112. Kyriakakis ¶147. Bedingfield uses “various audio techniques” such as “delay” to place the different participants in different “configurations within the sound field.” EX1005, 5:35-41. Bedingfield further describes applying “changes to delay” to enhance the “impression that a sound source has moved” relative to the listener. *Id.*, 3:56-59. The “delay” in this context is an *ITD*, commonly used to manipulate the perceived origin of a sound source in an audio scene. Kyriakakis ¶147; EX1027, 39-40. Thus, Bedingfield discloses *adjusting an ITD of the sound* delivered to the listener.

Pet. 29

Petitioner's Evidence: Its Expert Conclusory Opinion

b) 2[b]: Adjust ITD

Bedingfield discloses this limitation. Kyriakakis ¶¶145-147.

...

Bedingfield also discloses *adjusting an ITD of the sound that the listening person hears* in order to orient the listener within the sound field 112. Kyriakakis

¶147. Bedingfield uses “various audio techniques” such as “delay” to place the different participants in different “configurations within the sound field.” EX1005, 5:35-41. Bedingfield further describes applying “changes to delay” to enhance the “impression that a sound source has moved” relative to the listener. *Id.*, 3:56-59.

The “delay” in this context is an *ITD*, commonly used to manipulate the perceived origin of a sound source in an audio scene. Kyriakakis ¶147; EX1027, 39-40. Thus,

Bedingfield discloses *adjusting an ITD of the sound delivered to the listener*.

Pet. 29

Petitioner's Evidence: Its Expert Conclusory Opinion

147. Bedingfield also discloses *adjusting an ITD of the sound that the listening person hears* in order to orient the listener within the sound field 112. Bedingfield uses “various audio techniques” such as “delay” to place the different participants in different “configurations within the sound field.” Ex. 1005, 5:35-41. Bedingfield further describes applying “changes to delay” to enhance the “impression that a sound source has moved” relative to the listener. *Id.*, 3:58-59. A fixed SLP and a moving listener is functionally equivalent to a stationary listener and a moving SLP. Furthermore, a POSA would have readily understood that the “delay” in this context is an *ITD*, which is the delay in arrival time of a sound between two ears and is commonly used to manipulate the perceived origin of a sound source in an audio scene. Thus, Bedingfield discloses *adjusting an ITD of the sound* delivered to the listener.

Ex.1003, ¶147

“Delay” Means a Lot of Things in the Field

Q Changing topics, do you agree that the term "delay," by and of itself, could encompass a broad range of concepts in audio engineering and audio signal processing?

ATTORNEY TENNANT: Objection. Lack of foundation.

A I agree with that, yes.

Ex.2031, 79:17-80:1

Tellingly, Dr. Kyriakakis never engaged with the context of Bedingfield's disclosure to explain how the term “delay” necessarily means ITD. Dr. Kyriakakis

Sur-Reply, 22

Petitioner Relies on Conclusory Statements

POR, 5 Petitioner relies solely on conclusory statements from Dr. Kyriakakis. A “petitioner cannot employ mere conclusory statements. The petitioner must instead articulate specific reasoning, based on evidence of record, to support the legal conclusion of obviousness.” *In re Magnum Oil Tools Int’l, Ltd.*, 829 F.3d 1364, 1380 (Fed.Cir.2016). Further, as discussed below, his credibility is compromised by

Pet. 29 The “delay” in this context is an *ITD*, commonly used to manipulate the perceived origin of a sound source in an audio scene. Kyriakakis ¶147; EX1027, 39-40. Thus, Bedingfield discloses *adjusting an ITD of the sound* delivered to the listener.

Petitioner's Expert Opinion: "Conclusory & Unsupported"



Ex.2017, 24

Finally, we agree with Patent Owner that Dr. Kyriakakis's opinion that he "read[s] Li as disclosing that moving the sound source . . . takes place during a voice exchange between the computer program and a user/person" is conclusory and unsupported. *Id.* at 10–13 (quoting Ex. 1003 ¶ 104) (emphasis omitted). Again, the only support for Dr. Kyriakakis's opinion is the very brief "conference call" excerpt in Li, which we find insufficient for the reasons explained above. *See* Ex. 1003 ¶ 104 (citing

Ex.2017, 30

(as Petitioner argues in this proceeding). Petitioner and Dr. Kyriakakis provide no explanation why that would be the case, other than to state that listener 102's headphones are at both ears. *See* Pet. 58, 62; Ex. 1003 ¶¶ 206,

PHOSITA: Bedingfield's Delay is NOT ITD



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

Contrastingly, Dr. Begault meticulously analyzed the three mentions of “delay” in Bedingfield across four pages, determining that Bedingfield’s delay is not ITD. Ex.2011, ¶49. The first instance is linked to “sound effects” that help the

Bedingfield's First Delay

source. 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. In a virtual environment, this effect, combined with a decrease in volume and left-right positioning, may give the listener the impression that a sound source has moved "behind" the listener, even with a two channel stereo system. Additionally, changes to delay and echo may heighten this effect. Further, other parameter changes may provide the sensation that one participant is located behind another participant.

Ex.1005, 3:49-61

Petitioner's Expert Concedes Bedingfield's First Delay Can Be Things Other Than ITD

source. 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. In a virtual environment, this effect, combined with a decrease in volume and left-right positioning, may give the listener the impression that a sound source has moved "behind" the listener, even with a two channel stereo system. Additionally, changes to delay and echo may heighten this effect. Further, other parameter changes may provide the sensation that one participant is located behind another participant.

Ex.1005, 3:49-61

Q Or the sound's being perceived by the listener as in front. Is it -- can you move the sound behind the listener by changing the delay time?

MS. WANG: Objection, form.

Q For example, by changing the delay time of echo to what would be perceived as a delay for echo?

...

So theoretically in this hypothetical, if you change the response with echos to correspond to the frequency response of a sound source behind the listener, you would be able to mimic that.

Ex.2016, 229:14-230:16

PHOSITA: Bedingfield's First Delay CANNOT be ITD



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

listener.”). Ex. 1003 ¶147. Dr. Kyriakakis has confused the term “delay”, which can mean to a POSITA any number of repeated copies of a sound such as used in musical instrument processors (digital delays), with the microsecond interaural time differences that are captured only in a calibrated HRTF measurement in a laboratory.

Ex.2011, ¶47

source by our hearing. A very strong localization cue is the arrival-time difference between left and right ear signals, which positions the auditory event laterally” (p. 240). Note that these characteristics must be imposed

Ex.2011, ¶35

source. 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. In a virtual environment, this effect, combined with a decrease in volume and left-right positioning, may give the listener the impression that a sound source has moved “behind” the listener, even with a two channel stereo system. Additionally, changes to delay and echo may heighten this effect. Further, other parameter changes may provide the sensation that one participant is located behind another participant.

PHOSITA: Bedingfield's First Delay CANNOT be ITD



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

source. 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. In a virtual environment, this effect, combined with a decrease in volume and left-right positioning, may give the listener the impression that a sound source has moved "behind" the listener, even with a two channel stereo system. Additionally, changes to delay and echo may heighten this effect. Further, other parameter changes may provide the sensation that one participant is located behind another participant.

than a signal of higher volume." Then, he discusses a third sound effect: attenuation of higher frequencies to make a sound appear "behind" a listener. He concludes that echo and delay heighten this third effect "with a two-channel stereo system." However, a POSITA would recognize that ITD could not be useful for discrimination or disambiguating "front" from "rear" positions. In fact, the ITD would be the same for a front and rear position, such as shown in Figure 10 for positions *a* and *b*, thereby resulting in the well-known "cone of confusion." With reference

Bedingfield's Second Delay

within the sound field. For example, the participants may be placed in a circular configuration within the sound field. In other embodiments, the participants may be placed in other configurations within the sound field such as in a rectangular configuration or along a linear axis, using various audio techniques such as volume, balance, parametric equalization, delay, echo and the like to produce the effects described above with reference to FIG. 1. In yet other embodiments, the participants may be placed in a position relative to their known geographic location.

Ex.1005, 5:35-41

Bedingfield's Second Delay

Ex.1005, 5:35-41

within the sound field. For example, the participants may be placed in a circular configuration within the sound field. In other embodiments, the participants may be placed in other configurations within the sound field such as in a rectangular configuration or along a linear axis, **using various audio techniques such as volume, balance, parametric equalization, delay, echo and the like to produce the effects described above with reference to FIG. 1.** In yet other embodiments, the participants may be placed in a position relative to their known geographic location.

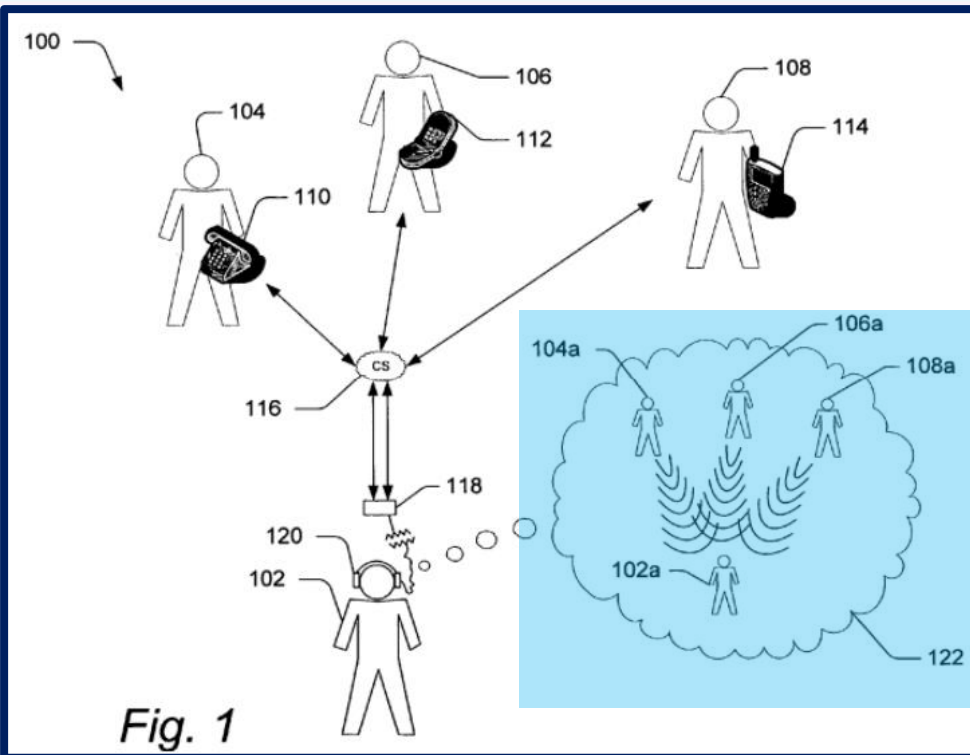


Fig. 1

The sound field 122 shown in FIG. 1 represents the perception resulting from the audio output of the multi-channel processing device 118 through the multi-channel speaker system 120 as heard by the listener 102. For example, if a listener 102a and participants 104a, 106a, 108a were in the same location and speaking to each other directly (without the communication system 116), then the listener 102a would hear sounds from the participants 104a, 106a, 108a through both of the listener's 102a ears, and thus have a specific balance, volume, and timbre for each speaker. In the environment 100, **the user 102 receives audio from the participants 104, 106, 108 through a communication system 116 in a substantially similar (or approximately similar) manner as if the listener was present between all participants 104, 106, 108.** The multi-channel processing device 118 configured

Ex.1005, 4:52-5:3

Bedingfield's Second Delay is NOT ITD

participants.” Bedingfield 4:52-5:3. Thus, “delay” here merely means a technique for simulating a room with participants for a listener, thereby creating a “perception” that the listener is among the participants, whereas ITD is typically used to adjust binaural sound as a function of listener’s head orientation. Ex.2011, ¶49; *see also*

POR 19-20

maintain the origin of the voice of the second individual at the sound localization point but alter an interaural time difference of the voice that the first individual hears in response to the first individual moving the head orientation with respect to the sound localization point.

Limitation 8[D]

Bedingfield's Second Delay is NOT ITD

Ex.1001, 9:24-29. But Bedingfield never discusses localization for the second instance of delay. Bedingfield, 4:52-5:41. Dr. Begault underscores that ITD, an important “localization cue,” measures “the arrival-time difference between left and right ear signals, positioning the auditory event laterally.” Ex.2011, ¶¶35-36. Thus, since Bedingfield’s delay is not used for localization, it cannot be ITD.

Sur-Reply, 24

Beddingfield's Third Delay is NOT ITD

pants. In some embodiments, the change volume at the block 316 may also include equalization controls (e.g., bass and treble, and more complex characteristics controls), digital delay controls, echo controls, and other audio controls to adjust the sound output of all of the participants.

Ex.1005, 6:15-17



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

These types of processors are commonly used by musicians and recording engineers to create special dramatic effects or to simulate room reverberation or outdoor environments (e.g., an echo from a canyon), including in multi-channel loudspeaker playback. A POSITA would not imagine such devices as being capable of simulation of ITDs using microsecond interaural time differences that are captured only in a calibrated HRTF measurement in a laboratory.

Ex.2011, ¶49

Petitioner's Reference to Commonplace

As explained for 2[b] and claims 17 and 18, Bedingfield discloses *alter[ing] an ITD of the voice that the first individual hears. See Sections VIII.D.2 and VIII.F.2 (Ground 1, 2[b] and 17) and Section IX.C (Ground 2, 18); Kyriakakis ¶218. A POSA would have appreciated it was commonplace to alter ITDs to maintain the origin of a sound while the user moves head orientation. Kyriakakis ¶218.*

Pet. 52

Petitioner's Reference to Commonplace

218. Furthermore, as I already explained above for 2[b] in Ground 1, claim 17 in Ground 1, and claim 18 in Ground 2, Bedingfield discloses *alter[ing] an ITD of the voice that the first individual hears*. See Sections IX.D.2, IX.F.1, X.A-C. As the POSA would have appreciated, and based on my discussion in the Background (Section VI), it was commonplace to alter ITDs to maintain the origin of a sound while the user moves head orientation.

Ex.1003, ¶218

Problem with Petitioner's Reference to Common Sense

POR, 5

First, it pivots on the conclusory and unsupported declaration of Dr. Kyriakakis offered “to supply” key limitations “missing from the prior art.” *Xerox Corp. v. Bytemark, Inc.*, IPR2022-00624, Paper 9, 5 (PTAB Aug. 24, 2022) (precedential) (citing *Arendi S.A.R.L. v. Apple, Inc.*, 832 F.3d 1355, 1361-62 (Fed.Cir.2016)).



Third, our cases repeatedly warn that references to “common sense”—whether to supply a motivation to combine or a missing limitation—cannot be used as a wholesale substitute for reasoned analysis and evidentiary support, especially when dealing with a limitation missing from the prior art references specified. Indeed,

Grounds Based on Li

PHOSITA: Localization During the Call is a Novel Aspect of '090 Claims



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

BY MS. NGUYEN:

Q. Do you believe the innovative concept of the '090 patent is to be able to place a sound localization point in an empty space during a telephone call?

MS. PARTRIDGE: Objection. Form.

THE WITNESS: I think that's one of the novel aspects of Claim 1.

Ex.1051, 53:18-24

What is a “Call”?

⁸ The Petition applies PO’s interpretation that *telephone call* is “voice communication” over an electronic network between at least two parties.” See EX1027, 4.

Pet. 19 n.8

Q And what I asked you was whether or not this included -- would be live voice exchange.

A So **there is live voice exchange during a telephone call**, but I was trying to say before we ended that other stuff like background noise and things could also go through there. **So it's not the only thing is what I'm trying to say.**

Ex.2016, 176:21-177:6

Petitioner Exclusively Relies on Li for the Call Terms

PO argues that “Li does not teach an electronic/telephone call involving a live voice exchange.” POR 61; *see also* POR 21-30. But Li discloses: “[T]he configurable 3D sound system 900 provides a method of communication among multiple people [located remotely from one another], for example, in a conference call, by placing different speaking users at different spots to mimic a real conference room environment.” EX1006, 21:60-22:5; *see also* Pet. 58-59.

Reply, 26

The Board Rejected This Argument Twice



Opp. 9–10. Again, as explained in the Decision, Li mentions a “conference call” only in one very brief portion, while describing in detail the recording of sound tracks that can be configured and played back, and there was insufficient explanation in the Petition of “how a voice *exchange* would occur during such a ‘conference call’ when modified to use the processed voice of a speaker.” *See* Dec. 21–22. We are not persuaded that we misapprehended or overlooked any argument in the Petition in reaching that conclusion.

IPR2023-01004, Paper 14, 6

Petitioner's Attempt to Distinguish Between Recorded and Pre-Recorded in Li Fails

Reply, 27

PO further counters that “Li consistently uses ‘recorded’ or ‘recordings’ to describe recorded sound playing back at a later time—not in real-time.” POR 26.

But as both the '090 Patent and Dr. Begault confirm, the plain meaning of “record” covers both live and non-live sounds.⁷ The '090 Patent repeatedly uses the term

Further, even if Petitioner's new argument was proper, we disagree that Li always distinguishes between the terms “recording”/“recorded” as referring to “sound captured and transmitted in real-time” and “pre-recording”/“pre-recorded” as “sound that has been stored and later played back.” See Req. Reh'g 1, 4. For example, Li discloses that “sound



Petitioner Expert's Arbitrary Distinction

Q But the whole point of Li is talking about recorded and prerecorded voice, correct?

A Right. Recorded as we said several times in other contexts here has to do with capturing, to me that means live. And prerecorded means it was recorded in advance and it is being played back.

Ex.2016, 282:16-22

opinion of Li's conference call. Dr. Kyriakakis drew an arbitrary distinction between

Li's use of recorded and prerecorded, stating that "recorded ... to me [] means live."

Ex.2016, 282:9-22. Indeed, Dr. Kyriakakis's statement here echoes the similar

"conclusory" statement in the 1004 IPR that "he reads Li as disclosing...", which

the Board rejected as "unsupported." Ex.2017 24 (citing Ex.2018 ¶104).

POR 24

Petitioner Expert's Inconsistent Opinions

beam pattern directions. The sound tracks recorded by the microphone array system 902 are stored in a memory or a storage device (not shown). The 3D sound processing appli-

Ex.1006, 17:20-22

Q So does this mean that the sound is live when it is stored in memory?

A When it is stored, no, it is not live.

Ex.2016, 281:21-284:1

Petitioner Expert's Inconsistent Opinions

phones, tablet computing devices, etc. The microphone array system 902 can be configured in a computing device 901 with a universal serial bus (USB) interface for applications in 3D sound recording. The multiple channel sound can be saved in one file in a portable device. Using the 3D sound processing application 1602, users can play the recorded audio as a 3D binaural sound or a 3D surround sound. The 3D sound pro-

Ex.1006, 34:56-63

Q Does this mean that the recorded sound is live when users can play the recorded audio after it was saved?

A This particular sentence says after it was saved. Of course it is not live, but I think we're in the context of this patent he has both, live and – so recorded and prerecorded.

Ex.2016, 284:15-21

POR 24-25 (citing Ex.1006, 34:56-63);

POR 24-25 (citing Ex.2016, 284:2-21)

DEMONSTRATIVE EXHIBIT – NOT EVIDENCE

Petitioner Expert's Inconsistent Opinions

tioned several feet apart and separated by empty space. When a stereo recording is played back, the recording from one microphone goes into the subject's left ear, while the recording from the other microphone is channeled into the subject's right ear. This gives a sense of the position of the sound as

Ex.1006, 1:54-58

Q I'm asking you about these specific paragraphs. And so when -- or those specific provisions here where it is Column 1, lines 54 through 56, when it is being played back is that live?

A That is what he -- in the abstract and elsewhere calls prerecorded, so it is not live two-way communication.

Ex.2016, 286:12-19

Petitioner Expert's Shifting Opinions

Q But the whole point of Li is talking about recorded and prerecorded voice, correct?

A Right. Recorded as we said several times in other contexts here has to do with capturing, to me that means live. And prerecorded means it was recorded in advance and it is being played back.

Ex.2016, 282:16-22

Q And how do you -- how do you -- how would you form your opinion that the recorded sound can be live?

A Well, put it a different way. I would form my opinion that it is not live if as in a couple of the examples you showed me, it says it was stored on a memory device or something of the type and played back from there. But if it is just, like, if I had a binaural microphone sitting here next to me and it was recording, it would be live sending it to you.

Ex.2016, 289:15-22

Petitioner Expert's Continued Inconsistency

Q And how do you -- how do you -- how would you form your opinion that the recorded sound can be live?

A Well, put it a different way. I would form my opinion that it is not live if as in a couple of the examples you showed me, it says it was stored on a memory device or something of the type and played back from there. But if it is just, like, if I had a binaural microphone sitting here next to me and it was recording, it would be live sending it to you.

Ex.2016, 289:15-22

ing devices to allow users to record and play 3D sound. The configurable 3D sound system disclosed herein allows individual users to record and reproduce 3D sound for playback on their headsets and home theater speaker systems, thereby allowing users to experience immersive 3D sound.

Ex.1006, 7:33-37

Q Okay. Does this mean, again, that the recorded sound is live?

A It doesn't say. It can be.

Ex.2016, 289:12-14

POR 25 (citing Ex.2016, 289:15-22);
POR 25-26 (citing Ex.1006, 7:33-37);
POR 25-26 (citing Ex.2016, 289:3-14)

PHOSITA: Li Has Nothing to Do with Live Voice Exchange



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

93.I disagree. The reason that Li distinguishes between recorded and pre-recorded is that Li discloses a method for recording sound, as the following examples show. Thus, when Li refers to “recorded”, it refers to sounds that were recorded and then stored by Li’s system for later playback, as opposed to “pre-recorded” which refers to sound recorded by a system other than Li’s.

94.The following section from Li describes “recorded” sounds that are consistently described as having been recorded using Li’s microphone system:

Ex.2011, ¶¶93-94

PHOSITA: Li Has Nothing to Do with Live Voice Exchange



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

100. A POSITA reading Li would understand that Li's conference call example must be read in the broader context of Li's disclosure about recorded and pre-recorded sounds. When Li gives numerous examples in which recorded and pre-recorded sounds are not live, the conference call example in Li cannot mean to a POSITA to involve live voice exchange with the users.

Ex.2011, ¶100

PHOSITA: Li Has Nothing to Do with Live Voice Exchange



DURAND RENE BEGAULT, Ph.D.
Advanced Controls and Displays Group
Human Factors Research and Technology Division
NASA Ames Research Center

101. A POSITA would understand that in Li's conference call, the listener simply listens to the recorded voices of speakers that are being played back after being processed based on spatial configuration selected by the listener. For example, a listener may wish to review a recording of a teleconference, an interactive class lecture with multiple participants, etc.. However, Li never addresses real-time telephonic conversation between two parties that involves real-time live voice exchange.

Ex.2011, ¶101

Petitioner Borrows Context from Dr. Begault's Testimony of the '090 Patent

PO further relies on IPR2023-1004 to allege Petitioner is precluded from arguing that Li discloses a live exchange. But the Board's finding in IPR2023-1004 pertained to a different limitation. *See* EX2017, 17-24; EX2021, 18-32. Moreover, the Board did not have Dr. Begault's admission that "record" can include live sound.

Reply, 28

added). Indeed, as Dr. Begault confirmed: "In the context of [the '090 Patent], ['record'] clearly is talking about a real-time interaction and not talking about 'record' in a sense that sometimes the word 'record' is used, to talk about pre-stored sound." EX1051, 55:6-19; *see also* EX2016, 282:9-22.

Reply, 27-28

The 1004 IPR Decision

arguing that Li discloses a live exchange. But the Board's finding in IPR2023-1004 pertained to a different limitation. See EX2017, 17-24; EX2021, 18-32. Moreover,

Reply, 28

⁵ In the 1004 IPR, Petitioner and Dr. Kyriakakis argued a computer program's voice is no more than the voice of a speaker processed by a computer program. Ex.2018 ¶104; Ex.2021 23-24; 57. Accordingly, Petitioner argued that the voice exchange between a user and a computer program is no more than a voice exchange between a speaker whose voice is processed by a computer program and another user. Ex.2021 23-24; 57. The Board assumed (without deciding) that Petitioner's

POR, 22 n.5

The 1004 IPR Decision (Continued)

arguing that Li discloses a live exchange. But the Board's finding in IPR2023-1004

pertained to a different limitation. See EX2017, 17-24; EX2021, 18-32. Moreover,

Reply, 28

argument was correct. Ex.2017 20. Thus, to the extent that Petitioner may argue that voice exchange in the '090 Patent's telephone/electronic call is different from the voice exchange at issue in the 1004 IPR, Petitioner is precluded from making argument, for it represented (and the Board assumed) that the voice exchange between a user and a computer program is the same as the voice exchange between a user and a speaker whose voice is processed by a computer program.

POR, 22 n.5

Petitioner's Li Grounds Are Deficient



limitation 1[b]. At this stage, we question whether Petitioner's approach of using its analysis of limitation 1[b] as a shorthand for ten other allegedly "similar" limitations that use different language and pertain to different steps, without any further explanation, is sufficient to meet its burden. *See* Pet. 63 n.16.

Decision, 31-32

Case IPR2023-01003

Patent No. 9,226,090

CERTIFICATE OF SERVICE

I hereby certify that on October 25, 2024, I caused a true and correct copy of the foregoing Patent Owner's Demonstratives to be served via electronic mail at PH-Meta-Eight-KHZ-IPR@paulhastings.com upon the following attorneys of record for the Petitioner:

Lisa K. Nguyen
Eric E. Lancaster
Paul Hastings LLP
1117 S. California Avenue
Palo Alto, CA 94303

David M. Tennant
Alan M. Billharz
Paul Hastings LLP
2050 M St., N.W.
Washington, DC 20036

Grace I. Wang
Paul Hastings LLP
200 Park Avenue
New York, NY 10166

Date: October 25, 2024

Respectfully submitted,

/Erick Robinson/

Erick Robinson, Esq. (Reg. No. 51,354)

BOCHNER PLLC

24 Greenway Plaza, Suite 1800

Houston, TX 77046

Tel: (646) 971-0685

Fax: (646) 343-9672

Attorneys for Eight KHZ LLC