

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM SERVICES LLC,
Petitioner,

v.

SMART SPEAKER LLC,
Patent Owner.

Patent No. 12,401,721

Inter Partes Review No. IPR2026-00276

REPLY DECLARATION OF HENRY H. HOUH

I, Henry H. Houh, make this declaration in connection with the proceeding identified above.

I. INTRODUCTION

1. I have been retained by counsel for Amazon.com Services LLC (“Amazon”) as a technical expert in connection with the proceeding identified above. I described my qualifications in detail in my opening declaration dated February 27, 2026, which I understand is exhibit EX1002 to this proceeding. A copy of my curriculum vitae was provided in EX1002 as Appendix A.

2. I submit this reply declaration (“Reply Declaration”) in support of Amazon’s Petition for *Inter Partes* Review of United States Patent No. 12,401,721 (“the ’721 patent,” EX1001). I have been asked to consider Smart Speaker LLC’s Patent Owner Response (Paper 13, “Response”) and the Declaration of Dr. Michael Shamos (EX2011) and provide rebuttal opinions regarding the same. In performing my analysis in this proceeding, I have considered the ’721 patent (EX1001), its prosecution history, materials cited and discussed in the Patent Owner Response and the Declaration of Dr. Michael Shamos. I also reviewed the transcript of Dr. Shamos’ September 30, 2026 deposition (EX1102), the references and materials cited in my opening declaration, and the materials cited in this Reply Declaration. In formulating my opinions, I have further relied upon my extensive experience in the relevant fields.

3. I am being compensated at my normal rate, plus reimbursement for expenses, for my analysis. My compensation does not depend on the content of my opinions or the outcome of this proceeding.

4. I have personal knowledge of the facts and opinions set forth in this declaration and, if called to testify, I could and would testify competently to them.

5. Although I am not a lawyer and will not provide any legal opinions, I have been advised to apply certain legal standards explained to me by counsel in forming my opinions. I set forth my understanding of those legal standards in my opening declaration. (*See* EX1002, ¶¶51-54.)

II. LEVEL OF ORDINARY SKILL AND RELEVANT BACKGROUND KNOWLEDGE

6. Dr. Shamos applied the same level of ordinary skill that I proposed in my opening declaration: a bachelor's degree in electrical engineering, computer engineering, or an equivalent field, together with two years of experience working on remote device control and/or networked devices, with an advanced degree potentially substituting for some work experience. (EX1002, ¶57; EX2011, ¶¶23-24.) There is therefore no material dispute concerning the level of ordinary skill.

7. The agreed POSITA would have brought practical knowledge of ordinary client-server networking and networked devices to the references. A POSITA would have understood, and Dr. Shamos agreed at his deposition, that wireless home routers, Internet-connected servers, Internet-connected servers

located outside the home, client devices sending audio to servers, servers returning data in response to human voice, speech recognition, stored-voice comparisons, microphone arrays for source localization, LED status indicators, and loudspeakers converting electrical energy to mechanical movement and sound were all known by January 2012. (EX1102, 31:4-33:7.)

8. Dr. Shamos also agreed that by January 2012 offloading computation to a remote device would lessen the computational load on the local device and could permit the local device to be smaller or use fewer components, and that remote processing could reduce local power consumption. (EX1102, 33:8-19.) These are the same types of engineering considerations I discussed in my opening declaration when explaining how a POSITA would have understood the asserted references.

III. REBUTTAL OPINIONS

9. I understand that Patent Owner and Dr. Shamos have argued that the prior art does not disclose certain limitations claimed in the '721 patent under their narrow interpretations of several claim elements. In general, Patent Owner and Dr. Shamos' opinions depend on reading the claims more narrowly than the plain and ordinary meaning and demanding details that the claims do not recite. Their opinions also repeatedly evaluate individual references in isolation rather than the combinations actually proposed in the Petition.

10. I disagree with Dr. Shamos' opinions for the reasons explained below. His deposition testimony confirms several technical points that are inconsistent with the premises of his declaration, including that ordinary information can operate the '721 patent's actuators, that remote server architectures were conventional by January 2012, and that several dependent claims do not contain the additional implementations he attributes to them.

A. GROUNDS 1-8: LIM (CLAIMS 1, 5-6, 10, 12, 16, 20-21, 23, 29)

1. Dr. Shamos Incorrectly Rewrites "Message" as A Specialized "Control Message."

11. Dr. Shamos repeatedly treats the claimed "first message" as though it must be a device-control instruction. For example, he distinguishes informational content from a "message" by requiring that the message "must be addressed to a specific device; must carry an instruction in a form the receiving device's hardware or software is designed to act on; and must produce actuation without user involvement." (EX2011, ¶¶38, 48-50.)

12. I disagree with Dr. Shamos' opinion because claim 1 itself distinguishes a generic "first message" and "second message" from the downstream "control message." (EX1001, Claim 1.) Claim 1 requires the first device to receive "first and second messages" from the server, and separately, to send to the second device "a control message that comprises a control data for

controlling a first actuator.” (*Id.*) A POSITA would not read the term “message” as the separate “control message.” The extra requirements imposed by Dr. Shamos for “message” do not reflect the ordinary technical meaning of the word “message,” and are contrary to the intrinsic and extrinsic evidence. I further do not agree that recited “control message” requires the Patent Owner’s construction, for example, the three-part criterion.

13. The specification confirms my opinion. The ’721 patent defines “message” broadly as “an ordered series of characters or bits intended to convey a package of information.” (EX1001, 148:30-35.) The specification also describes different kinds of information that messages may carry, including notifications. (*Id.*, 123:64-124:8.) I therefore disagree with Dr. Shamos’ opinion.

14. The specification’s definition of “message” is consistent with the plain and ordinary meaning as provided in the dictionary. Cambridge Dictionary defines “message” as “a short piece of written or spoken information that is given or sent to someone.” (EX1099, 4.) The information may be a speech-recognition result, an identifier, a confidence value, a database result, audio content, a status value, or a command. Dr. Shamos also testified that he applied the plain and ordinary meaning of the ’721 claim terms and did not believe any term required an express construction for purposes of this IPR. (EX1102, 39:22-40:15.) Claim 1 does not recite a particular packet format, opcode, addressing syntax, command

semantics, or requirement that the message itself contain a standalone hardware instruction. (EX1001, Claim 1.) I disagree that a POSITA applying ordinary technical meaning would nevertheless import into “message” the specialized device-addressing, command-semantics, and no-user-involvement requirements Dr. Shamos places on that term.

2. Lim’s Remote Speech Recognition Result Is the Claimed First Message.

15. Dr. Shamos opines that Lim does not disclose a server-returned first message. (EX2011, ¶48.) I disagree with Dr. Shamos’ opinion because Lim’s disclosed workflow requires the result of the remote speech processing to be communicated back to the controlling device to cause actuation.

16. Lim explains that “spoken syllables may be digitized and presented to a speech recognition means”, and that “all or part of said speech recognition processing . . . may be performed by programming resident on a remote server 500 accessible via a network 502 such as the Internet[.]” (EX1005, [0027].) Lim then describes different subsequent prompts depending on whether the appliance information is recognized or further information is required. (*Id.*, [0028]-[0030].) Lim discloses that these prompts can be audio outputs (e.g., “synthesized spoken message,” “buzzer”). (*Id.*, [0019], [0027].) When the recognition occurs remotely, the controlling device would receive the speech processing outcome before it can provide the responsive audio prompt.

17. That returned speech processing result is a “message” under the ordinary meaning and under the ’721 patent’s own generic definition. The speech processing result from the server is a message that triggers and dictates the next audio output. Thus, it is information sent from server 500 to the first device in response to the captured voice data sent for processing.

18. This was also ordinary technology by the relevant date. Dr. Shamos agreed that, by January 2012, it was known for client devices to send audio to servers and for servers to return data to client devices in response to receiving human voice. (EX1102, 31:17-32:2.) I cited Gruber, Chang, and Tu in my opening declaration as documentary examples confirming that ordinary server behavior, not as references supplying claim elements. (*See* EX1002, ¶¶118-121.)

19. Moreover, I note that Patent Owner takes contradictory positions on Lim. On one hand, Patent Owner accuses the Petition’s theory to be an “inference,” alleging that Lim does not disclose the server returning any message. (Response, 14; EX2011, ¶48.) Yet, in the next paragraph, Patent Owner concedes that Lim’s “server . . . returns *recognition data*” to the local device. (Response, 15.) Neither Patent Owner nor Dr. Shamos offers any plausible alternative account of what Lim’s server could return after processing user speech.

3. **Lim Sounds Digital Audio Content in Response to the Returned First Message**

20. Dr. Shamos further attacks that Lim does not disclose limitation 1[H] as the local device's prompt is determined by its "pre-programmed app logic" and not a server returned message. (EX2011, ¶¶48-49.) I disagree with Dr. Shamos' opinion because these are not mutually exclusive. Even if the local device has preprogrammed app logic (not conceded), it does not mean that the prompt is therefore independent of the server message. As discussed above, in Lim, depending on the result, the first device may repeat a prompt, request additional information, or move to the next step. (EX1005, [0027]-[0030].) Lim makes it clear that without the speech processing result (first message) from the server, the local device has nothing that triggers it to provide the next prompt.

21. The claim does not require the first message itself to "instruct[] a controlled device how to use that data to operate an actuator." (EX2011, ¶38.) Claim 1 requires only that the first device sounds digital audio content "in response to" the received first message. (EX1001, Claim 1.) Dr. Shamos' contrary view effectively adds requirements that the claim does not recite. The '721 patent summarizes the invention as "system and method in a building or vehicle for an actuator operation *in response to a sensor according to a control logic*. . . ." (EX1001, Abstract.) Thus, by the patent's own description, "logic" does not preclude the actuator operation to be "in response to" received data. Patent Owner insists that the message must be the "cause of the sounding." (Response, 15.) That

is precisely what Lim discloses. The speech processing result (first message) *is* the cause of the sounding. (See EX1002, ¶¶120-124, 133-134.)

22. Nor does Lim require that any speech synthesis occur locally. Lim expressly teaches that the prompts may take the form of an illuminated indicator, displayed text or image, a buzzer, or a “synthesized spoken message” produced using speech-synthesis means and a speaker, but it does not specify that the speech synthesis must occur at the local device. (See, e.g., EX1005, [0027].) And even if synthesis occurs locally, the synthesized output is still generated in response to—and dictated by—the speech-processing result received from the server.

4. Lim’s Recognition Result and Downloaded Command Code Data Both Satisfy the Second Message Limitation.

23. Dr. Shamos calls the recognition result used in Lim’s FIG. 6 workflow merely an “intermediate recognition gate” insufficient to qualify as the “second message.” (EX2011, ¶50.) I disagree with Dr. Shamos’ opinion.

24. The relevant sequence is straightforward. Lim’s FIG. 6 shows that once the speech processing result of the user command is received from the server, the local device is triggered to, for example, load the corresponding command-code information (at step 610) for transmission (at step 614) to the controlled device. (Pet., 31, quoting EX1005, [0035]-[0036].) Lim is clear that the speech processing result returned to the first device is the cause and trigger of the

transmission of control message to the appliances as the first device cannot know whether to proceed to the command-provisioning step unless it receives the result of the speech-recognition processing. Indeed, Patent Owner characterized that Lim's server act as an "off-screen *speech-recognition* engine and as the database from which *codesets* may be downloaded." (Response, 18.) That is precisely the architecture Claim 1 describes: a server-side response is received as the second message, and the first device thereafter sends a control message to the second device.

25. Dr. Shamos also rejects the Petition's alternative mapping to downloaded command-code data. (EX2011, ¶50.) I disagree with Dr. Shamos' opinion.

26. Lim expressly teaches that command-code data may be stored at a remote server and that, once the desired function has been uniquely identified, the indicated command-code data may be provisioned to the controlling device, including by "downloading the data itself" from the server. (EX1005, [0022], [0024], [0036].) Patent Owner does not dispute that this command codeset is a "message" downloaded from the "server" or that Lim's control message is sent "in response to" the codeset. (Response, 17; Pet., 32-33, 45-46; EX1005, [0024], [0036].) Thus, Patent Owner effectively concedes that the command codeset

discloses all elements of “send[ing] . . . a control message . . . in response to the received second message.”

27. Dr. Shamos instead argues that the command codeset is not *generated* by the server “a database retrieval of pre-packaged command codes [is] a provisioning operation, not a message generated and sent by the server . . .” (EX2011, ¶50.) But “generate” is not in the claim, and Patent Owner does not explain where the “generate” requirement comes from. The ’721 patent merely recites “*receive, from the server device . . . first and second messages*, in response to the sending of the respectively captured first and second human voice data.” (EX1001, Claim 1.) The claim language does not specify how the server obtains the second message in the first place. Dr. Shamos thus invents another requirement. Nor does the ’721 patent equate “generate” with “creating.” (*See, e.g.,* EX1001, 91:26-27 (listing “generate” separately from “create”).)

28. I note that the separate communication carrying different information in Lim satisfies the claim differentiation between “second message” and “control message” in the ’721 patent. In Lim, the server-to-first-device communication may carry the *command-code information or speech processing result*, and the first device then uses that information to transmit the *appliance-specific control message*. (EX1005, [0017], [0020]-[0021], [0036].)

29. Patent Owner argues that Lim’s FIG. 6 “does not itself disclose any server generating or sending messages.” (Response, 18.) I disagree. Patent Owner itself characterizes server as “off-screen” here. *Id.* Lim discloses the server’s speech processing and returned results with reference to FIG. 6 and throughout the specification. (EX1005, [0027]-[0030], [0035]-[0036].) A POSITA would read those disclosures together, especially when the paragraphs offer discussion of FIG. 6.

5. The Background References Document POSITA Knowledge, They Do Not Replace Lim.

30. Dr. Shamos argues that my citations to Gruber, Chang, Tu, Doherty, Chatterjee, and other background materials confirm that Lim has missing limitations. (EX2011, ¶¶52-53.) I disagree with Dr. Shamos’ opinion because the cited materials were used to document what a POSITA already knew when reading Lim, not to supply missing elements.

31. For example, Gruber, Chang, and Tu document that such server-returned speech processing results were routine. Similarly, Doherty, Roberts, and Yadav document conventional networking details such as IP and MAC addressing. Those sources explain ordinary implementation details that a POSITA would bring to Lim’s express disclosure of Internet and WLAN-connected devices.

32. The same is true of my citations to other background references (*e.g.*, Goldsmith, Sinclair, Roberts, Yadav) for the dependent claims in Ground 1. The

references are cited to document POSITA's knowledge of conventional implementations. (EX2011, ¶53.) For example, Lim discloses that the first/controlling device communicates with the remote server 500 via WiFi. (Pet., 48-49, quoting EX1005, [0027], Figure 5b.) A POSITA reading Lim would understand that WiFi is based on the IEEE 802.11 standard group and that the 802.11b standards use ISM Band II frequencies. Goldsmith is cited to corroborate that background knowledge, not to supply a missing limitation or create a new obviousness combination.

33. As another example, Lim itself discloses that the first device may include a display and an LED. (Pet., 46-47, quoting EX1005, [0019], [0023].) A POSITA reading Lim would understand that displays and LEDs are electric light sources meeting claims 26-27. Sinclair is cited merely to corroborate that background knowledge; it is not used to supply missing limitations.

34. Dr. Shamos himself acknowledged at deposition that a reference can be cited to establish background knowledge or motivation without itself being relied on as an enabling implementation of the claimed system. (EX1102, 18:6-19:18.) That testimony is consistent with how I used such background materials in EX1002.

35. Patent Owner identifies no deficiency in Lim. Lim supplies the "message" limitations, grounds 2-8 do not rely on secondary references for them.

B. GROUND 9-16: WILLIAMS-LIM (CLAIMS 1-15, 17-29, 31, 33)

1. Williams’ “Services 218” Disclose “Second Message.”

36. Dr. Shamos argues that Petition’s mapping of the “second message” fails because Williams-Lim does not disclose “a second message generated in response to a voice command and received by the first device to trigger a relayed control message.” (EX2011, ¶¶55-56.) I disagree. Patent Owner attempts to import a “relay” requirement into the claims, asserting that “the second [message] causes the first device to relay a ‘control message.’” (Response, 22.) Under this false premise, Patent Owner took a step further to equate the claimed “second message” with a “control message,” despite the claim’s express distinction between the two. (*Id.*, 23.) But claim 1 neither recites “relay” nor equates “second message” with “control message.” Instead, claim 1 requires only that the first device “receive[s], from the server device . . . first and second messages, in response to the sending of the respectively captured first and second human data,” and “send[s], to a second device over the WLAN, a control message that comprises a control data for controlling a first actuator in the second device . . . in response to the received second message.” The claim terms “second message” and “control message” are distinct and not interchangeable. The second message is a generic message that encompasses information, not a control message.

37. Dr. Shamos opines that Williams' services 218 are merely "informational content" and therefore are not messages. (EX2011, ¶56.) I disagree with Dr. Shamos' opinion because "informational content" and "message" are not mutually exclusive technical categories. *See supra* section III.A.1.

38. Williams explains that the server initiates or customizes services 218 to media devices 106 and application devices 220. (EX1006, [0036].) This requires communicating information related to the services 218 to that device. Thus, Williams' "services 218" disclose the recited "second message."

39. Williams teaches that the server can provide audio 222 for transmission to interface device 116 to be played on speaker 206, including speech-synthesized audio. (EX1006, [0037], [0045], [0048].) That server-provided audio is information transmitted to the first device, and the first device sounds it using its speaker. This disclosure is fully consistent with the ordinary meaning of a received "first message." Dr. Shamos describes that audio content is "informational content" and agreed at deposition that audio data is information. (EX2011, ¶38; EX1102, 37:11-16.)

40. Dr. Shamos then faults "services 218" for not "trigger[ing] a relayed control message." (EX2011, ¶56.) I disagree.

41. First, as discussed, claim 1 requires no “relay” of any message that Dr. Shamos improperly added. Those requirements are also contrary to the extrinsic and intrinsic evidence.

42. Second, Dr. Shamos is attacking Williams in isolation rather than addressing the proposed combination. In the combination, Lim supplies the client-device architecture in which the client device receives server communications and sends control commands to controlled devices. Williams is not relied upon independently to disclose the server-to-first-device transmission path or the subsequently sending of control message.

43. Williams also expressly applies its architecture to home-device control, contrary to Dr. Shamos’ assertions. (EX2011, ¶39.) Williams’ services 218 include home devices: “[s]ervices may also be provided in conjunction with other types of systems, including telephone systems, media players, automation systems and devices, and so forth.” (EX1006, [0036].) Paragraph [0051] states that “[a]n application might also be designed to allow a user to control various on-premises devices and equipment such as media equipment and home-automation devices . . . may accept spoken commands from a user regarding on-premises devices, and may respond by interacting with such devices through available network connections.” (*Id.*, [0051].)

2. Williams Paragraph [0051] Describes a Concrete System Capability, Not an Aspiration.

44. I disagree with Dr. Shamos' opinion that paragraph [0051] is merely "aspirational" and says nothing about implementation. (EX2011, ¶¶39, 57.) Paragraph [0051] is not a disconnected statement about a different system. It appears under the "Example Services" heading and beneath paragraph [0043] which states that "[t]he *architecture above* can be utilized to provide a wide variety of services and functionality, of which *the following are simple examples.*" (EX1006, [0043].) Thus, paragraph [0051] uses the same application architecture described in the preceding paragraphs and is one of the "simple examples" of the "wide variety of services and functionality." *Id.*

45. Williams already describes the implementation architecture with details of processing of audio commands, resulting provisions of services 218 to controlled devices, and architecture of network environments. (EX1006, [0011]-[0015], [0021]-[0022], [0036], [0046], [0051]-[0052], FIGs. 1-3.) Paragraph [0051] identifies home-device control as one such service that uses that architecture, thereby disclosing the message, messaging path, and form of communication.

3. Williams and Lim Supply Complementary Portions of the Combined Architecture.

46. Dr. Shamos argues that Williams sends services directly to media devices and therefore does not show interface device 116 receiving the second message and then transmitting the control message to the controlled device.

(EX2011, ¶¶56-58.) I disagree with Dr. Shamos' opinion because Ground 9 is a combination of Williams and Lim. Williams supplies the cloud-side responsive information and audio; Lim supplies the client-side control sequence in which a local first device receives a message and then transmits an appliance-specific command.

47. The combination therefore does not depend on Williams alone. A POSITA applying Lim's known communication technique to Williams would leverage the interface device 116 as a hub to transmit control signals to the target device. (Pet., 71-72, 84-92; EX1002, ¶¶324-328, 385-403.)

48. Nor does the combination somehow reduce to Lim alone. Reliance on a secondary reference (*e.g.*, Lim in Williams-Lim) does not reduce the entire ground to it. Further, as explained above, Williams contributes the specific server-side services architecture, including cloud-based server, responsive information, audio 222, and voice-control of on-premises devices, among other things. Lim contributes a known way for a local device to control heterogenous appliances in response to information received from a remote processing service. The resulting system reflects the teachings of both references.

4. The Combination¹ Has Concrete Engineering Advantages and a Reference-Based Motivation.

49. Dr. Shamos opines that the proposed combination is contrary to “engineering reality” because it adds a hop, an additional failure point, and additional connection-management burden. (EX2011, ¶60.) I disagree with Dr. Shamos’ opinion because simply counting communication hops does not account for the function performed by the local first device or the engineering tradeoffs of managing heterogeneous home devices.

50. First, the combination does not insert a new physical box into Williams. Interface device 116 already exists in Williams, is already network connected, and is already the endpoint through which the user supplies voice information to the remote applications. For the voice-control feature at issue, if interface device 116 is unavailable, the user’s spoken command cannot reach Williams’ server in the first place. Giving that existing endpoint Lim’s additional downstream-control functionality therefore does not add a wholly independent hardware dependency in the manner Dr. Shamos suggests.

51. Second, Lim supplies a concrete design need and solution: support for heterogeneous and legacy appliance-control mechanisms. Lim expressly teaches

¹ Patent Owner mischaracterizes the rationale for the Williams-Lim combination as a “Relay Rationale.” I never mention a “relay.”

that the controlling device may transmit commands using “IR, RF, Point-to-Point, or networked protocol[s].” (EX1002, ¶325; EX1005, [0017].) A remote Internet server cannot itself emit an infrared command in a user’s living room, and many local RF or point-to-point protocols are not directly exposed as Internet endpoints. A local control device is therefore a natural protocol boundary between a cloud service and heterogeneous appliances, permitting more flexible deployment and more reliable local control.

52. A POSITA implementing Williams’ stated goal of controlling “various on-premises devices and equipment” would recognize that not every target appliance would present the same network interface, command format, or cloud connectivity. (EX1006, [0051].) Lim teaches a practical way to solve that problem: the remote system provides information identifying the requested operation, while the local controller sends the appliance-specific command using the available local protocol. (EX1005, [0007], [0017], [0020]-[0021], [0036].)

53. Third, centralization changes where connection-management complexity resides. Dr. Shamos treats a local IR, RF, or WLAN communication as though it imposes the same management burden on the remote server as a separately authenticated Internet-facing appliance connection. A POSITA would not regard those as similar. With a local hub, the server can maintain a common cloud-facing relationship with one known first device while appliance-specific

addressing, discovery, command formats, and local link management remain inside the premises.

54. This separation also improves scalability in a concrete way. Adding another appliance need not require the remote server to establish another independently Internet-facing control relationship, implement that appliance's local protocol, or maintain another set of appliance-specific connection state. The appliance can instead be registered with, and controlled by, the local controller, which alone maintains the relationship with the server. That is the network-management benefit I described in my opening declaration. (EX1002, ¶326.) And such an architecture is well-known and predates the '721 patent by decades. (*Id.*, ¶¶324-328.)

55. Fourth, the firewall and security benefits are also concrete. Direct cloud control of each appliance requires each controlled appliance to provide some cloud-facing mechanism by which the remote service can reach or maintain a session with it. That can be implemented through inbound reachability, port forwarding, NAT traversal, or persistent outbound sessions, but in each case the Internet-facing communication, authentication, and connection state must be implemented and maintained for each appliance. A hub architecture can concentrate those functions in the first device while allowing the appliances themselves to remain locally reachable. (EX1002, ¶327.)

56. The engineering benefit of a hub architecture is that a POSITA can reduce the number of Internet-facing endpoints and centralize cloud authentication and encryption handling in the first device rather than duplicating those functions across every controlled appliance. That is particularly useful for legacy or resource-limited appliances that were not designed to communicate independently with a remote Internet service.

57. Fifth, Dr. Shamos' "additional failure point" criticism is incorrect. The first device becomes a shared local control point, but the architecture can eliminate multiple independent cloud-facing appliance sessions. And, as explained above, Williams' interface device 116 is already a shared dependency for the voice-command transaction because it captures the user input.

58. Dr. Shamos also asserts that the combination would "dismantle" Williams' direct server-to-device path. (EX2011, ¶61.) I disagree with Dr. Shamos' opinion because my proposed combination does not require removing Williams' ability to provide appropriate services directly to media or application devices. Williams can continue delivering content directly where that is useful. For the home-device-control functionality Williams expressly contemplates, however, a POSITA would have reason to use Lim's local-controller technique. The two paths are not mutually exclusive.

59. Williams itself supplies the design objective that makes Lim relevant: accepting spoken commands concerning various on-premises devices and interacting with those devices. (EX1006, [0051].) Lim addresses how a local device can use the results of a remote processing service's recognition of a user request or use server-downloaded command code data to transmit control message to different appliances. That reference-to-reference fit is a concrete motivation, not a hindsight reconstruction.

60. Finally, I did identify the finite alternatives relevant to the communication topology. In EX1002, ¶328, I explained that home appliances could be controlled either through direct server-to-device communications or by a local hub device that sends control signals to the appliances in response to messages from a remote server. Lim demonstrates the hub approach. Both architectures were established and predictable. Applying Lim's demonstrated client-side control technique to Williams' existing networked interface device therefore would have involved a routine allocation of known communication functions, with a reasonable expectation of success. (EX1002, ¶328.)

61. Dr. Shamos further objects to my citations to Van der Meulen, Wittenberg, Chatterjee, and Doherty. (EX2011, ¶59.) I cite them to corroborate conventional implementations and POSITA's background knowledge such as the

prevalence of WiFi controlled appliances. They are not substituting any missing architecture.

62. Dr. Shamos argues Grounds 10-16 fail because they inherit Ground 9's alleged deficiency. (EX2011, ¶62.) Because Ground 9 does not fail, Grounds 10-16 likewise survive.

C. THE DEPENDENT CLAIMS ARE UNPATENTABLE

1. Claim 2: Wittenberg and Howard Disclose Identifying A Specific Person by Voice.

63. Dr. Shamos opines that Wittenberg performs only authorization and therefore does not identify the voice of a specific person. (EX2011, ¶63.) I disagree with Dr. Shamos' opinion because Wittenberg performs that authorization by comparing characteristics of the received voice to stored voice information associated with authorized users (that is, specific individual or individuals). (EX1008, 5:61-6:6.) The biometric comparison that determines which stored voice information matches identifies the received voice as the voice represented by that stored information.

64. Dr. Shamos' declaration places weight on Wittenberg's supposed lack of an output or use of the speaker's individual identity. (EX2011, ¶63.) But at deposition Dr. Shamos agreed that claim 2 does not recite reporting the identity to a user or server and stated that claim 2 "requires a step of identifying it" and "doesn't require more." He clarified that "it" is "[t]he voice of a specific person."

(EX1102, 41:6-24.) I therefore disagree with Dr. Shamos' opinion because he demands an output or downstream use that claim 2 does not require.

65. Wittenberg's "stored voice information" is also not an abstract authorization token. When asked where a POSITA would understand such stored voice information could come from, Dr. Shamos identified as an obvious source that people record samples of their voices. (EX1102, 46:18-47:4.) Matching the received voice characteristics to an enrolled person's voice sample is a voice-based identification of that enrolled speaker.

66. Dr. Shamos misinterprets Wittenberg's express disclosures, saying that Wittenberg's authorization operation "might well be done by IP address or a password." (EX2011, ¶63.) But Wittenberg expressly provides that its authorization is through voice recognition processing, and that "*voice recognition processing* provides an additional degree of security," which I explained in my opening declaration as one of the reasons for combining Wittenberg's disclosure with Williams and Howard-Doherty. (EX1002, ¶¶199-206.)

2. Claim 3: Gruber Makes the Response Depend on Sensor Data.

67. Dr. Shamos opines that Gruber's sensor data may assist interpretation but does not make the server's first message itself "in response to" the sensor data. (EX2011, ¶64.) I disagree with Dr. Shamos' opinion because Gruber expressly uses sensor data to affect interpretation of the user's request and then generates the

dialogue response based at least in part on that sensor data. (EX1009, [0058], [0198]-[0199], [0218]-[0225], claims 42, 52.)

68. Dr. Shamos confirmed this point at deposition. When asked how a POSITA would understand claim 52 of Gruber, he explained that the system is interpreting user input to perform a task and that claim 52 generates a dialogue response “based at least in part on the received context information.” He further testified that “[t]he response is based on the received context information. It is not just any old random response.” (EX1102, 48:11-49:15.)

69. In the proposed combination, Lim and Williams-Lim supply the server-side processing and responsive first message, while Gruber teaches using sensor data—such as location, ambient light, or temperature—to influence that processing. Indeed, I explained in my opening declaration that, in the combination, Lim’s server supplements speech processing with sensor data and “transmit the speech processing result(s) (e.g., first message) back to the first device.” (EX1002, ¶216.) The combination thus involves the first message sent in response to voice processing (claim 1) and the voice processing is based on sensor-data (e.g., GPS location, light, or temperature) (claim 3).

3. Claim 4: Gruber Discloses Conventional Electronic Temperature and Light Sensors.

70. Dr. Shamos opines that Gruber’s “thermometer” and “light sensor” are too generic to disclose the sensor types recited in claim 4. (EX2011, ¶65.) I

disagree with Dr. Shamos' opinion because a POSITA would read those sensors in the context of Gruber's electronic device and would plainly understand that a thermometer is a "thermoelectric sensor that responds to a temperature . . . using conduction, convection, or radiation," and that a light sensor is a "photoelectric sensor that responds to . . . visible or . . . invisible light." Claim 4 recites conventional manners of operations for these components.

71. Gruber identifies device sensors including a light sensor and thermometer that supply sensor data as context. (EX1009, [0222]-[0225].) A POSITA implementing those sensors in an electronic device would use ordinary electronic temperature and light-sensing components that respond to the physical phenomena they measure. Fraden (EX1030) documents that conventional technical understanding, it does not replace Gruber's disclosure of the thermometer and light sensor or introduce new components.

72. The prosecution history confirms why a POSITA would not demand the heightened level of disclosure Dr. Shamos requires. The Board previously treated a reference disclosing a sensor that sensed temperature as disclosing the corresponding claim 4 subject matter. (EX1004, 239-240, 371-372; EX1100, [0028], [0052].) My technical opinion is consistent with that ordinary understanding of the sensor terminology.

4. Claims 7-8: Benesty Supplies the Directional-Array Estimation Capability.

73. Dr. Shamos states that claim 8 requires particular array geometry, signal-processing algorithms, and processing resources, and criticizes the Petition for not specifying those implementation details. (EX2011, ¶67.) I disagree with Dr. Shamos' opinion because claim 8 does not recite those specific requirements, and Benesty expressly teaches directional microphone-array techniques that perform the claimed estimation functionality.

74. At deposition, Dr. Shamos agreed that the "or" in claim 8 ordinarily permits any one of the listed estimated quantities, that claim 8 does not require a specific localization algorithm, that it does not expressly require a numerical accuracy threshold, and that it does not require any processing-resource threshold. (EX1102, 56:1-57:14.) Those admissions are inconsistent with using unspecified geometry, algorithm, accuracy, or resource details as a reason to reject the Petition's mapping. (EX2011, ¶67.)

75. Benesty does not merely teach adding multiple microphones. It teaches that microphone arrays provide directional gains and describes beam scanning and time-delay-estimation as widely used approaches for speech-source localization. (EX1010, 1021, 1044.) Those techniques provide the estimation functionality recited by claim 8, including direction-of-arrival information.

76. The motivation is likewise concrete. Lim and Williams-Lim capture voice in a home environment where reverberation, background noise, and changing

speaker position are ordinary design problems. Benesty's directional-array techniques address those problems by improving pickup and source localization. Dr. Shamos agreed that microphone arrays for source localization were known by January 2012. (EX1102, 32:11-14.) A POSITA would have had a direct reason to apply Benesty's known techniques to the voice-capture devices in Lim and Williams-Lim. This is supported by the prosecution history of the '710 parent patent. I note that the Board previously found use of directional microphone arrays, including Direction-of-Arrival functionality, to be a predictable improvement for detecting speech while attenuating sounds from other directions. (EX1004, 372-373.)

5. Claims 10-12: Stored Network Addresses Were Routine Device-Identification Mechanisms.

77. Dr. Shamos objects that the Petition cites Doherty, Roberts, and Yadav when discussing conventional IP and MAC addressing. (EX2011, ¶¶53, 66.) I disagree with Dr. Shamos' opinion because Lim expressly discloses networked first devices, WLAN/Internet communication, and multiple forms of memory. (EX1005, [0019], [0023], [0027], [0038], FIG. 5b.) The cited networking materials document the conventional addressing properties that a POSITA would understand those networked devices to use.

78. For Claim 10, Lim discloses ROM, RAM, nonvolatile memory, and smart-device memory including ROM, RAM, and Flash. (EX1005, [0019],

[0023].) A POSITA implementing the disclosed IP-connected device would understand that the device's IP configuration is maintained in memory accessible to the networking software. Doherty provides an example showing networked devices displaying their associated IP addresses. (EX1007, 211, FIG. 10-44, 265; EX1002, ¶¶167-168.) Indeed, I note that during prosecution of the '710 parent patent, the Board found that making networked devices addressable and storing those addresses in device memory were obvious implementations of ordinary TCP/IP networking. (EX1004, 373-374.)

79. For Claim 11, conventional WLAN devices used MAC-layer addresses. Doherty and Yadav document that wireless network interfaces have unique MAC addresses and that common IEEE address forms included MAC-48, EUI-48, and EUI-64. (EX1007, 99, 166; EX1026, 2:50-59; EX1002, ¶169.) Dr. Shamos also ignored the express teachings of references in the Grounds. For both Grounds 3 and 11, Gruber expressly discloses "MAC addresses" as "signatures of connected networks." (EX1009, [0218]-[0225].) For Claim 12, Roberts and Doherty document the equally conventional use of IPv4 addresses and static or dynamic address assignment. (EX1029, 127; EX1007, 277; EX1002, ¶¶170-171.) Those references corroborate the technical knowledge a POSITA would bring to Lim and Williams-Lim; they do not substitute a missing network architecture or first device component for Lim or Williams-Lim.

6. Claims 13-15: Goldsmith Provides Conventional Wireless-Network Alternatives.

80. Dr. Shamos opines that Goldsmith is merely a protocol survey, not the network architecture (e.g., “via the WLAN” and “over the Internet”), and does not provide a reason to select the claimed wireless standards. (EX2011, ¶¶70-71.) I disagree with Dr. Shamos’ opinion because the design need is already present in Lim and Williams-Lim which discloses the network architecture: the first device is a wireless device that communicates over WLAN/Internet connections and may be implemented as a smartphone. (Pet., 13-15, 24-28, 72-74, 80-84; EX1002, ¶¶78-85, 106-120, 328-341, 360-381.) Goldsmith identifies established standards available to implement that existing wireless communication requirement.

81. Goldsmith expressly discusses IEEE 802.11 technologies, operation in ISM bands, and the recited cellular families including W-CDMA, UMTS, GSM, and EDGE. (EX1040, 45-46, 55-56, 65-66.) Dr. Shamos confirmed those technologies such as IEEE 802.11, ISM-band networks, and cellular were known. (EX1102, 32:15-21; EX1040, 45-46, 55-56, 65-66.) Selecting among these established standards for a device already required to communicate wirelessly was a routine design choice with predictable results. The Petition’s citation to other consumer-device examples corroborates that conventional implementation and support rationales to combine, it does not replace Lim, Williams-Lim, or Goldsmith. Dr. Shamos also ignored my explanations that a POSITA would

understand that such implementations were technically feasible and commonplace, including other rationales such as the finite number of predictable solutions, reasonable expectations of success, and known design incentives and market forces. (EX1002, ¶¶266-284, 500-518.)

7. Claims 18-20, 23: Chatterjee Supplies the Recited Appliance Types, Not the Server Architecture.

82. Dr. Shamos criticizes Chatterjee because its exemplary appliances are directed from a user's mobile device rather than from the claimed Internet server. (EX2011, ¶72.) I disagree with Dr. Shamos' opinion because the Petition did not rely on Chatterjee to replace the server-control pathway, which is supplied by Lim or Williams-Lim. Chatterjee supplies the known remotely-controllable appliances to which the primary combination's voice-control architecture is applied. (Pet., 64-67, 102-103; EX1002, ¶¶285-305, 519-529.)

83. Chatterjee expressly identifies remotely controllable ovens, dishwashers, refrigerators, HVAC equipment, and other home appliances with device controllers integrated into the appliance. (EX1012, Abstract.) Chatterjee expressly disclosed the appliances' ability to be readily remotely controlled. (EX1012, [0012]; EX1002, ¶¶290-292, 519-522.) A POSITA would have understood that extending an existing remote home-control system to these already-controllable appliances predictably yields server-based voice control of those appliance types. (*Id.*)

84. Further, Dr. Shamos attacks Chatterjee for lacking disclosures of “how [Chatterjee’s] appliances would be integrated into” Lim or Williams’ architecture. (EX2011, ¶72.) But I note that claims 18-20, and 23 do not specify how the appliance is integrated with the controlled device, they merely recite laundry lists of conventional appliances. The Petition and my opening declaration explained how Chatterjee discloses these appliances. (Pet., 64-67, 102-103; EX1002, ¶¶285-305, 519-529.)

8. Claim 25: A Remote Server External to the Home Was an Ordinary Implementation

85. Dr. Shamos objects that the Petition cites background references when explaining why Lim’s remote server would be external to the building and why a portable first device would be used within a home. (EX2011, ¶53.) I disagree with Dr. Shamos’ opinion because Lim itself repeatedly distinguishes remote server 500 from local components and states that speech recognition may be performed “remotely at a Web server.” (EX1005, [0007], [0024], [0027].)

86. Lim also teaches that the first device may be a smartphone or other portable smart device used to control home appliances. (EX1005, [0018], [0023].) A POSITA would naturally understand such a portable controller to be used within the home being controlled. The cited Howard example documents that conventional use. (EX1021, 1:53-58, 2:19-26; EX1002, ¶183.)

87. The use of an external server for resource-intensive processing was likewise conventional. Dr. Shamos agreed that, by January 2012, Internet-connected servers outside a home were known and remote computation could reduce local processing burden and power requirements. (EX1102, 31:9-16, 33:8-19.) Those facts corroborate my opinion that Lim’s “remote” Web server would naturally be implemented outside the user’s building. (EX1002, ¶184.)

9. Claims 5-6, 26-28: Rosenberger Supplies a Conventional Light Source Actuator.

88. Dr. Shamos opines that Rosenberger’s LED indicates locally determined speech-recognition states and therefore does not satisfy the dependent claims. (EX2011, ¶¶73-74.) I disagree with Dr. Shamos’ opinion because Claims 5-6 and 26-27 do not require the additional actuator or light source itself to perform server-side recognition or to operate directly from a server message. They require the first device to include the recited actuator/light-source structure.

89. Rosenberger expressly discloses a multi-colored LED status light used to communicate device states and modes. (EX1013, 8:8-16, 8:20-32.) Rosenberger teaches that, “[u]pon receiving and recognizing a speech trigger phrase,” its voice-controlled device “changes its status light 34” to prompt the user for a command. (*Id.*) Lim itself also uses illuminated visual feedback during voice interaction. As I explained in my opening declaration, a POSITA would have recognized the same benefit in audio-based systems such as Lim and Williams-

Lim, and would have included Rosenberger's known multi-color status indication in the first device to "indicate different device states and modes" and "improve the usability of the device." (Pet., 67-69, 100-101; EX1002, ¶¶309-316, 472-476.)

Incorporating Rosenberger's into Lim or Williams-Lim provides immediate visual feedback concerning listening, processing, status, or mode without changing where speech recognition occurs. Dr. Shamos agreed that LED status indicators were known output components by January 2012. (EX1102, 32:22-33:1.)

90. I also disagree with Dr. Shamos' Claim 28 criticism because Rosenberger is not relied on to supply Claim 28's added synthesized-audio limitation. Rosenberger supplies the light-source limitation of Claim 27. Claim 28's additional sounding requirement is supplied by Lim's "synthesized spoken message" in Ground 8 and by Williams' speech-synthesized audio 222 in Ground 12. (EX1005, [0027]; EX1006, [0037], [0045]-[0050]; Pet., 67-69, 99-101.)

10. Claims 9, 29, 30 and 33: Rossing Confirms Conventional Acoustic Implementations.

91. Dr. Shamos opines that Rossing merely shows that different microphone constructions existed and supplies no reason to choose one. (EX2011, ¶68.) I disagree with Dr. Shamos' opinion because Lim and Williams-Lim already require a microphone to capture voice. Rossing identifies the conventional microphone constructions available to implement that already-required function,

including condenser, electret, dynamic, ribbon, carbon, and piezoelectric types.

(EX1014, 757-759.)

92. The design need therefore does not have to be invented from the claim. The primary systems need a working microphone. Selecting a known microphone structure designed to perform voice capture yields the ordinary and predictable result for which that structure was developed. Rossing also identifies conventional omnidirectional microphones. (EX1014, 309.)

93. Further, the claim itself specifies no particular design or choice; it strings together a laundry list of conventional microphone features joined by “or”. (EX1001, Claim 9.) “[W]herein the microphone . . . is based on the sensing an incident sound-based motion of a diaphragm or a ribbon, *or* wherein the microphone comprises a condenser, an electret, a dynamic, a ribbon, a carbon, *or* a piezoelectric microphone.” (*Id.*) Patent Owner cannot recite a broad list of conventional microphone alternatives and then fault the prior art for disclosing these features. (Response, 36.) As the patent merely recites laundry lists of conventional components, and textbooks such as Rossing describe precisely those components, a POSITA would readily review such sources to implement Lim or Williams-Lim’s microphones.

94. The same reasoning applies to Claim 29. Lim and Williams-Lim already include speakers. The ’721 patent itself recognizes that a loudspeaker is a

sounder, and textbooks such as Rossing describe conventional loudspeaker transduction types including magnetic/dynamic, piezoelectric, condenser/electrostatic implementations. (EX1001, 105: 61-106:4; EX1014, 763.) A POSITA implementing the existing speaker requirement would select among those familiar implementations according to ordinary design considerations, with a reasonable expectation that the speaker would perform its known sound-producing function.

95. Claim 30 is even more direct. Lim expressly discloses a “buzzer” among the feedback devices of controlling device 100. (EX1005, [0019]; EX1002, ¶¶192-193.) Thus, the claim does not depend on Rossing or another outside reference to supply the recited acoustic device.

96. For Claim 33, I disagree with Dr. Shamos’ opinion because Rossing expressly teaches the omnidirectional alternative recited by the claim. (EX2011, ¶69.) Rossing explains that sound radiates from a loudspeaker in all directions and that loudspeakers are “almost unavoidably omni-directional at low frequencies.” (EX1014, 765; EX1002, ¶¶242, 258-259.) Claim 33 is disjunctive: it permits an omnidirectional, unidirectional, or bidirectional pattern. It does not require a specially engineered, selectable, or frequency-invariant radiation pattern.

97. My opinion is not that every loudspeaker is omnidirectional under every operating condition. Rossing itself explains that directionality varies with

frequency. The relevant point is that Rossing expressly teaches operation having the omnidirectional radiation pattern that Claim 33 permits. Dr. Shamos identifies no technical requirement in the claim that excludes Rossing's low-frequency omnidirectional operation.

D. THE BACKGROUND REFERENCES DOCUMENT POSITA KNOWLEDGE OR SUPPORT MOTIVATION TO COMBINE; THEY DO NOT REPLACE THE ASSERTED COMBINATIONS

98. Dr. Shamos repeatedly treats any citation documenting ordinary engineering knowledge as though the cited document becomes another reference in the asserted combination. I disagree with Dr. Shamos' opinion because my substantive mappings remain to the references identified in the asserted grounds; the additional materials document how a POSITA would understand and implement those express disclosures and augment the rationales for motivation to combine.

99. Examples include remote servers returning speech-recognition results, conventional IP and MAC addressing, ordinary sensor implementations, standard wireless protocols, and conventional microphone or loudspeaker mechanisms. These were well-developed technologies by January 2012. Documentary sources confirming that background knowledge and supporting the rationales for motivation to combine, do not change the engineering architecture supplied by Lim, Williams, or the identified secondary reference.

100. This distinction is especially important when a reference already discloses the functional component at issue. For example, Lim expressly discloses memory, WLAN/Internet connectivity, a microphone, a speaker, a buzzer, and remote server processing. References such as Doherty, Rossing, and Roberts are useful evidence of how a POSITA would implement or understand those familiar components and augment the rationales for motivation to combine. They are not being used to replace Lim.

101. For these reasons, I disagree with Dr. Shamos' conclusion that the cited background materials demonstrate technical gaps in the asserted references. They instead confirm that the Petition and my opening declaration viewed the express prior-art disclosures through the knowledge of the agreed skilled artisan having education and practical experience in networked devices and remote device control, rather than as if the references were being read by a layperson with no engineering background. Dr. Shamos' own deposition confirms that the relevant networking, server, voice-processing, microphone-array, LED, and loudspeaker technologies were known by January 2012. (EX1102, 31:4-33:19.) My use of background references is consistent with that technical reality.

IV. CONCLUSION

102. For the reasons above and those set forth in my opening declaration, I disagree with Dr. Shamos' opinions that the asserted grounds fail to disclose or

render obvious the challenged limitations. His analysis repeatedly adds technical requirements that the claims do not recite, reads the asserted references more narrowly than a POSITA would, and evaluates secondary references as though each had to reproduce the entire primary-reference architecture.

103. Nothing in the Patent Owner Response, Dr. Shamos' declaration, or deposition testimony changes my opinions that the challenged claims are rendered obvious by the grounds set forth in the Petition.

* * *

I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and that these statements were made with knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code.

Dated: October 9, 2026

A handwritten signature in black ink that reads "Henry H. Houh." The signature is written in a cursive style with a period at the end.

Henry H. Houh