

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. 12,401,720

Inter Partes Review No. IPR2026-00280

**PETITION FOR *INTER PARTES* REVIEW
UNDER 35 U.S.C. §§ 311-319 AND 37 C.F.R. § 42.100 *et seq.***

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. THE '720 PATENT	2
A. Summary	2
B. Effective Filing Date	3
C. Prosecution History	3
D. Level of Ordinary Skill in the Art	4
III. CLAIM CONSTRUCTION	5
IV. PRECISE REASONS FOR RELIEF REQUESTED	5
A. Ground 1: Lindahl and Williams Render Obvious Claims 1-2, 4-8, 11-26, and 29.....	8
1. Summary of Lindahl	8
2. Summary of Williams	10
3. A POSITA would have been motivated to combine Lindahl and Williams.....	12
4. Claim 1	14
5. Claim 2.....	34
6. Claims 4/5/6	35
7. Claims 7/8/11/12	36
8. Claims 13/14/15	37
9. Claims 16/17/18/19	37
10. Claims 20/21	38
11. Claim 22	39
12. Claims 23/24/25/26/29.....	40
B. Ground 2: Lindahl, Williams and iPhone User Guide Render Obvious Claims 1-2, 23-26, 29	41
1. Summary of iPhone User Guide	41
2. Claim 1	42
3. Claim 2.....	44

TABLE OF CONTENTS
(continued)

		Page
4.	Claim 23	45
5.	Claims 24/25/29	46
6.	Claim 26	46
C.	Ground 3: Lindahl, Williams and Doherty Render Obvious	
	Claims 4-5	47
1.	Summary of Doherty	47
2.	Claims 4/5	47
D.	Ground 4: Lindahl, Williams and Rossing Render Obvious	
	Claims 7-8, 11-12	48
1.	Summary of Rossing	48
2.	Claims 7/8	48
3.	Claims 11/12	49
E.	Ground 5: Lindahl, Williams and Benesty Render Obvious	
	Claims 9-10	51
1.	Summary of Benesty	51
2.	Claims 9/10	51
F.	Ground 6: Lindahl, Williams and Goldsmith Render Obvious	
	Claims 20-21	53
1.	Summary of Goldsmith	53
2.	Claim 20	53
3.	Claim 21	54
G.	Ground 7: Lindahl, Williams, and Slaby Render Obvious	
	Claims 14, 23, 28-29	54
1.	Summary of Slaby	54
2.	Claim 14	55
3.	Claims 23/28/29	55

TABLE OF CONTENTS
(continued)

	Page
H. Ground 8: Williams, Lim and Benesty Render Obvious Claims 1, 4-26, 29.....	57
1. Summary of Lim	57
2. A POSITA would have been motivated to combine Williams, Lim and Benesty.....	61
3. Claim 1	64
4. Claims 4/5/6	84
5. Claims 7/8/11/12	85
6. Claims 9/10	85
7. Claims 13/14/15	85
8. Claims 16/17/18/19	86
9. Claims 20/21	87
10. Claim 22	88
11. Claims 23/24/25/26/29	88
I. Ground 9: Williams, Lim, Benesty and iPhone User Guide Render Obvious Claim 2	89
J. Ground 10: Williams, Lim, Benesty and Doherty Render Obvious Claims 4-5	90
K. Ground 11: Williams, Lim, Benesty and Rossing Render Obvious Claims 7-8, 11-12	90
L. Ground 12: Williams, Lim, Benesty and Goldsmith Render Obvious Claims 20-21	91
M. Ground 13: Williams, Lim, Benesty, and Slaby Render Obvious Claims 14, 23, 28-29.	91
N. Ground 14: Lindahl and Evermann Render Obvious Claims 1-2, 4-8, 11-26, 29.	92
1. Summary of Evermann	92

TABLE OF CONTENTS
(continued)

	Page
2. A POSITA would have been motivated to combine Lindahl and Evermann.....	94
3. Claim 1	95
4. Claim 2.....	101
5. Claims 4/5/6	102
6. Claims 7/8/11/12	102
7. Claims 13/14/15	102
8. Claims 16/17/18/19	103
9. Claims 20/21	103
10. Claim 22	103
11. Claims 23/24/25/26/29	104
V. NOTICES AND STATEMENTS.....	104

Exhibit List

Exhibit Description	Exhibit #
U.S. Patent No. 12,401,720 to Binder	1001
Expert Declaration of Dr. Henry Houh	1002
File History of U.S. Patent No. 12,401,720	1003
File History of U.S. Patent No. 11,128,710	1004
U.S. Patent Pub. No. 2013/0132094 to Lim	1005
U.S. Patent Pub. No. 2012/0198339 to Williams <i>et al.</i>	1006
Jim Doherty <i>et al.</i> , “Wireless Home Networking Simplified” (Cisco Press, 2007).	1007
iPhone User Guide For iOS 5.0 Software, available at https://web.archive.org/web/20111107154723/http://support.apple.com/manuals/en/iPhone_User_Guide.pdf	1008
U.S. Patent Pub. No. 2012/0265528 to Gruber <i>et al.</i>	1009
Jacob Benesty <i>et al.</i> , “Springer Handbook of Speech Processing” (Springer-Verlag Berlin Heidelberg, 2008).	1010
U.S. Patent Pub. No. 2010/0088100 to Lindahl	1011
U.S. Patent Pub. No. 2008/0154608 to Evermann <i>et al.</i>	1012
U.S. Patent No. 8,340,975 to Rosenberger	1013
Thomas D. Rossing <i>et al.</i> , “Springer Handbook of Acoustics” (Springer Science+Business Media, LLC, 2007).	1014
Michele Mazzucco <i>et al.</i> , “Profit-Aware Server Allocation for Green Internet Services” (2010 18 th Annual IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems) (available at https://ieeexplore.ieee.org/abstract/document/5581586)	1015

Exhibit Description	Exhibit #
Kenneth J. Christensen <i>et al.</i> , “The next frontier for communications networks: power management” (2004) (available at https://www.sciencedirect.com/science/article/abs/pii/S0140366404002427)	1016
Lidong Wang <i>et al.</i> , “MMSEmail: Delivering Emails to Mobile phone through MMS” (2009 IEEE International Conference on Multimedia Information Networking and Security) (available at https://ieeexplore.ieee.org/abstract/document/5370492)	1017
Ian Sinclair <i>et al.</i> , “Electronic and Electrical Servicing” (Elsevier Limited, 2 nd Edition, 2007)	1018
U.S. Patent No. 8,838,261 to Stewart <i>et al.</i>	1019
U.S. Patent No. 8,751,056 to Slaby <i>et al.</i>	1020
U.S. Patent Pub. No. 2011/0299412 to Diab <i>et al.</i>	1021
David Allan, “Effects of Popular Music in Advertising on Attention and Memory” (Journal of Advertising Research, Vol. 46, December 2006) (available at https://www.tandfonline.com/doi/abs/10.2501/S0021849906060491)	1022
Robert J. Shimonski <i>et al.</i> , “Sniffer Pro Network Optimization and Troubleshooting Handbook” (Syngress Publishing, Inc., 2002).	1023
U.S. Patent No. 8,200,798 to Yadav <i>et al.</i>	1024
RESERVED	1025
RESERVED	1026
RESERVED	1027
RESERVED	1028
P.A. Roberts <i>et al.</i> , “IP address management” (BT Technology Journal, Vol. 18, 2000) (available at	1029

Exhibit Description	Exhibit #
https://link.springer.com/article/10.1023/A:1026749131441)	
Jacob Fraden, “Handbook of Modern Sensors” (Springer Science+Business Media, LLC, 2010)	1030
RESERVED	1031
U.S. Patent No. 7,590,542 to Williams <i>et al.</i>	1032
RESERVED	1033
M. Needham, “Smart Home Market on Track for Rebound Thanks to Emerging Regions, According to IDC,” https://my.idc.com/getdoc.jsp?containerId=prUS52632124 , IDC Research (Oct. 1, 2024)	1034
U.S. Patent No. 7,593,718 to Gorday <i>et al.</i>	1035
U.S. Patent Pub. No. 2006/0135067 to Dunko	1036
David Chen et al., “Low-Cost Asset Tracking using Location-Aware Camera Phones” (Applications of Digital Image Processing XXXIII, Proceedings Volume 7798, 2010) (available at https://www.spiedigitallibrary.org/conference-proceedings-of-spie/7798/77980R/Low-cost-asset-tracking-using-location-aware-camera-phones/10.1117/12.862426.short)	1037
Yu-Shuo Chang et al., “CSR: a Cloud-assisted Speech Recognition Service for Personal Mobile Device” (2011 International Conference on Parallel Processing) (available at https://ieeexplore.ieee.org/document/6047199)	1038
Zhemin Tu et al., “Speech Recognition Over the Internet Using JAVA” (1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings, ICASSP99) (available at https://ieeexplore.ieee.org/document/758414)	1039
Declaration of Tony Corsini on the scanned copies of pages of Jim Doherty <i>et al.</i> , “Wireless Home Networking Simplified” (Cisco Press, 2007).	1040

Exhibit Description	Exhibit #
Declaration of Catherine Vassikova on the scanned copies of pages of Jacob Benesty <i>et al.</i> , “Springer Handbook of Speech Processing” (Springer-Verlag Berlin Heidelberg, 2008).	1041
Declaration of Rachel J. Watters on the scanned copies of Chapter 1 and Appendix D of Andrea Goldsmith, “Wireless Communications” (Cambridge University Press, 2005).	1042
Declaration of Tony Corsini on the scanned copies of pages of Thomas D. Rossing <i>et al.</i> , “Springer Handbook of Acoustics” (Springer Science+Business Media, LLC, 2007).	1043
Declaration of Gordon MacPherson on the IEEE publications Michele Mazzucco <i>et al.</i> , “Profit-Aware Server Allocation for Green Internet Services” (2010 18 th Annual IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems) and Lidong Wang <i>et al.</i> , “MMSEmail: Delivering Emails to Mobile phone through MMS” (2009 IEEE International Conference on Multimedia Information Networking and Security)	1044
Declaration of Catherine Vassikova on the scanned copies of pages of Kenneth J. Christensen <i>et al.</i> , “The next frontier for communications networks: power management” (2004).	1045
Declaration of Catherine Vassikova on the scanned copies of pages of Lidong Wang <i>et al.</i> , “MMSEmail: Delivering Emails to Mobile phone through MMS” (2009 IEEE International Conference on Multimedia Information Networking and Security).	1046
Declaration of Tony Corsini on the scanned copies of pages of Ian Sinclair <i>et al.</i> , “Electronic and Electrical Servicing” (Elsevier Limited, 2 nd Edition, 2007).	1047
Declaration of Mina Ching on the webpage https://web.archive.org/web/20111107154723/http://support.apple.com/manuals/en/iPhone_User_Guide.pdf , Internet Archive	1048
RESERVED	1049

Exhibit Description	Exhibit #
Declaration of Tony Corsini on the scanned copies of pages of P.A. Roberts <i>et al.</i> , “IP address management” (BT Technology Journal, Vol. 18, 2000)	1050
Declaration of Tony Corsini on the scanned copies of pages of Jacob Fraden, “Handbook of Modern Sensors” (Springer Science+Business Media, LLC, 2010)	1051
RESERVED	1052
Declaration of Tony Corsini on the scanned copies of pages of David Chen <i>et al.</i> , “Low-Cost Asset Tracking using Location-Aware Camera Phones” (Applications of Digital Image Processing XXXIII, Proceedings Volume 7798, 2010)	1053

Claims Listing

Claim	Limitation #	Limitation
1	1[PR]	A device comprising:
	1[A]	multiple microphones for capturing first human voice data;
	1[B]	a Wireless Local Area Network (WLAN) transceiver for transmitting digital data to, and for receiving digital data from, a WLAN;
	1[C]	a speaker for sounding audio data;
	1[D]	an electric light source for emitting light for emitting light;
	1[E]	an enclosure that houses the multiple microphones, the WLAN transceiver, the speaker, and the electric light source;
	1[F]	and one or more processors programmed with computer program instructions that, when executed, cause the device to:
	1[G]	send, to a server device over the Internet via the WLAN by the WLAN transceiver, at least part of the first human voice data captured by the multiple microphones;
	1[H]	play, by the speaker, an audio data received from the server device over the Internet via the WLAN by the WLAN transceiver;
	1[I]	and emit, by the electric light source, in response to data received from the server over the Internet via the WLAN by the WLAN transceiver,

Claim	Limitation #	Limitation
	1[J]	wherein the device is addressed in the WLAN using an Internet Protocol (IP) address that uniquely identifies the device in the WLAN or in the Internet.
2	2	The device according to claim 1, further comprising, in the enclosure, a rechargeable battery for powering the device.
3	3	The device according to claim 1, wherein the enclosure is a wearable enclosure.
4	4	The device according to claim 1, further being addressable using the IP address in the WLAN or in the Internet using distinct locally administered addresses or a universally administered digital addresses.
5	5	The device according to claim 4, wherein the IP address is a layer 3 address and is static or dynamic IP address that is IPv4 or IPv6 type address.
6	6	The device according to claim 1, further comprising a volatile or non-volatile memory that stores the IP address.
7	7	The device according to claim 1, wherein the speaker comprises an electromagnetic loudspeaker, a piezoelectric speaker, an electrostatic loudspeaker (ESL), a ribbon or planar magnetic loudspeaker, or a bending wave loudspeaker.
8	8	The device according to claim 1, wherein the speaker is operative for omnidirectional, unidirectional, or bidirectional pattern audio emitting.

Claim	Limitation #	Limitation
9	9	The device according to claim 1, wherein the multiple microphones are configured to improve far-field audio reception.
10	10	The device according to claim 9, wherein the multiple microphones are arranged as a directional microphones array operative to estimate a number, magnitude, frequency, Direction-Of-Arrival (DOA), distance, or speed of a phenomenon impinging the microphones array.
11	11	The device according to claim 1, wherein each of the multiple microphones comprises, or consists of, an omnidirectional, unidirectional, or bidirectional microphone that is based on sensing of an incident sound-based motion of a diaphragm or a ribbon.
12	12	The device according to claim 1, wherein each of the multiple microphones comprises, or consists of, a condenser, an electret, a dynamic, a ribbon, a carbon, or a piezoelectric microphone.
13	13	The device according to claim 1, wherein the light is a visible light for illumination or indication.
14	14	The device according to claim 1, wherein the light is a non-visible light that is infrared, ultraviolet, X-rays, or gamma rays.
15	15	The device according to claim 1, wherein the electric light source consists of, or comprises, a Solid-State Lighting (SSL), a Light Emitting Diode (LED), an Organic LED (OLED), a polymer LED (PLED), or a laser diode.

Claim	Limitation #	Limitation
16	16	The device according to claim 1, further comprising a memory that stores digital audio content, and the one or more processors are further programmed with computer program instructions that, when executed, cause the device to playing by the speaker the stored digital audio content.
17	17	The device according to claim 16, wherein the digital audio content is pre-recorded or synthesized audio content.
18	18	The device according to claim 16, further operative to play a simulating of a voice of a human being or to play music.
19	19	The device according to claim 18, wherein the digital audio content comprises a syllable, a word, a phrase, a sentence, a short story, or a long story, using male or female voice.
20	20	The device according to claim 1, wherein the WLAN is according to, or compatible with, IEEE 802.11-2012, IEEE 802.11a, IEEE 802.11b, IEEE 802.11g, IEEE 802.11n, or IEEE 802.11ac standard.
21	21	The device according to claim 1, wherein the WLAN uses a wireless communication over a licensed or an unlicensed radio frequency band, that is an Industrial, Scientific and Medical (ISM) radio band.
22	22	The device according to claim 1, wherein the enclosure is a handheld enclosure.

Claim	Limitation #	Limitation
23	23	The device according to claim 1, further comprising in the enclosure an actuator that is directly or indirectly affecting, changing, producing, or creating a physical phenomenon, the actuator that is coupled to the WLAN transceiver for being activated or controlled in response to data received from the WLAN.
24	24	The device according to claim 23, wherein the actuator consists of, or comprises, a motion actuator that causes linear or rotary motion.
25	25	The device according to claim 23, wherein the actuator consists of, or comprises, an electromagnetic coil or an electromagnet operative for generating a magnetic or electric field.
26	26	The device according to claim 23, wherein the actuator consists of, or comprises, an electrical signal generator.
27	27	The device according to claim 23, wherein the actuator consists of, or comprises, a chemical or an electrochemical actuator that is configured for producing, changing, or affecting a matter structure, properties, composition, process, or reactions.
28	28	The device according to claim 23, wherein the actuator consists of, or comprises, a thermoelectric actuator that is a heater or a cooler and is configured for affecting a temperature of a solid, a liquid, or a gas object, by conduction, convection, force convention, thermal radiation, or by transfer of energy by phase changes.

Claim	Limitation #	Limitation
29	29	The device according to claim 23, wherein the physical phenomenon comprises temperature, humidity, pressure, audio, vibration, light, motion, sound, proximity, flow rate, electrical voltage, or electrical current.

Petitioner Amazon.com Services LLC respectfully petitions for *inter partes* review of claims 1-2, 4-26, and 28-29 of U.S. Patent No. 12,401,720 (EX1001) per 35 U.S.C. §§311-319 and 37 C.F.R. §42.100 *et seq.*

I. INTRODUCTION

One billion smart devices are projected to be shipped by 2028. (EX1034.) The industry that manufactures and sells these devices drives domestic manufacturing and supports many U.S. jobs. Ensuring that this industry remains vital and unencumbered by assertions of unpatentable claims is critical to the U.S. economy. Smart Speaker has asserted such patents. The claimed invention of the '720 patent covers well-known and routine smart-device functionalities. The Director should institute review on the '720 patent to protect our robust patent system and a critical segment of our nation's economy.

Review is particularly appropriate here, because the Examiner did not conduct a thorough examination. The Examiner *never* issued *any* Office Action and his search was defective. (*Infra* §II.C.) Devices capable of receiving and responding to audio commands were prevalent. (*E.g.*, EX1011, EX1006, EX1005.) The recited network architecture simply reflected the conventional configuration of home wireless networks at the time. (*Id.*; EX1007.)

The challenged claims are unpatentable for the reasons set forth herein, as supported by Dr. Houh. (EX1002.)

II. THE '720 PATENT

A. Summary

The '720 patent has one independent claim and 28 dependent claims. The dependent claims span everything from standard networking protocols, to conventional microphones, to generic actuators, to routine lighting/audio components. (EX1001, claims 2-28.) The breadth of these conventional components is reflected in the number of grounds presented in the petition.

Like the claims, the patent's specification is a vague aggregation of off-the-shelf technologies at the time of the invention. "Actuators" could be speakers, "sounders," displays, coolers, heaters, lights, motors, pumps, etc. (*E.g.*, EX1001, 94:5-23; 96:12-97:31; 99:7-23, 100:10-101:8; 104:11-32; 105:51-106:55; 109:45-110:28.) Various network technologies were described to be "known in the art." (*E.g.*, EX1001, 2:34-38; 6:42-43; 6:66-7:6; 9:35-38.) The specification provides no technical details of how these well-known parts are integrated.

A paragraph in the Summary Section describes voice control over a network at a high level of generality. It describes "a router," "a first device in the building (or vehicle) . . . operative to transmit a digital data corresponding to the sensed voice to the router," and "[a] second device may be operative to execute actuator commands received from the router." (EX1001, 32:60-33:10.) Further, a "control server may be operative to produce actuator commands in response to [a] voice

processor output” (*Id.*, 33:10-14.) FIG. 19 depicts a high-level flowchart of “controller involving voice processing.” (*Id.*, 55:66-67, 134:38-135:30, 135:34-38.)

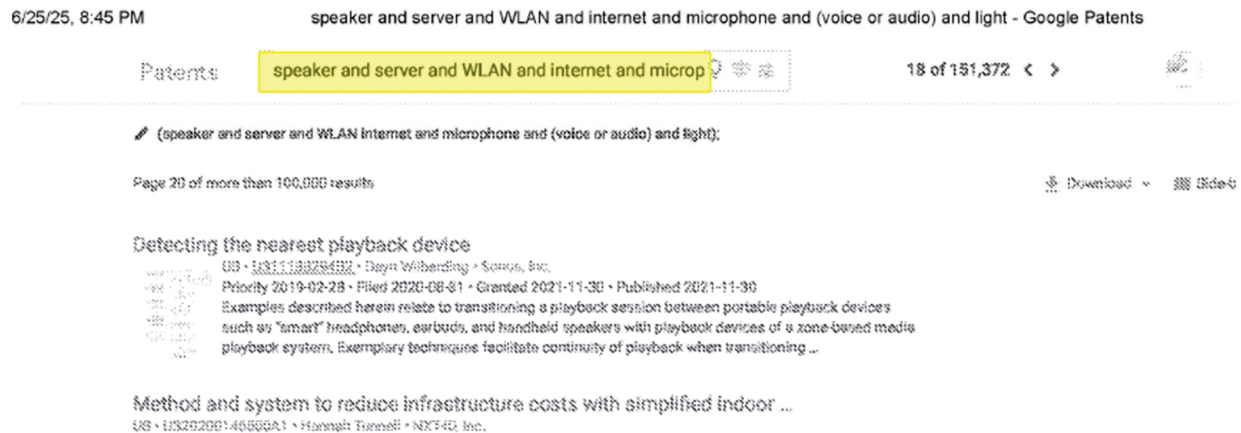
B. Effective Filing Date

The ’720 patent purportedly claims priority to a provisional application filed on January 9, 2012, which Petitioner treats as the earliest effective filing date only for purposes of this proceeding. (EX1002, ¶18.)

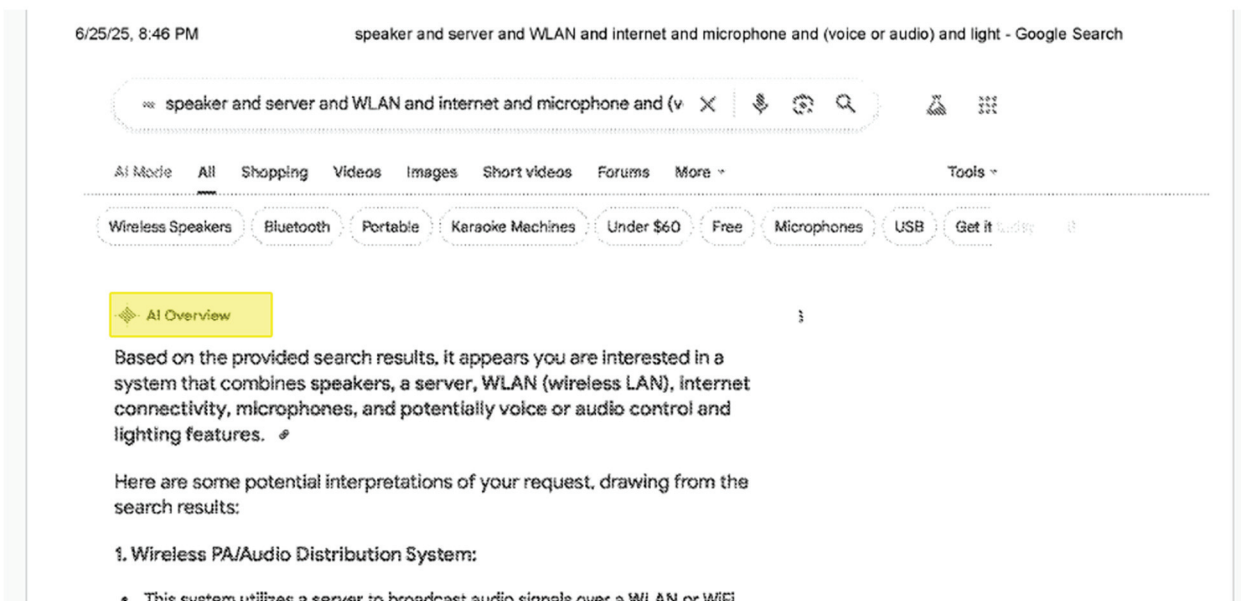
C. Prosecution History

The examination was not thorough. No Office Action was issued. (EX1003.) The Examiner allowed the claims within **2 months** of filing, citing the entirety of claim 1 as the reason for allowance. (*Id.*, 170-176.)

The Examiner performed a defective search on Google Patents, yielding over 100,000 results. His search was *not* limited by any priority date. (EX1003, 212-213.) Nor did it cover anything other than on conventional components: “**speaker** and **server** and **WLAN** and **internet** and **microphone** and (**voice** or **audio**) and **light**.” *Id.* His PE2E searches were also at a high level of generality. (*Id.*, 221-227.)



The Examiner also included Google’s “AI Overview” as part of his search strategy and results. The AI Overview was from the same flawed search string and does not constitute relevant searching. (EX1002, ¶¶36-41.)



D. Level of Ordinary Skill in the Art

A person of ordinary skill in the art (POSITA) in the field of the '720 patent would have a bachelor's degree in electrical engineering, computer engineering, or

equivalent, and two years of experience working on remote device control and/or networked devices. An advanced graduate degree may make up for the work experience indicated above. (EX1002, ¶¶47-49.) This Petition is supported by Dr. Houh, who meets these requirements. *Id.*

III. CLAIM CONSTRUCTION

Claim terms are presumed to carry their ordinary and customary meaning. The Board construes the terms only to the extent necessary to resolve the issue at hand: here, the invalidity of the claims under 35 U.S.C. §103. The Board need not construe any of the claims because, as set forth in the Petition, the claims are invalid under 35 U.S.C. §103 under any reasonable interpretation, and thus, no claim construction is necessary to resolve any issue before the Board. (EX1002, ¶50.)

IV. PRECISE REASONS FOR RELIEF REQUESTED

Pursuant to 37 C.F.R. §42.104(b), Petitioner requests cancellation of the challenged claims based on the following references (EX1002, ¶¶51-62). The Examiner never found, and the Applicant never cited any of these references during prosecution.

The following references are prior art *at least* under the following provisions.

Prior Art Reference	Abbreviation (Exhibit No.)	Prior Art Basis
U.S. Publication No. 2010/0088100	“Lindahl” (EX1011)	§102(a), (b), and (e)
U.S. Publication No. 2013/0132094	“Lim” (EX1005)	§102(e)
U.S. Publication No. 2012/0198339	“Williams” (EX1006)	§102(e)
U.S. Publication. No. 2008/0154608	“Evermann” (EX1012)	§102(a), (b), and (e)
Jim Doherty <i>et al.</i> , “Wireless Home Networking Simplified” (Cisco Press, 2007).	“Doherty” (EX1007)	§102(a) and (b)
“iPhone User Guide For iOS 5.0 Software” (Apple, 2011).	“iPhone User Guide” (EX1008)	§102(a)
Jacob Benesty <i>et al.</i> , “Springer Handbook of Speech Processing” (Springer-Verlag Berlin Heidelberg, 2008).	“Benesty” (EX1010)	§102(a) and (b)
Andrea Goldsmith, “Wireless Communications” (Cambridge University Press, 2005).	“Goldsmith” (EX1042)	§102(a) and (b)
Thomas D. Rossing <i>et al.</i> , “Springer Handbook of Acoustics” (Springer Science+Business Media, LLC, 2007).	“Rossing” (EX1014)	§102(a) and (b)

Prior Art Reference	Abbreviation (Exhibit No.)	Prior Art Basis
U.S. Patent No. 8,751,056	“Slaby” (EX1020)	§102(e)

The statutory grounds for the challenge of each claim are set forth below.

All of the statutory citations are pre-AIA.

Ground	35 U.S.C. §	Claims	Prior Art Reference(s)
1	103	1-2, 4-8, 11-26, 29	Lindahl and Williams
2	103	1-2, 23-26, 29	Lindahl, Williams, and iPhone User Guide
3	103	4-5	Lindahl, Williams, and Doherty
4	103	7-8, 11-12	Lindahl, Williams, and Rossing
5	103	9-10	Lindahl, Williams, and Benesty
6	103	20-21	Lindahl, Williams, and Goldsmith
7	103	14, 23, 28-29	Lindahl, Williams, and Slaby
8	103	1, 4-26, 29	Williams, Lim, and Benesty
9	103	2	Williams, Lim, Benesty, and iPhone User Guide
10	103	4-5	Williams, Lim, Benesty, and Doherty
11	103	7-8, 11-12	Williams, Lim, Benesty, and Rossing

Ground	35 U.S.C. §	Claims	Prior Art Reference(s)
12	103	20-21	Williams, Lim, Benesty, and Goldsmith
13	103	14, 23, 28-29	Williams, Lim, Benesty, and Slaby
14	103	1-2, 4-8, 11-26, 29	Lindahl and Evermann

Below, Petitioner discusses why the challenged claims are unpatentable under the statutory grounds raised, including how and where the prior art satisfies each element of each challenged claim, as required by 37 C.F.R. §42.104(b)(4).¹ Petitioner’s showing establishes that there is a reasonable likelihood that it will prevail on each ground of unpatentability to each challenged claim.

A. Ground 1: Lindahl and Williams Render Obvious Claims 1-2, 4-8, 11-26, and 29.

1. Summary of Lindahl

Lindahl was filed on October 2, 2008, and published on April 8, 2010, and is prior art under §102(a), (b), and (e). (EX1011, 1.) Lindahl is assigned to Apple Inc. “[D]evice 12 may receive a voice command to ‘*find more music like this*’” and “upload the voice command and relevant contextual information . . . to

¹ Petitioner also shows how each claim preamble is satisfied but does not concede any is limiting.

computing equipment 14.” (EX1011, [0042] (emphasis added).²) “The *computing equipment* may be a *server* (e.g., an internet server).” (*Id.*, [0028].) It “may receive and process the voice command” and “respond by displaying search results, purchase recommendations, etc.” (*Id.*; EX1002, ¶¶63-65.)

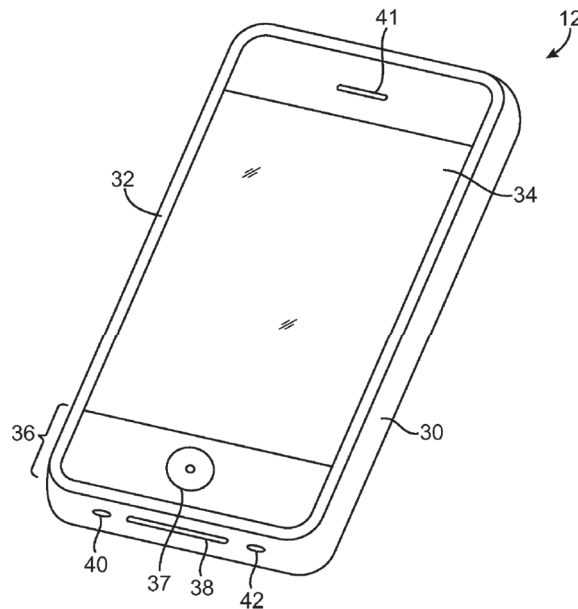


FIG. 2

Lindahl’s device includes the recited hardware elements in claim 1, including “*microphones*,” “*liquid-crystal display (LCD) screens* or other screens, *light-emitting diodes (LEDs)*,” and “*speakers and other devices for creating sound*,” and “*transceiver circuitry*.” (EX1011, [0048]-[0049], [0052]; EX1002, ¶¶66-70.)

² All emphasis and annotations are added unless otherwise stated.

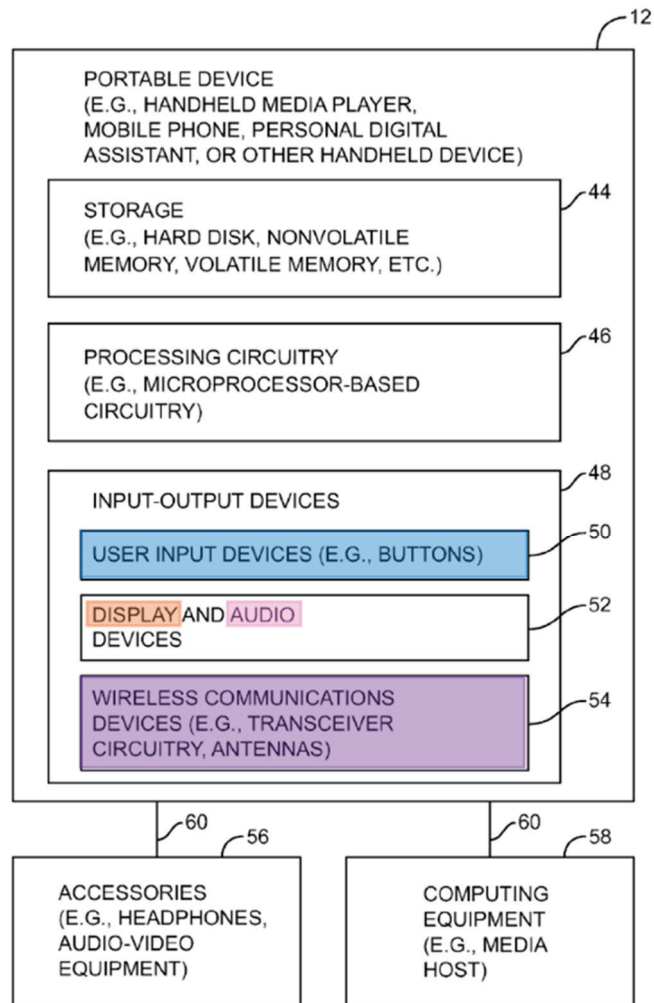


FIG. 3

2. Summary of Williams

Williams was filed on January 28, 2011, and is prior art under §102(e).

(EX1006, 1.) Williams is assigned to Amazon Technologies, Inc. In Williams,

“[a] user may ask a question, such as ‘*What’s the recipe for deviled eggs?*’” and

the system can “provide *audio answers* to spoken questions.” (EX1006, [0048].)

Referring to FIG. 1 below, “[a]n individual interface device 116 may include a microphone 202 that is configured to detect ambient noise, sounds, and *speech*.” (EX1006, [0023]; EX1002, ¶¶71-72.)

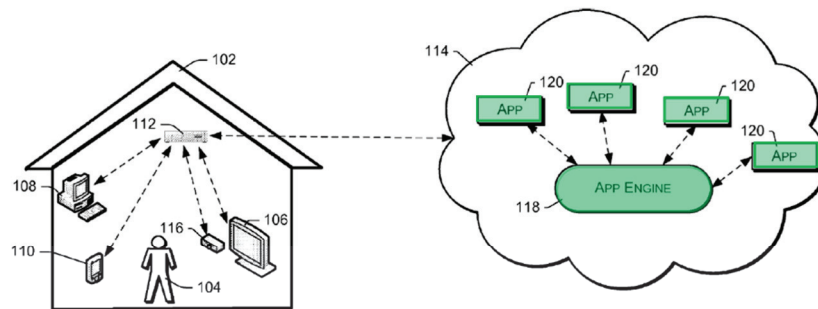


FIG. 1

Williams discloses “a *server* 400 that may be used to implement the functionality of the application engine 118 and/or any of the applications 120.” (EX1006, [0052].) “In operation, the application engine 118 [of the *server* 400] receives audio-based information from the interface devices 116.” (*Id.*, [0022].) The *server* can “provide *audio* 222 for eventual transmission to the interface device 116, to be played on the *speaker* 206.” (*Id.*, [0037]; EX1002, ¶¶73-74.)

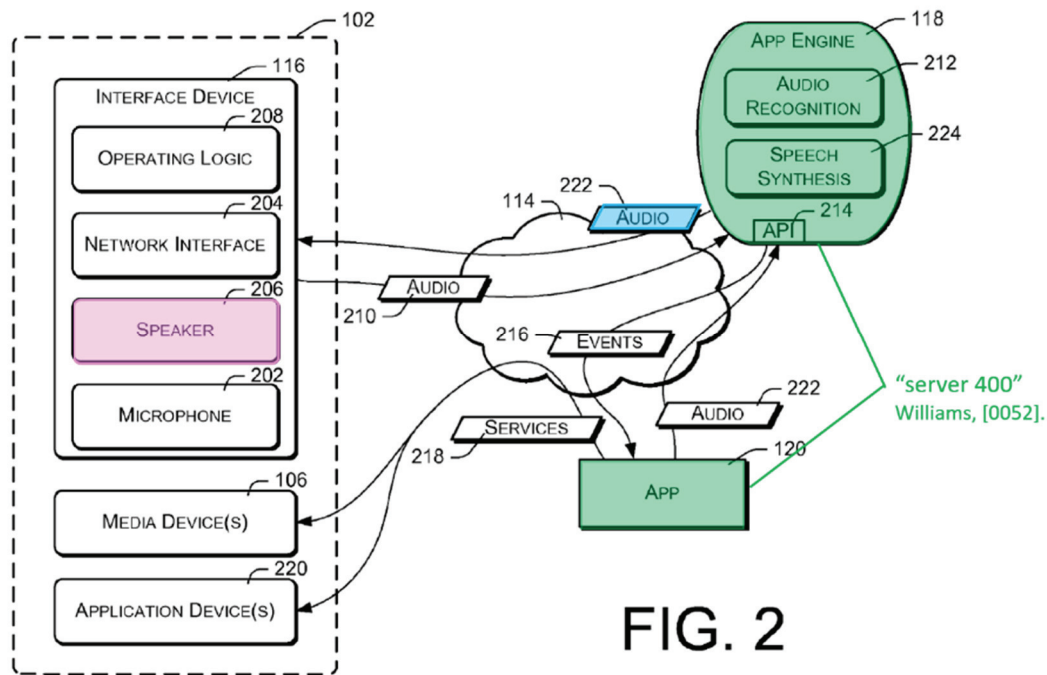


FIG. 2

3. A POSITA would have been motivated to combine Lindahl and Williams

It would have been obvious for a POSITA to combine Lindahl and Williams. Lindahl and Williams are analogous to the '720 patent and both relate to processing/responding to human speech commands via a remote server. (EX1011, [0073], [0043]; EX1006, [0027], [0036]; EX1001, 32:60-64; EX1002, ¶¶75-78.)

A POSITA would have been motivated to combine Lindahl and Williams, resulting in a system where Lindahl's server (*e.g.*, computing equipment 14/18) transmits a response in the form of audio data (*e.g.*, "audio 222") to the user device for playback as taught by Williams. This way, the device can respond to any voice command using an audio response (*e.g.*, speech, music via a speaker) as taught by Williams, a visual response (*e.g.*, via a display) as taught by Lindahl, or both

depending on the context. Multi-modal feedback improves the usability of the device. Transmitting audio (*e.g.*, rather than transmitting text to be converted into audio locally) would reduce or eliminate the need for local text-to-speech conversion at the device, thereby avoiding additional software or hardware resources and improving battery life. Lindahl expressly discusses the advantages of offloading intensive processing to the remote server, given the relatively limited processing capability of the local user device. (EX1011, [0035].)

While audio transmission may potentially require more bandwidth than text transmission, it would have been an acceptable tradeoff given its numerous advantages. Audio transmission does not demand significant bandwidth to begin with. Further, transmitting audio would improve user experience by ensuring the audio received is consistent (*e.g.*, in voice, tone, pronunciation, timing, effects) across every device even if they have different software and/or hardware configurations. (EX1002, ¶¶78-79.) Server-side audio generation is more sophisticated, resulting in higher-quality audio. *Id.* Moreover, transmitting audio provides better security, ensuring that the user hears exactly what was intended and not affected by user settings, malicious manipulation, or missing voices. *Id.*

Further, it would have been obvious to try combining the teachings of Lindahl and Williams. (EX1002, ¶80.) There are only a finite number of identified and predictable solutions for transmitting information to a device for

audio output, either sending audio directly or sending other forms of data (*e.g.*, text) to be converted into audio locally. *Id.* Transmitting audio directly represents a routine design choice between known alternatives. *Id.* The combination is straightforward and readily achievable by a person of ordinary skill in the art, as demonstrated by successful implementations in the field. *Id.* Williams demonstrates the feasibility and benefits of direct audio transmission, and a POSITA would have had a reasonable expectation of success in applying that approach to Lindahl’s architecture. *Id.*

4. Claim 1

a. 1[PR]: “A device comprising:”

Lindahl-Williams discloses a device. For example, Lindahl discloses “[a] *user device 12* [that] may be used to receive voice commands (*e.g.*, to record a user’s voice).” FIG. 2 shows “*user device 12*,” *e.g.*, a smartphone. (EX1011, [0025], [0027], [0038]; EX1002, ¶¶81-83.)

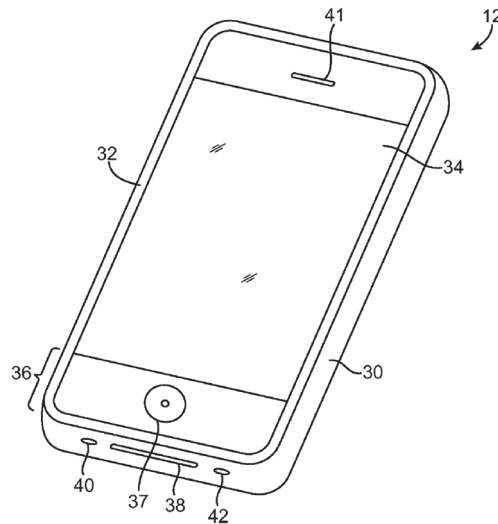


FIG. 2

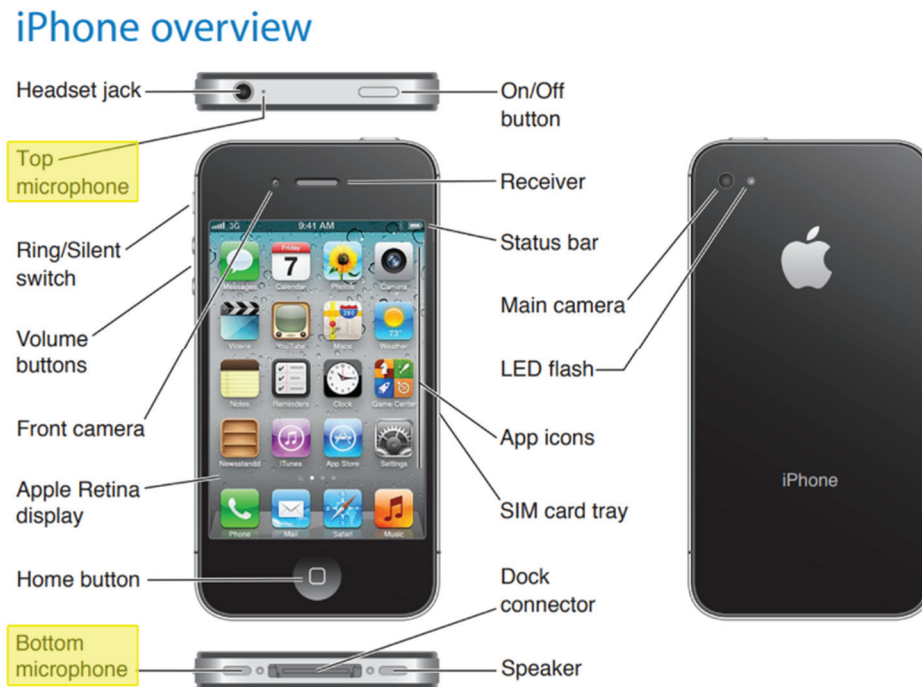
Williams also discloses this element. (*Infra* §IV.H.3.a; EX1002, ¶84.)

b. 1[A]: “multiple microphones for capturing first human voice data;”

Lindahl-Williams discloses or suggests *multiple microphones* for capturing *first human voice data*. Lindahl’s device 12 “may be used to capture *voice commands* from a user” and includes “user input devices 50 such as . . . *microphones*, cameras, etc.” (EX1011, [0034], [0023], [0048].) Lindahl discloses several exemplary voice commands such as “*find more music like this*,” “*identify this car*,” etc. (EX1011, [0042], [0013], [0063], [0067].) These voice commands are *first human voice data*. (EX1002, ¶¶85-88.)

It would have been obvious that Lindahl’s “*microphones*” capture the *human voice data*. Lindahl states that “[a] user can control the operation of user

device 12 by supplying *commands* through *user input devices 50*”, which includes “*microphones*”. (EX1011, [0048].) A POSITA would have understood that each of Lindahl’s microphones would capture voice command that is uttered near the device. (EX1002, ¶¶87-89.) It was common knowledge before 2012 that using multiple microphones improve audio capturing and reduce background noise, thereby enhancing speech recognition (*e.g.*, by Siri). (EX1010, 1021; EX1002, ¶¶90.) Consistent with Lindahl (assigned to Apple Inc.), many iPhones in/before 2012 had multiple microphones for capturing audio. (*Id.*; EX1008, 9 (“Top microphone” and “Bottom microphone”).)



Williams also discloses this element. (*Infra* §IV.H.3.b; EX1002, ¶91.)

- c. 1[B]: “a Wireless Local Area Network (WLAN) transceiver for transmitting digital data to, and for receiving digital data from, a WLAN;”

Lindahl-Williams discloses a *WLAN transceiver* for transmitting and receiving digital data to and from a WLAN.

Referring to FIG. 3 below, Lindahl’s device 12 comprises “[w]ireless communications devices 54,” which has “radio-frequency (RF) *transceiver circuitry*” and “may be used to support local and *remote wireless links*. Examples of local wireless links include . . . *Wi-Fi* ® (*IEEE 802.11*).” (EX1011, [0049], [0052].) In other words, Lindahl’s “*transceiver circuitry*” includes a transceiver for supporting Wi-Fi. (EX1002, ¶¶92-94.) It was common knowledge that WiFi is a type of WLAN. (EX1042, 600-601; EX1013, 7:48-51; EX1002, ¶¶94-95.) The ’720 patent confirms the same. (*Id.*; EX1001, 8:25-32.) Thus, Lindahl’s WiFi transceiver discloses a *WLAN transceiver*. (EX1002, ¶94.)

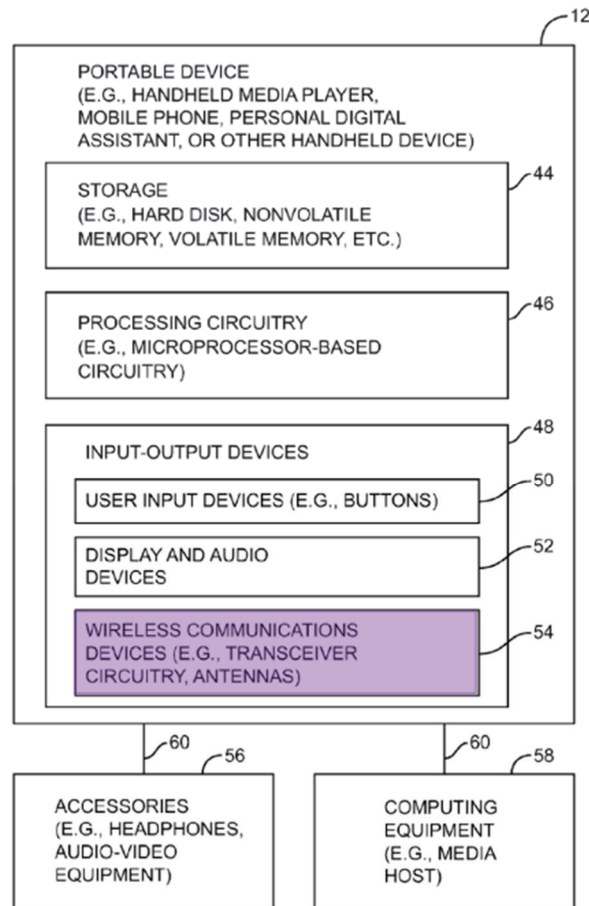


FIG. 3

It was common knowledge that a “WLAN” transceiver would be for transmitting digital data to, and for receiving digital data from a “WLAN” as recited in the claim. (EX1035, 3:17-20, 3:38-40; EX1036, [0013]-[0015]; EX1002, ¶96.) Lindahl confirms this, stating that “[u]ser device 12 may handle communications over one or more wireless communications bands such as local area network bands” and “based on wireless communications technology such as . . . Wi-Fi®.” (*Id.*; EX1011, [0038], [0027], [0044], [0030]-[0031].) And as discussed below, Lindahl’s device transmits digital data (*e.g.*, voice commands) to

and receives digital data (*e.g.*, information) from a remote server over a WLAN.

(EX1002, ¶¶97-98; *infra* §§IV.A.4.h-IV.A.4.j.)

Williams also discloses this element. (*Infra* §IV.H.3.c; EX1002, ¶99.)

d. 1[C]: “a speaker for sounding audio data;”

Lindahl-Williams discloses this element. In FIG. 2, Lindahl’s device 12 includes “[p]ort 41 [which] may be, for example, a *speaker* port.” (EX1011, [0040], FIG. 2.) For example, by 2012, iPhones were equipped with an earpiece receiver speaker (*e.g.*, positioned toward the top part of the device). (EX1008, 9, 29, 148; EX1002, ¶100.)

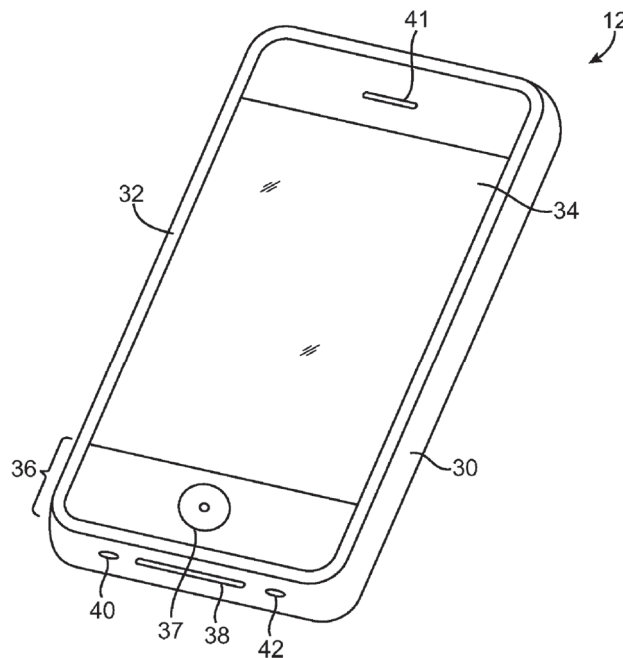


FIG. 2

All speakers are for sounding audio data. (EX1002, ¶¶101-102; *e.g.*, EX1001, 105:64-106:7.) Lindahl’s speaker does this. Referring to FIG. 3, Lindahl’s device 12 comprises “*speakers* . . . for *creating sound*.” (EX1011, [0048].)

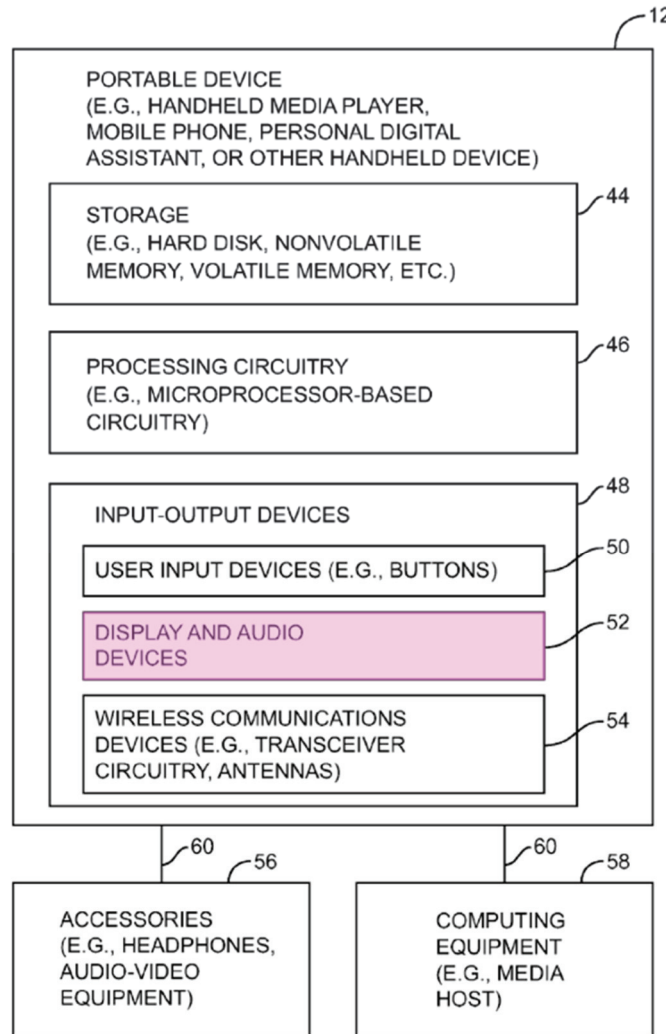


FIG. 3

Williams also discloses this element. (*Infra* §IV.H.3.d; EX1002, ¶103.)

e. 1[D]: “an electric light source for emitting light for emitting light;”

The '720 patent states, an “electric light source may be based on Solid-State Lighting (SSL) such as a *Light Emitting Diode (LED)*[.]” (EX1001, 39:19-26.) Further, “a display . . . may consist of *an array (e.g., matrix) of light emitters* or light reflectors[.]” (EX1001, 25:33-36.) “An *LED-backlit LCD* is a flat panel display that uses *LED backlighting* [.]” (EX1001, 110:64-111:3, 111:11-23; EX1002, ¶¶104-105.)

Lindahl-Williams discloses this element. Consistent with the patent, Lindahl’s device 12 includes “[d]isplay and audio devices 52 [which] may include *liquid crystal display (LCD) screens* [display 34] or other screens, *light-emitting diodes (LEDs)*, and other components that present visual information and status data,” as shown in FIG. 3. (EX1011, [0048], [0039]; EX1002, ¶106.)

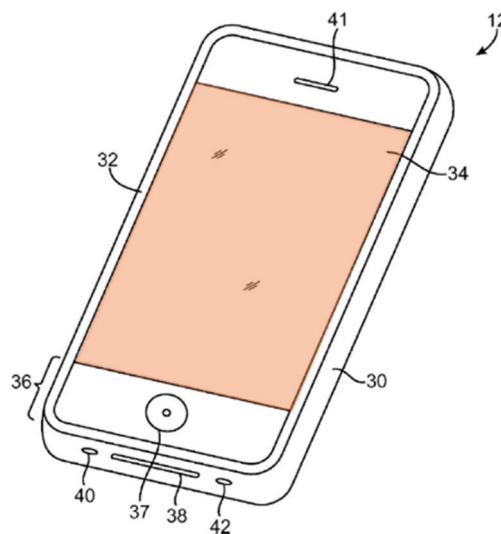


FIG. 2

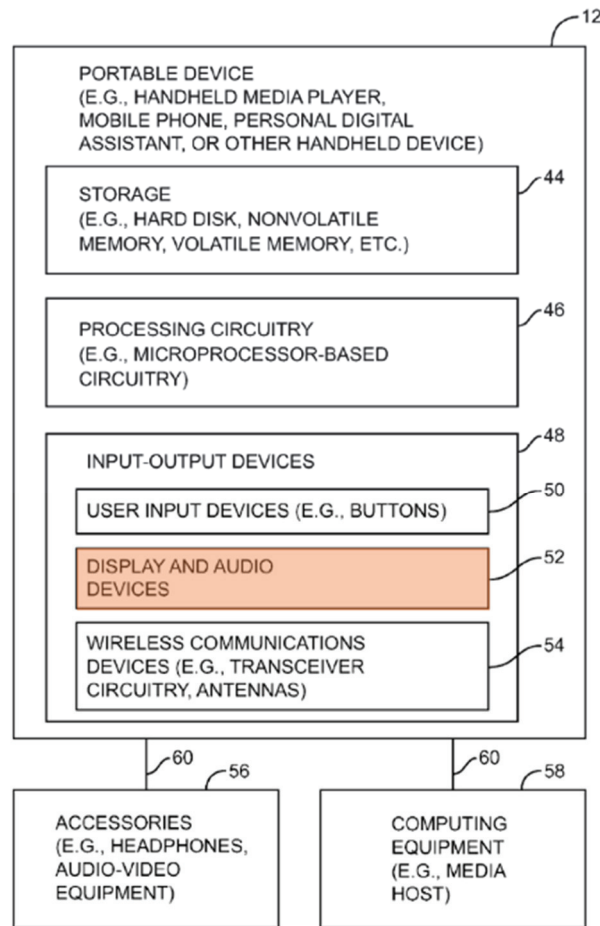


FIG. 3

- f. **1[E]: “an enclosure that houses the multiple microphones, the WLAN transceiver, the speaker, and the electric light source;”**

Lindahl-Williams discloses this element. For example, as shown in FIGs. 1 and 2, Lindahl’s device 12 houses the WLAN transceiver (e.g., *wireless communications devices 54*), speaker (e.g., *speaker port 41 or speakers in “display and audio devices 52”*), microphones (e.g., *microphones in “display and audio devices 52”*), and electric light source (e.g., *display 34, LCD screens or*

other screens, or LEDs in “display and audio devices 52”) within an enclosure.

(EX1011, [0039]-[0040], [0048]-[0049], FIGs. 1-2; EX1002, ¶¶107-108.)

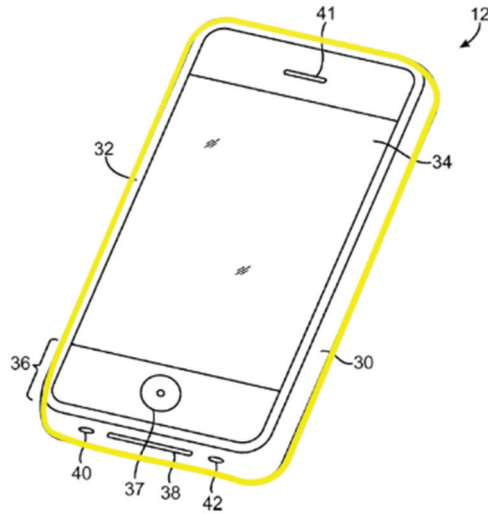


FIG. 2

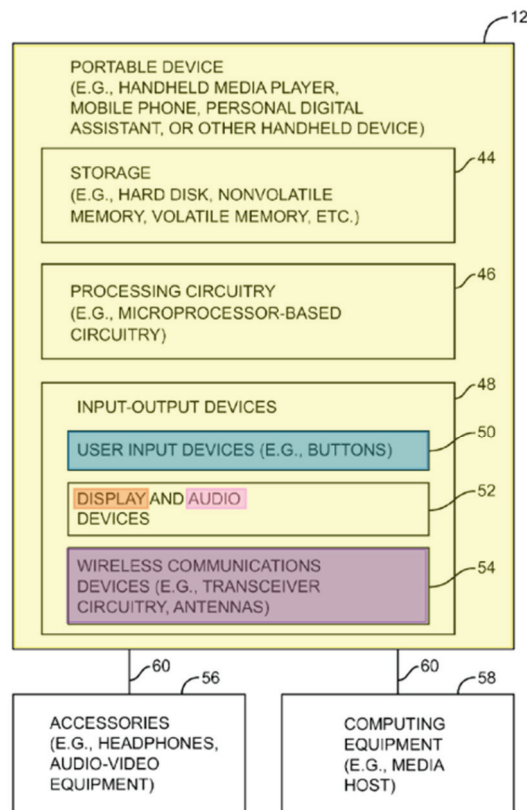


FIG. 3

Williams also discloses this element. (*Infra* §IV.H.3.f; EX1002, ¶109.)

- g. 1[F]: “and one or more processors programmed with computer program instructions that, when executed, cause the device to:”**

Lindhahl-Williams discloses this element. For example, referring to FIG. 3, Lindahl’s device 12 comprises processing circuitry 46 “used to run software.” (EX1011, [0046].) “Processing circuitry 46 may be used to control the operation of user device 12,” including the recited operations. (*Id.*; *infra* §§IV.A.4.h-IV.A.4.j; EX1002, ¶110.)

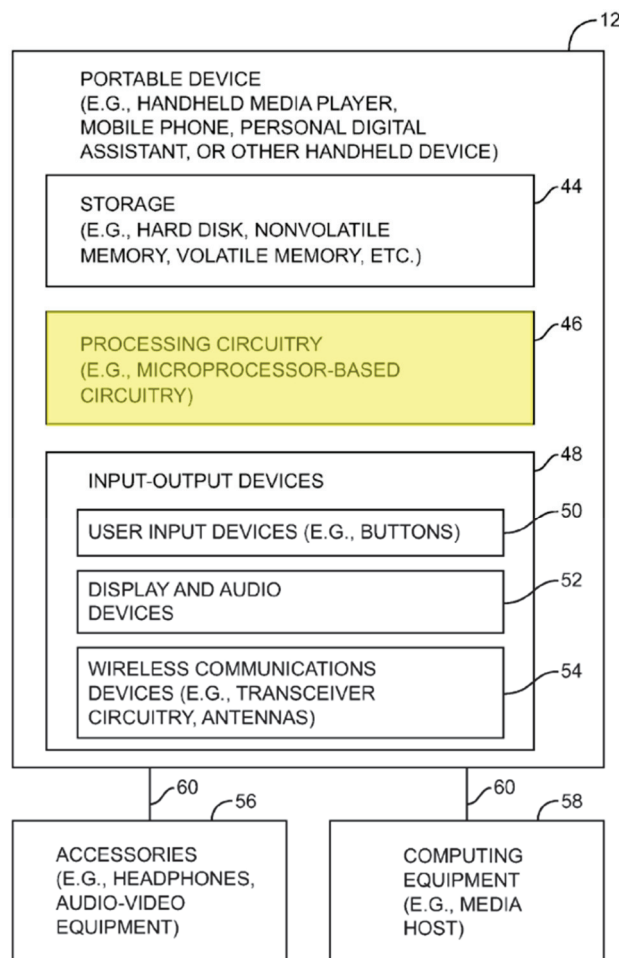


FIG. 3

Williams also discloses this element. (*Infra* §IV.H.3.g; EX1002, ¶111.)

- h. 1[G]: “send, to a server device over the Internet via the WLAN by the WLAN transceiver, at least part of the first human voice data captured by the multiple microphones;”**

Lindahl-Williams discloses this. First, Lindahl discloses a *server device*.

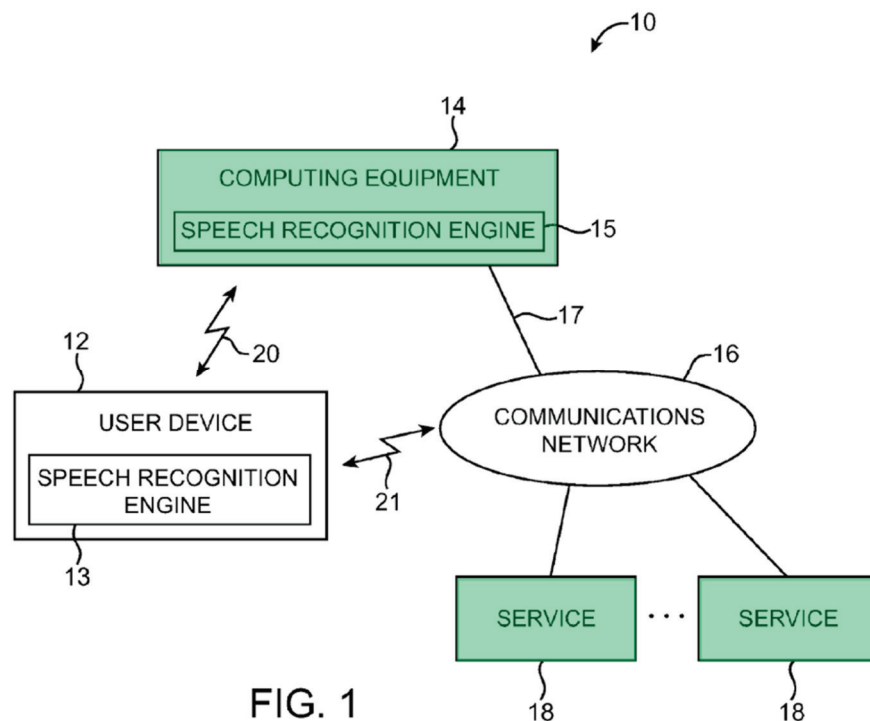
Lindahl discloses a “[c]omputing equipment 14 [which] may include . . . a *server*.”

(EX1011, [0028].) “The computing equipment may be a *server (e.g., an internet*

server).” *Id.* As another example, Lindahl discloses a “[c]omputing equipment 18

(e.g., *one or more servers*) [which] may be associated with one or more online

services.” (EX1011, [0029]; EX1002, ¶112.)



Lindhahl explains that server processing of voice commands is advantageous as “*computer servers* [] generally have relatively large amount of processing and storage capabilities compared to portable devices such as user device 12” and thus “the probability that a given voice command can be successfully processed and fulfilled will generally increase.” (EX1011, [0082], [0012]; EX1002, ¶¶118-120.)

Lindhahl discloses sending at least part of the captured *first human voice data* to the *server*. As shown in FIG. 6, at step 74, “user device 12 may record a *voice command*.” (EX1011, [0065].) At step 78, “[u]ser device 12 may upload recorded *voice commands* and stored contextual information to *computing equipment 14* or *equipment 18* using any suitable communications path.” (EX1011, [0068], [0070]; EX1002, ¶¶113-116.)

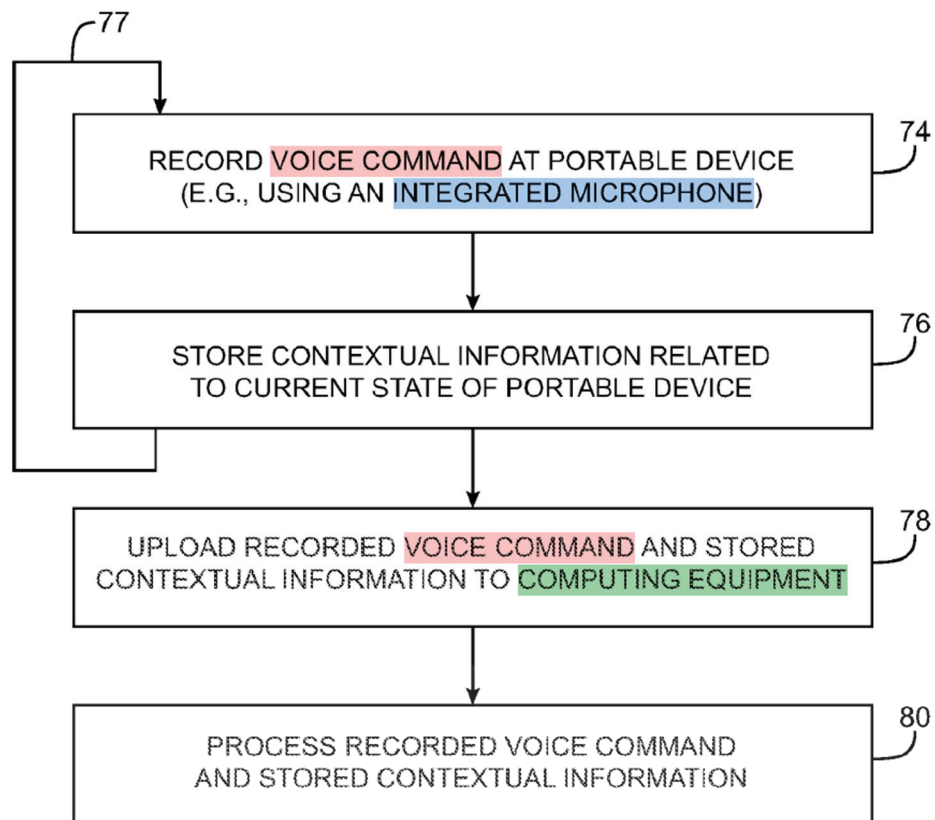
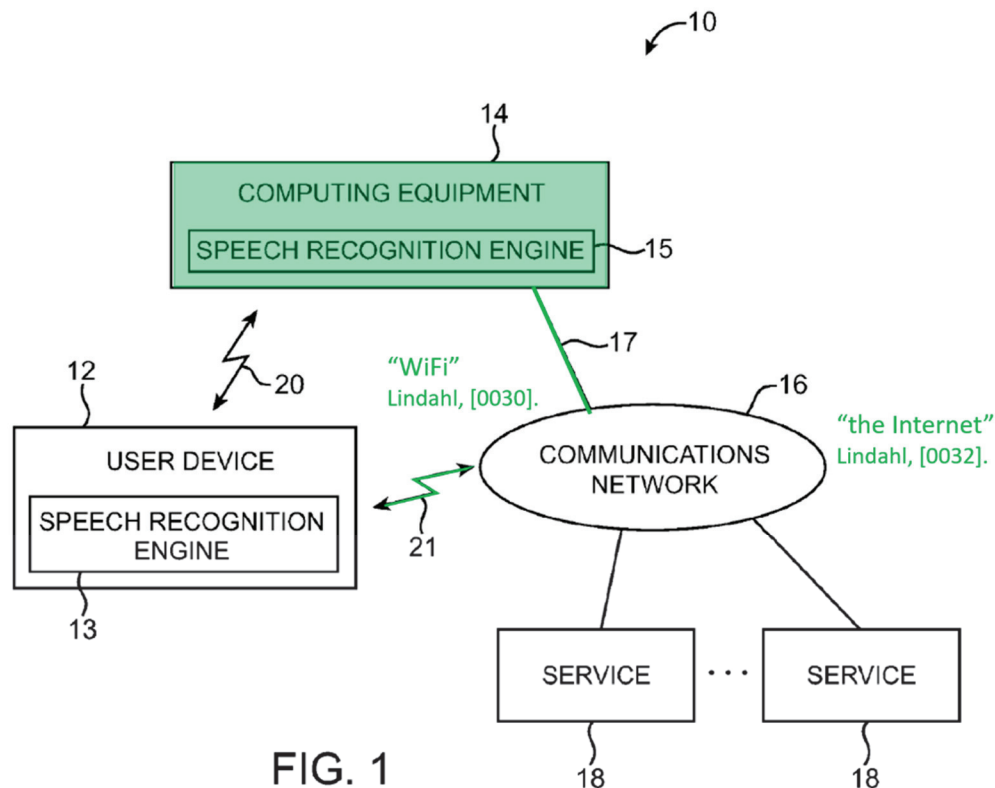


FIG. 6

It would have been obvious that the “*voice command*” would be at least part of the first human voice data captured by Lindahl’s “*microphones*.” (*Supra* §IV.A.4.b) For example, Lindahl’s multiple microphones would simultaneously capture the “*voice command*.” Alternatively, at least one of Lindahl’s multiple microphones captures the “*voice command*,” which is part of all audio data captured by the multiple microphones (“first human voice data”). (EX1002, ¶117.)

Lindhahl discloses sending the voice command to a server device over the Internet via the WLAN by the WLAN transceiver. Referring to FIG. 1 below, Lindahl's *voice commands* are transmitted to "*equipment 14* . . . through *communications network 16 over paths 17 and 21*, or may upload them to *equipment 18* over *network 16*." (EX1011, [0068].) Lindahl discloses that "network 16 may be based on . . . networks such as . . . *the Internet*" and "communication paths in system 10" (e.g., path 20, path 17, and 21) may be based on "*Wi-Fi®*." (EX1011, [0030], [0032]; EX1002, ¶¶121-124.) It would be obvious to a POSITA that the transmission of voice commands over the Internet via the WLAN is by the *WLAN transceiver*. (EX1011, [0052]; EX1002, ¶125.)



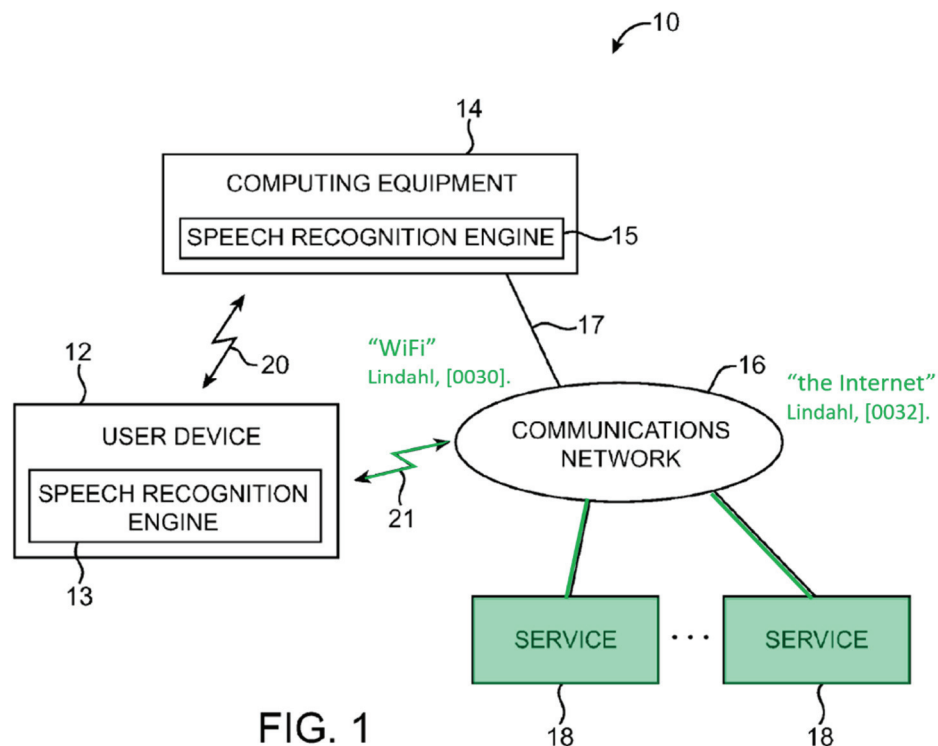


FIG. 1

Williams also discloses this element. (*Infra* §IV.H.3.h; EX1002, ¶126.)

- i. **1[H]: “play, by the speaker, an audio data received from the server device over the Internet via the WLAN by the WLAN transceiver;”**

Lindahl-Williams discloses playing, by the *speaker*, an *audio data* received from the *server* device *over the Internet via the WLAN* by the *WLAN transceiver*.

First, Lindahl discloses or at least suggests receiving *audio data* from the *server* device. After sending voice commands to the server device (*i.e.*, computer equipment) for processing, “*computing equipment 14* may transmit *data signals* to user device 12.” (EX1011, [0031].) “*Equipment 14* may, for example, transmit *information* to device 12 in response to voice commands transmitted by device 12

to system 14.” (EX1011, [0031], [0009], Abstract.) It would have been obvious to a POSITA that such “information” would include audio data. (EX1002, ¶¶127-129.)

Lindahl discloses that such transmission of *audio data* by the *server device* is *over the Internet via the WLAN*. As discussed, the sending of voice commands to the server is over the Internet via the WLAN, it would be obvious that the receipt of audio data from the server is through the same communication path. (EX1002, ¶¶130-131.) Lindahl confirms this, stating that it is “particularly beneficial” for “*computing equipment 14 or 18* [to] transmit *information related to processing and responding* to the voice command to user device 12” when the devices are connected “through long-range communications paths . . . such as the *Internet*.” (EX1011, [0071]-[0072], [0033].) It would have been obvious to a POSITA that the receipt of audio data is by the *WLAN transceiver*. (*Supra* §IV.A.4.c; Lindahl, [0052]; EX1002, ¶131.)

To the extent that Lindahl does not expressly disclose receiving audio data and playing it by the speaker (which is not conceded), Williams discloses this. (EX1002, ¶132.) For example, Williams’ device receives audio data from the *server* via the WLAN over the Internet by the WLAN transceiver. Referring to FIG. 2, Williams’ *cloud server* “processes the *audio-based information* to determine various data about user activities, status, environment, commands, etc.”

and “provide[s] **audio 222** for eventual transmission to the interface device 116.” (EX1006, [0022], [0037], [0048], [0045].) Williams discloses playing, by the speaker, the received audio data. Specifically, the cloud server “provide[s] **audio 222** for eventual transmission to the interface device 116, *to be played on the speaker 206.*” (EX1006, [0037], [0045], [0048]; EX1002, ¶133.) Williams discloses several examples of interface device 116 providing an **audio response** generated by the **cloud server**. (EX1006, [0045], [0048], [0050]; EX1002, ¶134.) For example, it can “provide **audio answers to spoken questions**” or “**query the user regarding movie selections and times.**” (EX1006, [0048], [0045].)

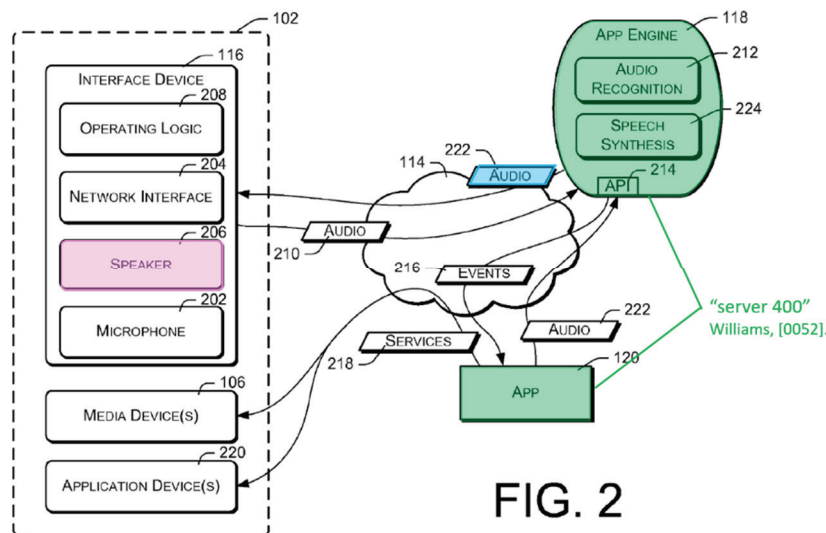


FIG. 2

As shown in FIGs. 1 and 2, the **audio 222** is sent by the **cloud server** over the network 114 (i.e., “the Internet”) via the home WLAN (via router 112). (*Supra* §§IV.A.4.c, IV.A.4.h; EX1006, FIG. 1, [0015], [0037], [0052].) Alternatively, the **wireless network 114** itself includes both WLAN and the Internet and thus

discloses the device receiving audio data “over the Internet via the WLAN”.

(*Supra* §§IV.A.4.c, IV.A.4.h; EX1006, [0015].) It would be obvious that the audio data is sent by the *WLAN transceiver*. (*Supra* §IV.A.4.c; EX1006, [0023]; EX1002, ¶¶135-137.)

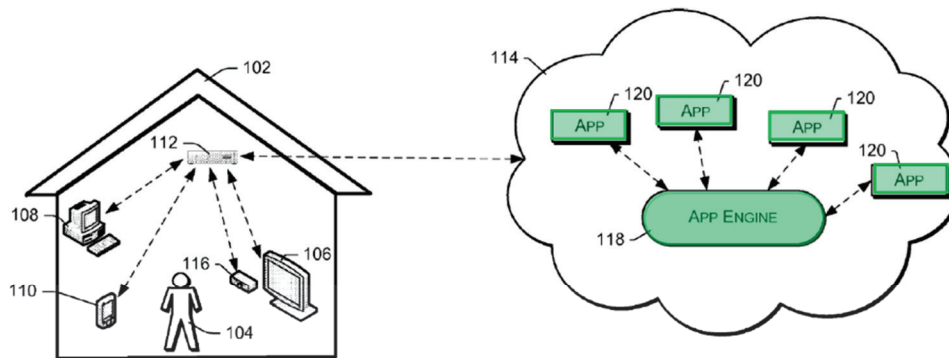


FIG. 1

A POSITA would have been motivated to combine Lindahl and Williams, resulting in a system where Lindahl’s server (*e.g.*, computing equipment 14/18) transmits a response in the form of audio data (*e.g.*, “audio 222”) to the user device for playback as taught by Williams. (*Supra* §IV.A.3.) Multi-modal feedback improves the usability of the device. (EX1002, ¶138.)

- j. 1[I]: “and emit³, by the electric light source, in response to data received from the server over the Internet via the WLAN by the WLAN transceiver,”

Lindahl discloses emit[ting] by the *electric light source*, in response to *data* received from the *server*. (*Supra* §IV.A.4.i; EX1011, [0040], [0048].) For example, “[t]he *data* received by device 12 from *equipment 14* in response to voice commands and contextual information may be used by device 12 to carry out requests associated with the voice commands.” (EX1011, [0043].) “[A]fter processing the voice command and contextual information, *computing equipment 14* may transmit *results associated with the voice command* to user device 12 which may then *display* the *results*.” (EX1011, [0043]; EX1002, ¶139-141.) A POSITA would have understood displaying the results/data involves emitting light. (EX1002, ¶142; EX1011, [0048].)

Lindahl’s device 12 receives *data* from the *server* device (*e.g., computing equipment 14 or 18*) over the Internet via the WLAN by the *WLAN transceiver*. (*Supra* §§IV.A.4.c, IV.A.4.i; EX1011, Abstract, [0009], [0072], [0031], [0030], [0052]; EX1002, ¶140.)

³ The limitation appears to be missing the word “light” after “emit[.]”

- k. **1[J]: “wherein the device is addressed in the WLAN using an Internet Protocol (IP) address that uniquely identifies the device in the WLAN or in the Internet.”**

Lindhahl-Williams discloses this. As discussed, Lindahl’s device 12 comprises a WLAN transceiver and communicates with the remote server over the Internet via the WLAN. (*Supra* §§IV.A.4.c, IV.A.4.h.) A POSITA would have understood, and considered obvious, that this involves associating Lindahl’s device with an IP address that uniquely identifies it in the WLAN. (EX1002, ¶¶143-144.) The IP address may be an assigned IP address for the device (*e.g.*, by the router), the router’s IP address, a combination of the two, or a globally unique IP address. (*Id.*; EX1007, 99, 108, 175, 210, 340.) These IP addresses uniquely identify the device within the user’s WLAN, and further these devices would also be addressed using the router’s globally unique IP address when communicating with Internet servers using the Network Address Translation (“NAT”) feature of the router to communicate with devices outside of the user’s WLAN. *Id.*

Williams also discloses this element. (*Infra* §IV.H.3.k; EX1002, ¶145.)

5. Claim 2

Lindhahl’s device is a mobile phone, as shown in FIG. 2. (*Supra* §IV.A.4.a; EX1011, [0027], [0044], FIG. 2.) Lindahl discloses that it includes “**battery-based** static or dynamic random-access-memory.” (EX1011, [0045].) Thus, a POSITA

would understand that Lindahl's handheld mobile phone would include a rechargeable battery within its enclosure for usability. (EX1002, ¶¶146-147.)

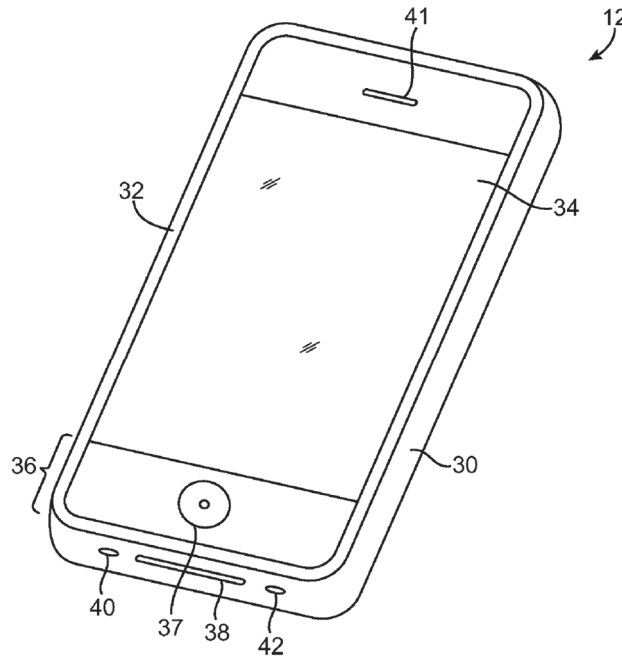


FIG. 2

6. Claims 4/5/6

Regarding claim 4, Lindahl's device is addressed in the WLAN using IP addresses. (*Supra* §§ IV.A.4.c, IV.A.4.k) Regarding claim 4, it was well-known, and a POSITA would find it obvious, that IP address would be a distinct locally administered address addressable in the WLAN or a universally administered digital address addressable in the WLAN and in the Internet. (*Id.*; EX1002, ¶¶148-149; EX1007, 99, 108, 175, 210, 340.) Further, a POSITA would understand (and obvious) that Lindahl's device has one or more MAC addresses for networking,

and these are universally administered digital addresses addressing the device both in the WLAN and in the Internet. (EX1007, 99, 166; EX1023, 24, 218, 312; EX1002, ¶¶150-151.)

Regarding claim 5, it was well-known (and obvious) that IP addresses are layer 3, either static or dynamic, and of IPv4 type. (EX1029, 127; EX1007, 277; EX1023, 11, 13; EX1024, 2:60-62; EX1002, ¶¶152-153.)

Regarding claim 6, it was common knowledge that a networked device locally stores the IP address. (EX1007, 211, Figure 10-44, 265.) Lindahl's device includes "nonvolatile memory" and "volatile memory." (EX1011, [0045].) A POSITA would understand (and obvious) that Lindahl's "nonvolatile and volatile memory" would locally store the IP address to achieve wireless connectivity. (EX1002, ¶¶154-157; EX1011, [0045], [0060]; EX1009, [0218]-[0225]; *e.g.*, EX1007, 211, Figure 10-44, 265.)

7. Claims 7/8/11/12

A POSITA would have understood that these claims recite standard acoustic devices (*e.g.*, microphone, speaker), and the patent confirms this. (EX1001, Claim 9, 17:47-54, 105:48-106:62; EX1002, ¶¶158-161.) Lindahl's device comprise standard microphones and speakers. (EX1011, [0006], [0040], [0048], FIGs 2-3; *Supra* §§IV.A.4.b, IV.A.4.d, IV.A.4.i.) A POSITA would understand that each of

Lindahl's microphones and speakers operate as claimed. (EX1014, 757-758, 309, 763, 765; EX1018, 122-123; EX1002, ¶¶158-161.)

8. Claims 13/14/15

Lindahl-Williams discloses these claims. Lindahl's user device 12 comprises "*liquid crystal display (LCD) screens* or other screens, *light-emitting diodes (LEDs)*, and other components that present visual information and status data." (*Supra* §§IV.A.4.e, IV.A.4.j; EX1011, [0048].) A POSITA would understand display screens and LEDs are or include electric light sources emitting visible light for illumination or indication and the '720 patent confirms the same. (EX1001, 89:50-57, 109:45-50; 110:64-111:3, 111:11-23; EX1002, ¶¶162-163, 165.)

For claim 14, a POSITA would understand that LEDs and LCD displays also emit non-visible light, *e.g.*, infrared radiation. (EX1030, 126; EX1002, ¶164.)

9. Claims 16/17/18/19

Lindahl-Williams discloses these claims. Regarding claim 16, Lindahl's device has "music player functionality" and may perform a "media playback operation" to play "a music file or a video file". (EX1011, [0024], [0027], [0034], [0061], [0042].) A POSITA would understand (and it would be obvious) that Lindahl's device stores the media files in its local memory and Lindahl's "processing circuitry 42" are programmed with instructions that plays the media

files on its speaker when executed. (EX1011, [0046]; EX1008, 60-61; EX1002, ¶¶166-167.)

Regarding claims 17-19, Lindahl’s “music file[s] or video file[s]” are pre-recorded audio content. (EX1011, [0034].) Lindahl’s device comprising “media players” and “music player” plays music, which a POSITA would understand to comprise a syllable, a word, a phrase, a sentence using male or female voice. (EX1011, [0024], [0027], [0034], [0042]; EX1002, ¶¶168, 170, 172.)

Additionally or alternatively, Williams discloses that “audio [222] may be in the form of text, in which *speech synthesis or text-to-speech conversion* may be performed by the application engine 118.” (EX1006, [0037].) Williams’ speaker “us[es] the *speech synthesis functionality* of the system to *query the user regarding movie selections and times*”. (EX1002, ¶¶167, 169, 171, 173; EX1006, [0037], [0045]-[0046] and [0050].) Thus, a POSITA would find it obvious that Lindahl-Williams, which receives audio 222, would play synthesized digital audio content, simulates the voice of a human being, and sounds syllables/words/phrases using human voice to query the user. (*Supra* §IV.A.4.i; EX1002, ¶¶166-173.)

10. Claims 20/21

Lindahl-Williams discloses or suggests these claims. For example, Lindahl discloses its device communicates with the remote server over the *Internet* (*i.e.*, network 16) and *via WiFi* (*i.e.*, path 17, 20 or 21). (*Supra* §§IV.A.4.c, IV.A.4.h;

EX1011, [0025], [0032], [0026], [0030].) Consistent with the '720 patent, a POSITA would understand that WiFi is a form of WLAN based on the IEEE 802.11 standard group comprising 802.11a/b/g standard. (*Supra* §IV.A.4.c; EX1042, 600-601; EX1013, 7:48-51; EX1001, 8:25-32; EX1002, ¶¶174-175.) It was well-known that 802.11b WLANs are ISM Band II frequencies. (EX1042, 22-23; EX1002, ¶176.) A POSITA would have found it obvious that Lindahl-Williams implements this for wireless networking. *Id.*

11. Claim 22

Lindahl's "user device 12 may be provided in the form of a *handheld device*." (EX1011, [0027], [0044], [0023]; EX1002, ¶177.)

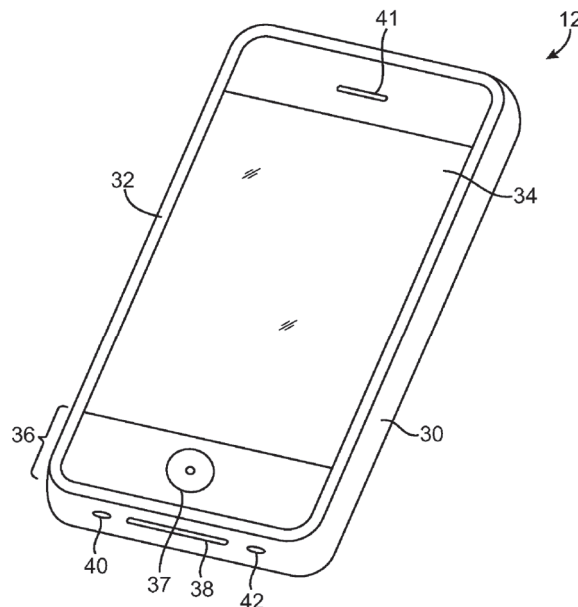


FIG. 2

12. Claims 23/24/25/26/29

Lindhahl discloses multiple speakers. Lindahl's user device 12 comprises "[d]isplay and audio devices 52 . . . such as *speakers and other devices for creating sound.*" (EX1011, [0048].) A POSITA would understand that each speaker or a speaker system is an actuator (*e.g.*, operating using the moving coil principle). (EX1018, 122-123; EX1014, 763; EX1002, ¶¶178-182.) The '720 patent confirms that speakers are actuators. (EX1001, 105:51-106:7.) It would have been obvious that each speaker/speaker system is coupled to the WLAN transceiver and controlled in response to data received from the WLAN in Lindahl-Williams. (*Supra* §IV.A.4.i; EX1002, ¶183.)

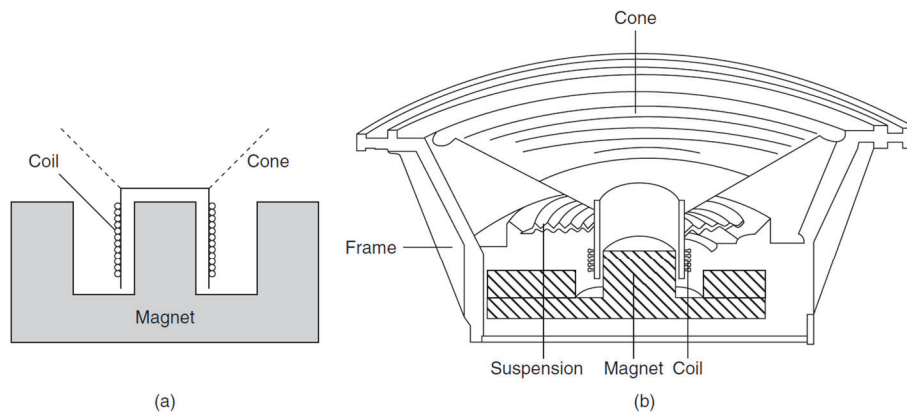


Figure 10.8 Moving-coil loudspeaker: (a) cross-section of coil and magnet, and (b) cutaway view of complete unit

Regarding claims 24/25/29, it was well-known that a speaker is a motion actuator that causes “linear motion” (EX1014, 903) and the “dynamic or moving-coil loudspeaker” has an electromagnetic coil creating a magnetic or electric field that can “move axially.” (EX1014, 18; EX1018, 122-123.) The '720 patent

confirms the same. (EX1001, 105:64-106:7; EX1002, ¶¶184-187, 191.)

Speakers/speaker systems are configured to directly or indirectly affect, change, produce, or create audio/sound and electrical current. *Id.*

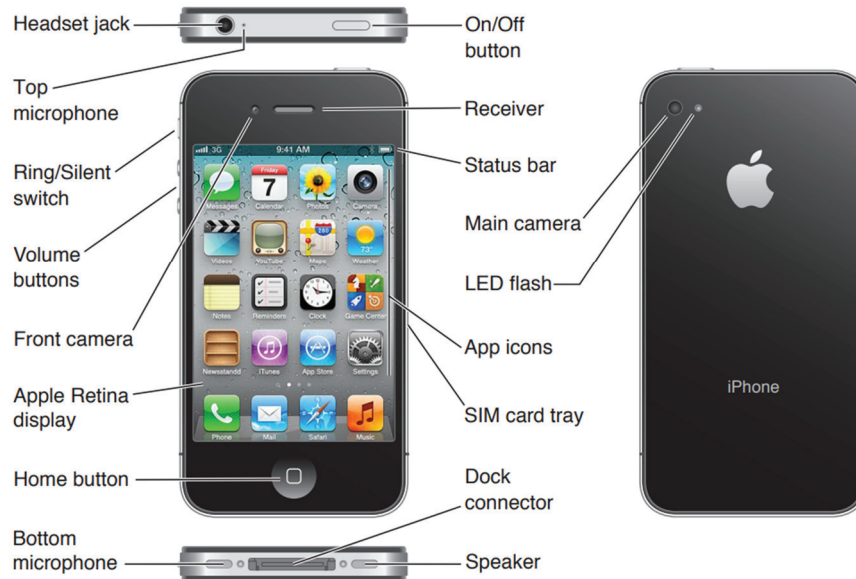
Regarding claim 26, it was well-known that a speaker system comprises a digital-to-analog converter (“DAC”) for converting received digital audio into analog signals for playback. (EX1014, 714, 522; EX1018, 170; EX1002, ¶¶188-190.) A DAC is an actuator generating electrical signals, consistent with the ’720 patent. (*Id.*; EX1001, 34:25-28, 107:13-20.)

**B. Ground 2: Lindahl, Williams and iPhone User Guide
Render Obvious Claims 1-2, 23-26, 29**

1. Summary of iPhone User Guide

iPhone User Guide was published/made publicly available at least by November 7, 2011, and is prior art under §102(a). (EX1048; EX1002, ¶¶192-196.)

iPhone overview

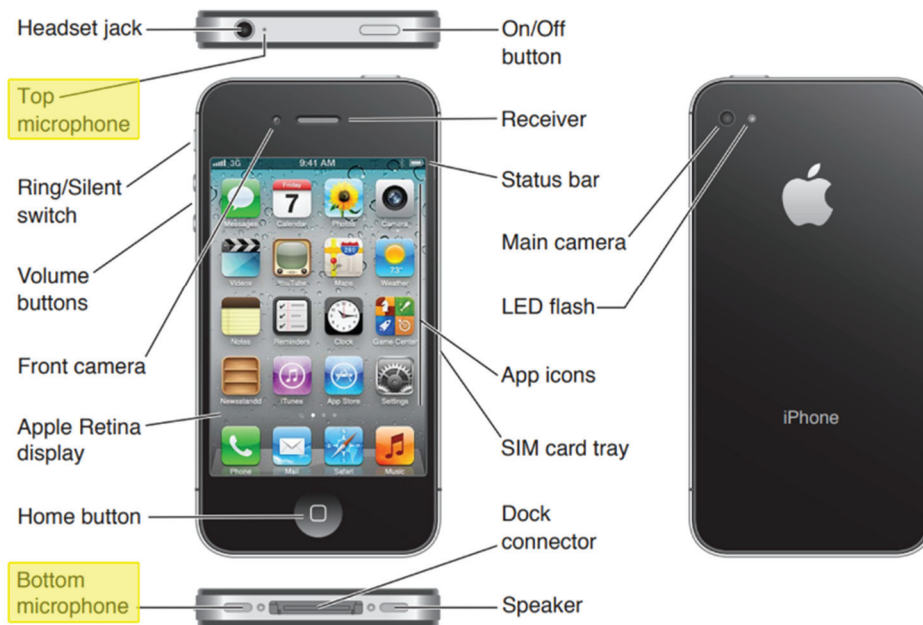


2. Claim 1

a. 1[A]/1[E]/1[G]

Claim 1 is obvious in further view of iPhone User Guide. (EX1002, ¶¶197-199.) iPhone User Guide discloses a smart phone with multiple microphones for capturing audio. (EX1008, 9 (“Top microphone” and “Bottom microphone”).) These multiple microphones are enclosed in an enclosure. (EX1002, ¶¶201-202.)

iPhone overview



It would have been obvious that the audio captured by the multiple microphones is sent to Lindahl-Williams' server over the Internet via the WLAN by the WLAN transceiver for processing. (*Supra* §IV.A.4.h; EX1002, ¶¶203-205.) It would have been obvious to combine Lindahl-Williams with iPhone User Guide. (EX1002, ¶200.) At the time of invention, the iPhone was a prevalent type of mobile phone. *Id.* Lindahl itself discloses a mobile telephone device and thus suggests a motivation to combine. (EX1011, [0027], [0044], FIG. 2.) A POSITA would be motivated to look to iPhone User Guide for implementing mobile phones and incorporate iPhone User Guide's multiple microphones in Lindahl-Williams with a reasonable expectation of success to improve usability of the device and the quality of captured audio. (EX1002, ¶200.) Mobile phones were known to include

multiple microphones by 2012. *Id.* Customer expectations, design incentives, and market forces would prompt their applications in Lindahl-Williams. *Id.*

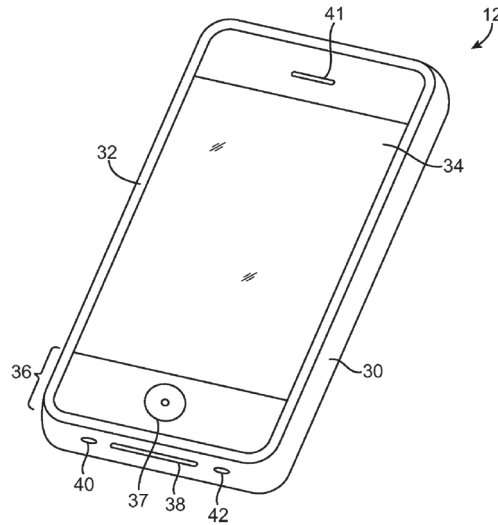


FIG. 2

3. Claim 2

Claims 2, 23-26, 29 are obvious in further view of iPhone User Guide.

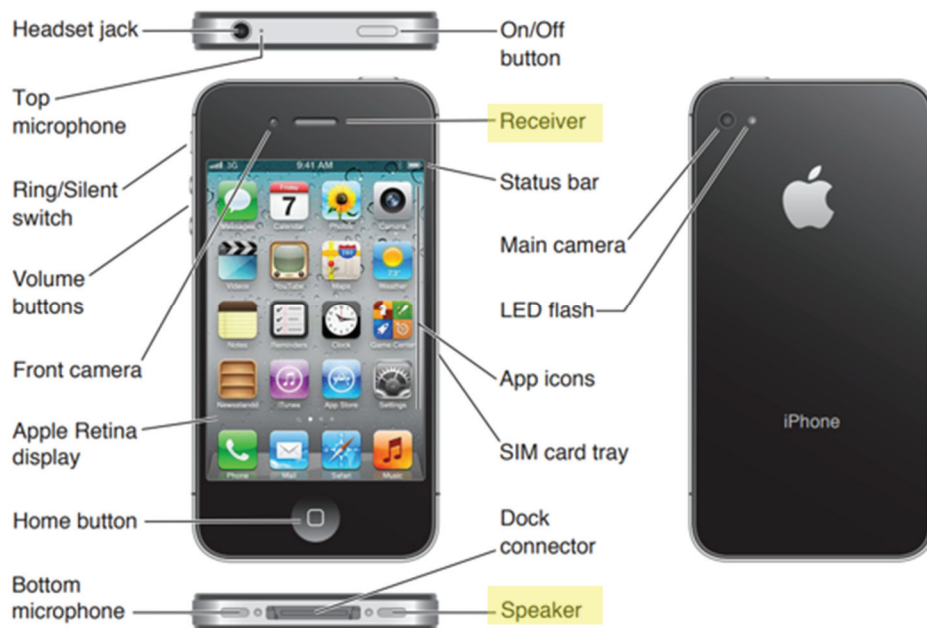
(EX1002, ¶¶206-209.) iPhone User Guide discloses that “iPhone has an *internal rechargeable battery*.” (EX1008, 34.)

It would have been obvious to combine Lindahl-Williams with iPhone User Guide. (*Supra* §IV.B.2; EX1002, ¶¶210-212.) For example, iPhone User Guide’s rechargeable battery would be incorporated into Lindahl-Williams’ portable phone to improve device usability and serving as a straightforward and obvious substitution for alternative power supplies yielding predictable results. (EX1002, ¶¶211-212.)

4. Claim 23

iPhone User Guide discloses multiple speakers in the iPhone, specifically, a “Receiver” and a “Speaker”. (EX1008, 9.) A POSITA would understand that these are two speakers. (EX1019, 3:20-22, 3:52-55; *supra* §IV.A.4.d; EX1018, 122-123; EX1014, 763; EX1002, ¶¶213-218.)

iPhone overview



iPhone User Guide discloses FaceTime video calls. “FaceTime video calls (iPhone 4 or later) let you see as well as *hear the person you’re talking to*. . . . No setup is needed, but you must have a *Wi-Fi connection to the Internet*.” (EX1008, 45.) A POSITA would understand that the iPhone speaker (*e.g.*, the loudspeaker at the bottom) is activated or controlled in response to data received from the WLAN

(e.g., audio stream received via Wi-Fi during a FaceTime call). (EX1002, ¶¶219-220.) It also follows that the speaker would be coupled to the WLAN transceiver (e.g., via processors.) *Id.* Further, the iPhone speaker is directly or indirectly affecting, changing, producing, or creating sound. *Id.*

As discussed, it would have been obvious to combine Lindahl-Williams with iPhone User Guide. (*Supra* §IV.B.2; EX1002, ¶¶221-224.) For example, the combination would result in one of Lindahl's speakers operating being coupled to the WLAN transceiver to be actuated by data received from the WiFi connection, as taught by iPhone User Guide, to provide audio or video calls. The combination can be achieved without undue effort and with a reasonable expectation of success due to the widespread adoption of multi-speaker design and WLAN-based calls. *Id.*

5. Claims 24/25/29

As discussed, these claims describe how a conventional speaker operates. (*Supra* §IV.A.12; EX1014, 18, 903; EX1018, 122; EX1001, 105:64-106:7.) The combination discloses these claims. (*Supra* §IV.B.4; EX1002, ¶¶225-232, 234-236.)

6. Claim 26

It would have been obvious that Lindahl-Williams-iPhone User Guide system would include a DAC—an electrical signal generating actuator—for

converting received “digital audio stream” into analog signals for playback. (*Supra* §IV.A.12; EX1006, [0037]; EX1011, [0087]; EX1002, ¶233.)

C. Ground 3: Lindahl, Williams and Doherty Render Obvious Claims 4-5

1. Summary of Doherty

Doherty was published in 2007 and is prior art under §102(a) and (b).

Doherty is entitled “Wireless Home Networking Simplified” and was published by Cisco to explain setting up wireless home networks. (EX1007, Introduction; EX1002, ¶¶237-238.)

2. Claims 4/5

Claim 1 is obvious in view of Lindahl-Williams. (*Supra* §IV.A.4.) Claims 4-5 are obvious in further view of Doherty. (EX1002, ¶239.) Doherty discloses that each networked device is associated with “a unique identifier called a MAC address.” (EX1007, 99, 166.) It was common knowledge that a MAC address is a universally administered address addressing the device both in the WLAN and in the Internet. (EX1023, 24, 218, 312.) Doherty discloses and it is well-known that IP addresses are layer 3, either static or dynamic, and of IPv4 type. (EX1029, 127; EX1007, 277; EX1023, 11, 13; EX1024, 2:60-62; EX1002, ¶¶240-243, 248.)

It would have been obvious for a POSITA to look to Doherty to implement conventional wireless networks in Lindahl-Williams. (EX1002, ¶¶244-248.)

MAC and IP addressing are fundamental and ubiquitous features in networking

and would be routine and predictable design choices. Adoption of these addressing schemes is motivated by strong market incentives, product design requirements, and a reasonable expectation of success. (EX1002, ¶246.)

D. Ground 4: Lindahl, Williams and Rossing Render Obvious Claims 7-8, 11-12

1. Summary of Rossing

Rossing was published by at least July 23, 2007, and is prior art under §102(a), (b). (EX1043, 1.) Rossing is a textbook entitled “Handbook of Acoustics.” (EX1014, Title.) Rossing was published by Springer and was well-known among POSITAs. (EX1002, ¶¶249-256.) A POSITA would have looked to Rossing for details regarding conventional acoustic devices. *Id.*

2. Claims 7/8

Claim 1 is obvious in view of Lindahl-Williams. (*Supra* §IV.A.4.) Claims 7-8, 11-12, 26 are obvious in further view of Rossing. (EX1002, ¶257.)

Regarding claim 7, Rossing states that “many of the same transduction methods that have been used in microphones have at least been tried for loudspeakers including *magnetic (also known as dynamic), piezoelectric, and condenser.*” (EX1014, 763; EX1002, ¶¶258-259.) Rossing discloses “an *electrostatic loudspeaker.*” *Id.* Regarding claim 8, given Rossing, a POSITA would have understood that the speaker operates by emitting audible sound waves

that are of an omnidirectional, unidirectional, or bidirectional pattern. (EX1014, 765; EX1002, ¶¶262-263)

It would have been obvious to a POSITA to look to Rossing, a textbook on conventional electroacoustics, for implementation details of speakers and microphones in Lindahl-Williams with a reasonable expectation of success without undue effort. (EX1002, ¶¶260-261.) Lindahl discloses standard speakers and thus suggests a motivation to combine. *Id.*

3. Claims 11/12

Regarding claim 11, Rossing states that “[c]ommon to virtually all microphones ever made is a diaphragm” and that “the sound waves impinging on a diaphragm cause movement of the diaphragm, and it is this movement that is sensed and converted to an electrical audio signal.” (EX1014, 757.) Rossing discloses that an ordinary microphone is commonly omnidirectional. (EX1014, 309; EX1002, ¶¶264-267.)

Regarding claim 12, Rossing discloses types of microphones such as “the ribbon microphone and the condenser microphone.” (EX1014, 758.) Rossing discloses piezoelectric microphones and carbon microphones. (*Id.*; EX1002, ¶¶270-273.)

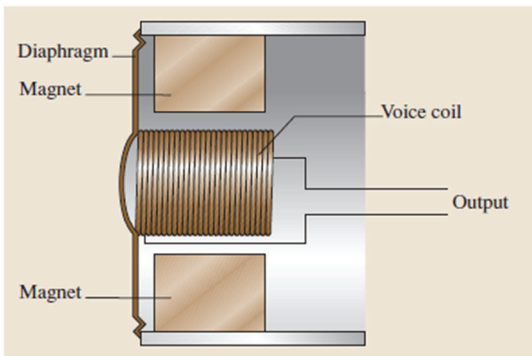


Fig. 18.10 A dynamic microphone uses a coil of wire attached to a diaphragm, suspended in a magnetic field (After [18.24])

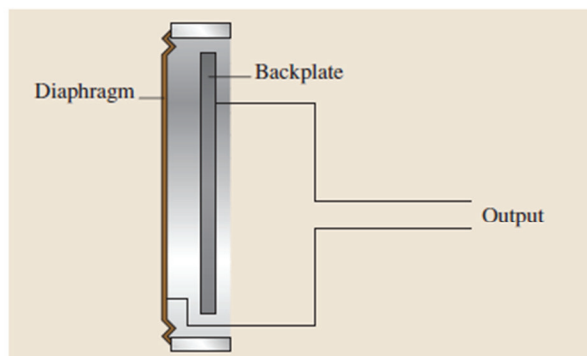


Fig. 18.9 A condenser microphone consists of a very thin diaphragm suspended parallel to a backplate (After [18.24])

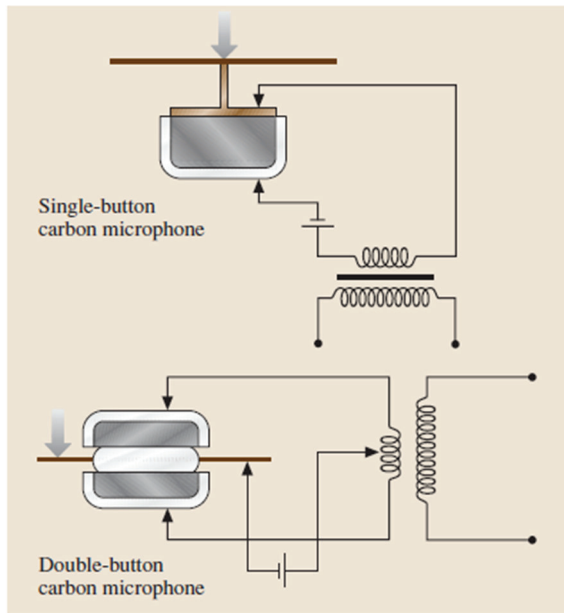


Fig. 18.11 A carbon microphone (After [18.25])

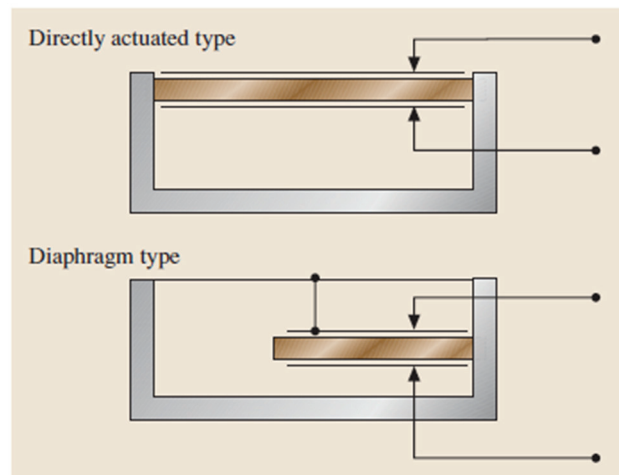


Fig. 18.12 Piezoelectric microphones (After [18.29])

Further, Rossing discloses that “the *electret condenser microphone*” which is “*practical for multi-microphone arrays.*” (EX1014, 759.) It would have been obvious to combine the teachings of Rossing with Lindahl-Williams. (*Supra* §IV.D.2; EX1002, ¶¶268-269, 274-275.) Lindahl discloses standard microphones and thus suggests a motivation to combine. *Id.*

E. Ground 5: Lindahl, Williams and Benesty Render Obvious Claims 9-10

1. Summary of Benesty

Benesty was published by at least December 11, 2007, and is prior art under §102(a), (b). (EX1041, 1.) Benesty is a textbook titled “Springer Handbook of Speech Processing”. (EX1010, Title.) Benesty provides an overview of well-known speech processing techniques. (EX1010, 1021; EX1002, ¶¶276-279.)

Benesty was published by Springer and was well-known among POSITAs. (EX1002, ¶280.) A POSITA would have looked to Benesty for techniques to, for example, implement/improve Lindahl’s audio signal capturing and processing. *Id.*

2. Claims 9/10

Claim 1 is obvious in view of Lindahl-Williams. (*Supra* §IV.A.4.) Claims 9/10 are obvious in further view of Benesty. (EX1002, ¶¶281-292.) Benesty discloses that, “[i]t is shown that *microphone arrays* have the ability to offer directional gains that can *significantly improve the quality of signal pickup in reverberant and noisy environments*.” (EX1010, 1021.) Benesty teaches that microphone arrays improve far-field audio reception (*e.g.*, where “it is not [] possible or desirable to place the microphone very close to the talker’s mouth”). *Id.*

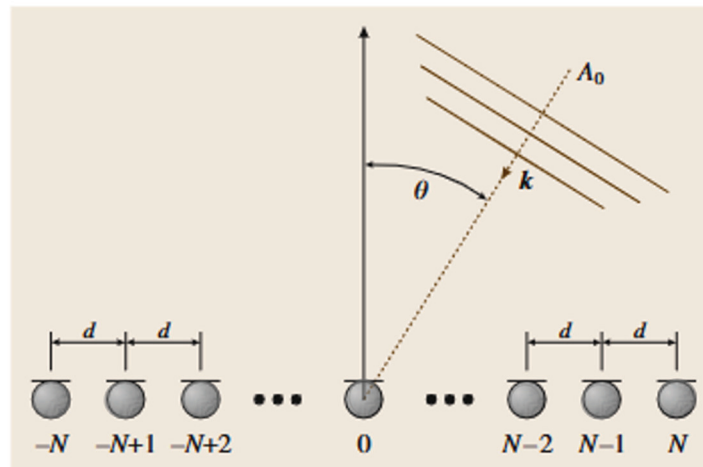


Fig. 50.1 Schematic of uniformly spaced linear array containing $2N + 1$ microphones spaced by d

Regarding claim 10, Benesty discloses that “[f]or speech source localization, microphone array beam scanning and *time delay estimation* (TDE)-based localization methods are the two most widely used approaches” and that “steering a microphone array beamformer and scanning across a room for the highest-energy output leads to an *estimate of the direction* of an active speech source.” (EX1010, 1044.) Benesty also discloses that “a set of relative time differences of arrival (TDOAs) for different microphone pairs are calculated.” (*Id.*; EX1002, ¶¶289-291.)

A POSITA would have been motivated to combine Lindahl-Williams and Benesty to incorporate Benesty’s microphone array to improve audio capturing/processing and with a reasonable expectation of success. (*Supra* §IV.A.4.b; EX1011, [0048]; EX1008, 9; EX1002, ¶¶286-288.) Lindahl itself already discloses multiple microphones and thus suggests a motivation to combine.

Id. For example, Lindahl-Williams-Benesty would yield the predictable results of improving audio capturing, especially for far-field audio reception where the microphone is not placed close to the mouth of the user. (EX1002, ¶288.)

F. Ground 6: Lindahl, Williams and Goldsmith Render Obvious Claims 20-21

1. Summary of Goldsmith

Goldsmith was published by at least January 20, 2006, and is prior art under §102(a), (b). (EX1042, 2.) Goldsmith is a textbook titled “Wireless Communications” and was published by Cambridge University Press. (EX1042, Title; EX1002, ¶¶293-297.) A POSITA would have looked to Goldsmith for knowledge about wireless communication protocols taught/suggested in Lindahl-Williams. *Id.*

2. Claim 20

Claim 1 is obvious in view of Lindahl-Williams. (*Supra* §IV.A.4.) Claims 20-21 are obvious in further view of Goldsmith. (EX1002, ¶298.) Goldsmith discloses various IEEE802.11 standards, including “IEEE 802.11a wireless LAN standard” and “802.11g.” (EX1042, 15-16; EX1002, ¶¶300-302.)

A POSITA would have been motivated to look to Goldsmith for conventional wireless networking techniques and protocols to implement Lindahl-Williams. Lindahl discloses WiFi and thus suggests a motivation to combine. (EX1011, [0025]-[0026]; EX1002, ¶¶303-306.) WLAN over IEEE 802.11a/b/g

was a known and prevalent wireless communication technique and its incorporation in Lindahl-Williams would yield predictable results with a reasonable expectation of success. (EX1002, ¶¶305.) Further, customer expectations, strong market incentives, and product design requirements would prompt the implementation in Lindahl-Williams. *Id.*

3. Claim 21

Goldsmith discloses that 802.11b/g WLANs are ISM Band II frequencies. (EX1042, 22-23; *supra* §IV.F.2; EX1002, ¶¶307-308.)

Table 1.2: Unlicensed U.S. spectrum allocations	
Band	Frequency
ISM band I (cordless phones, 1G WLANs)	902–928 MHz
ISM band II (Bluetooth, 802.11b and 802.11g WLANs)	2.4–2.4835 GHz
ISM band III (wireless PBX)	5.725–5.85 GHz
U-NII band I (indoor systems, 802.11a WLANs)	5.15–5.25 GHz
U-NII band II (short-range outdoor systems, 802.11a WLANs)	5.25–5.35 GHz
U-NII band III (long-range outdoor systems, 802.11a WLANs)	5.725–5.825 GHz

G. Ground 7: Lindahl, Williams, and Slaby Render Obvious Claims 14, 23, 28-29

1. Summary of Slaby

Slaby was filed on November 30, 2011, and is prior art under §102(e). (EX1020, 1.) Slaby discloses “a *mobile device*” with “a temperature sensitive touchscreen.” (EX1020, 3:36-45; EX1002, ¶¶309-310.)

2. Claim 14

Claim 1 is obvious in view of Lindahl-Williams. (*Supra* §IV.A.4.) Claims 14, 23, 28 are obvious in further view of Slaby. (EX1002, ¶311.) Slaby discloses a “temperature sensitive touchscreen” with “multiple thermal energy emitter/detector devices,” which are “capable of generating and emitting thermal energy.” (EX1020, 6:2-21, 3:36-45, *see also*, 32:36-45.) These thermal energy emitter/detector devices would emit non-visible light such as infrared through thermal radiation. (EX1030, 126; EX1002, ¶¶311-312.) It would have been obvious to combine Lindahl-Williams with Slaby. (*Id.*; *infra* §IV.G.3.)

3. Claims 23/28/29

Slaby’s “thermal energy emitter/detector devices” generate thermal energy “in response to the application of the current” and thus are actuators that is directly or indirectly affecting, changing, producing, or creating heat/temperature. (EX1020, 6:2-21, 3:36-45, *see also*, 32:36-45; EX1002, ¶313.)

Regarding claim 23, Slaby’s thermal energy emitter/detector devices are in the phone enclosure because they are a part of the “temperature sensitive touchscreen.” (EX1020, 6:2-21, 3:36-45, *see also*, 32:36-45.) Further, Slaby’s thermal energy emitter/detector devices are coupled to the WLAN transceiver for being activated or controlled in response to data received from the WLAN. Slaby discloses a WLAN transceiver because Slaby’s device supports “WiFi

connectivity” and has a “[t]ransceiver 310” “for receiving and transmitting wireless signals.” (EX1020, 7:58-66, 27:55-59; EX1002, ¶¶314-315.)

Moreover, the thermal energy emitter devices (1) “heat up [to] produce[] a corresponding thermal image” in response to image “downloaded *wirelessly*” (2) “selectively activat[e]” to perform device authentication in response to “authentication request received” *wirelessly*, (3) pre-tune (*e.g.*, “increase or decrease”) the temperature of the user interface in response to “a received *broadcast* of weather data,” (4) perform “heat generating tasks” in response to received WLAN data indicating whether the device is “within range of a *WiFi* node.” (EX1020, 17:11-17, 18:3-15, 6:46-52, 20:9-14, 23:42-46, 20:19-28, 29:50-56, 30:21-29, 31:51-67, 32:25-31; EX1002, ¶¶315-319.) Slaby discloses that the thermal energy emitter devices themselves may “heat up” or “a heating element” such as “a resistor” may be placed to elevate “the thermocouple junction temperatures.” (*Id.*, EX1020, 31:51-67.)

Regarding claims 28-29, Slaby discloses thermoelectric actuators that are heaters (*e.g.*, thermal energy emitter/detector devices and heating elements) and coolers (*e.g.*, “thermoelectric cooling system or a liquid coolant system”). (EX1020, 32:36-44.) They are configured for directly or indirectly affecting or changing the temperature as claimed. (EX1030, 121, 125; EX1002, ¶¶324-330.)

It would have been obvious to combine Lindahl-Williams with Slaby. (EX1002, ¶¶320-323.) For example, Lindahl-Williams-Slaby would result in a system where Slaby's thermal energy emitter/detector devices or touchscreen are incorporated in Lindahl-Williams, thereby resulting in a temperature-sensitive touchscreen with additional enhanced functionalities (*e.g.*, device authentication, display of thermal images, auto-bias of use interface temperature to improve detection capabilities). *Id.* Lindahl itself discloses a touch screen and thus suggests a motivation to combine. (*Id.*; EX1011, [0048], [0039]; EX1020, 5:47-50, 5:65-6:2.) It would have been obvious to try, and a POSITA would reasonably expect it to be a solution to improve touchscreen detection and ensure proper touchscreen operation across operating conditions. (EX1002, ¶323.)

H. Ground 8: Williams, Lim and Benesty Render Obvious Claims 1, 4-26, 29.

1. Summary of Lim

Lim was filed on November 17, 2011, and is prior art under §102(e). (EX1005, 1.) Lim relates to “controlling devices such as universal remote controls” and “appliances to be controlled.” (EX1005, [0007], [0001].) Lim discloses that “a controlling device may be provisioned with means to receive voice input” *Id.* Lim's controlling device can be device 100 (FIG. 1a) or smart device 110 in the form of a “smart phone” (FIG. 1b). (*Id.*, [0017]-[0018], [0023]; EX1002, ¶331.)

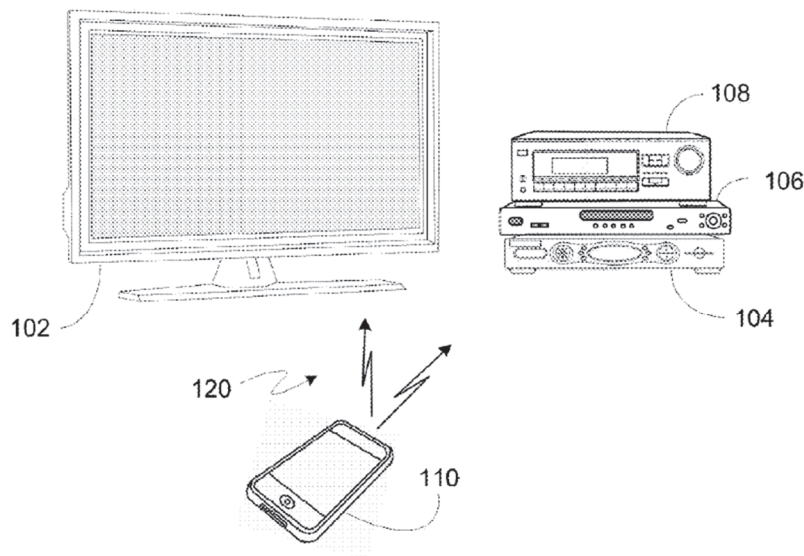


Figure 1b

Lim's device receives speech commands and responds with a *visual or audio output*. For example, the device can "prompt the user to speak the brand name of the appliance" and respond to user's speech inputs with "subsequent prompts for voice input [that] may take the form of an *illuminated LED and/or "speak" symbol, a displayed text or image, a synthesized spoken message*" (*Id.*; EX1002, ¶332.)

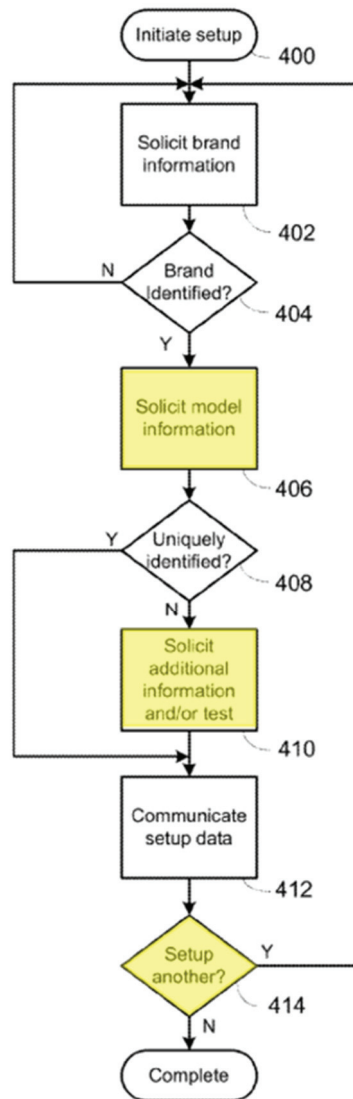


Figure 4

In Lim, the speech commands are transmitted to a *remote server 500* for processing. (EX1005, [0027], [0034]-[0035].) Further, Lim discloses (1) the *server* is accessible via the *Internet* and (2) the controlling device connects to the *Internet* via *WiFi* through a gateway device such as a router. (*Id.*; EX1002, ¶333.) An exemplary network architecture is shown in FIG. 5b below. The “*remote*

server 500 is accessible via a network 502 such as the *Internet* or PSTN” and “access to such network [502] may be . . . via a *local network connection 508*, such as *WiFi or Bluetooth*, through a gateway device such as a router (not illustrated).” (EX1005, [0027]; EX1002, ¶¶333-334.)

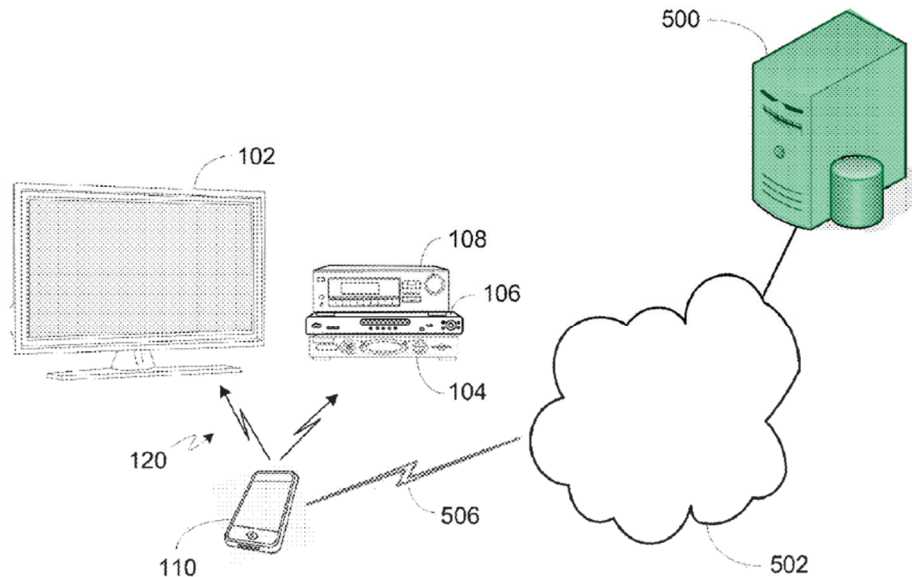


Figure 5b

Lim’s controlling device has a microphone, a display, a speaker, various sensors, and WiFi/Bluetooth capabilities. (EX1005, [0023], FIG. 3; EX1002, ¶335.)

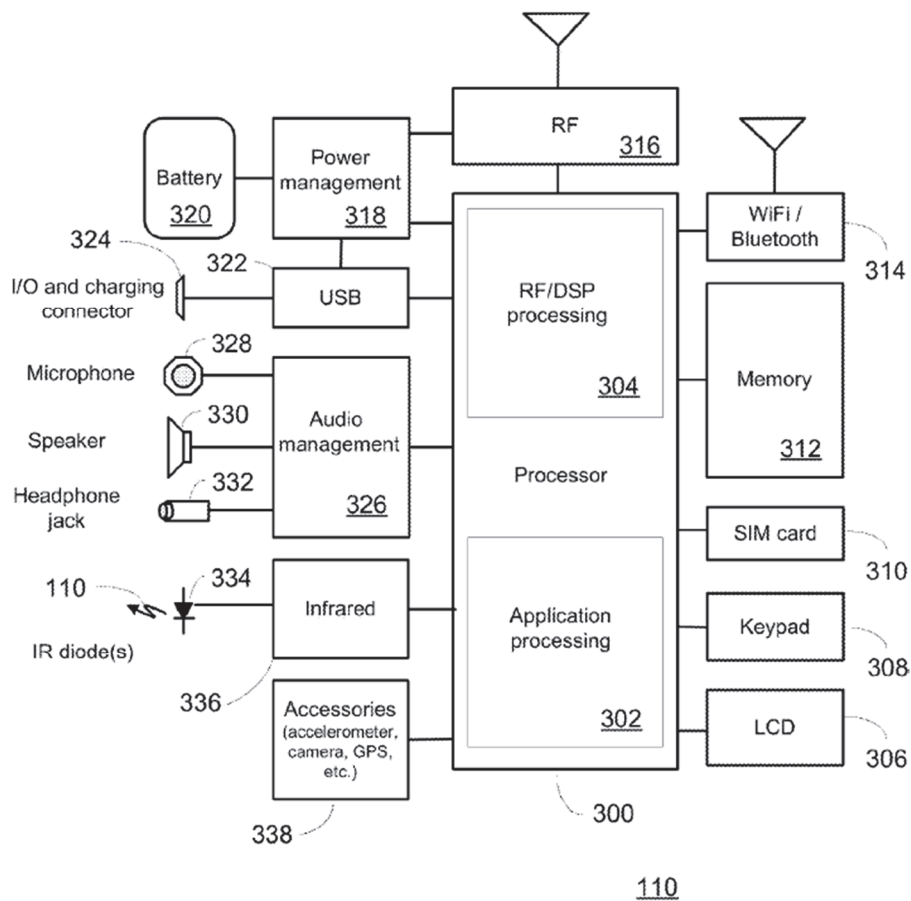


Figure 3

2. A POSITA would have been motivated to combine Williams, Lim and Benesty

At the time of the invention, it would have been obvious for a POSITA to combine Williams, Lim, and Benesty. (EX1002, ¶¶336-344.) They are analogous art to the '720 patent and relate to speech processing and playback. (*Id.*; EX1006, [0027], [0036]; EX1005, [0007]; EX1010, 1021, 980; EX1001, 105:48-52, 105:61-106:4.)

First, a POSITA would have been motivated to combine Williams with Lim. (EX1002, ¶¶337-338.) Williams teaches providing audio feedback in response to a voice command. (EX1006, [0037].) Lim teaches providing visual feedback in response to a voice command. For example, the visual output may “take the form of an illuminated *LED* and/or ‘*speak*’ *symbol*, a *displayed text or image*[.]” (EX1005, [0027], [0019].) For example, Williams-Lim would result in a system where a display as taught by Lim is incorporated into Williams’ audio interface device to provide visual feedback (*e.g.*, texts, images) in response to data (*e.g.*, processing results of voice commands) received from the server. As another example, the Williams-Lim combination would result in a system where LEDs taught by Lim are incorporated into Williams’ audio interface device as status lights to indicate device states and modes and/or to prompt a user as taught by Lim, in response to data received from the server. (EX1002, ¶338.)

Incorporating a display and/or LEDs in Williams’ audio interface device yields predictable result of improving usability by enabling multimodal feedback. (EX1002, ¶¶339-340.) Visual feedback would allow the device to present information (*e.g.*, device status, prompts, confirmation messages) that may be difficult to convey by audio alone or that may be missed in noisy environments. *Id.* These benefits align with common design incentives and market forces in

consumer electronics toward multimodal user interfaces, and would have prompted application of Lim's display/LEDs to Williams' interface devices. *Id.*

It would have been obvious to try combining Lim's display with Williams, as there is a finite number of identified, predictable solutions for providing feedback to users, indicating to a user the device state or mode, or prompting for additional input, such as visual cues, auditory cues, tactile cues, physical cues, without undue effort and with a reasonable expectation of success. (EX1002, ¶341.)

Further, it would have been obvious to a POSITA to combine Benesty with the Williams-Lim system, for example, by incorporating Benesty's multiple microphones in Williams-Lim to capture human voice data, yielding the predictable results of effectively improving audio capturing. (EX1002, ¶¶342-344.) It would have been obvious to try combining the microphone array described in Benesty with Williams-Lim, as there was a finite number of potential solutions to improve audio signal collection, such as additional software processing or incorporating hardware components. *Id.* The combination is straightforward and readily achievable by a person of ordinary skill in the art. The above-discussed benefits would prompt Benesty's application in the same field based on design incentives and other market forces. *Id.*

3. Claim 1

a. 1[PR]: “A device comprising:”

Williams-Lim-Benesty discloses this. Both Williams and Lim disclose a device. For example, Williams discloses an audio interface device, *e.g.*, a smart-home hub. “[T]he user’s home 102 may be equipped with one or more on-premises *audio monitoring devices 116* . . . having a microphone and a network interface”. (EX1006, [0016]; EX1002, ¶¶345-346.) “The *interface device* may also have a speaker or other form of audio output.” (*Id.*; EX1006, [0004], [0023], FIG. 2.) Lim also discloses this element. (*See* EX1005, [0007], [0017], [0018], [0023]; EX1002, ¶347.)

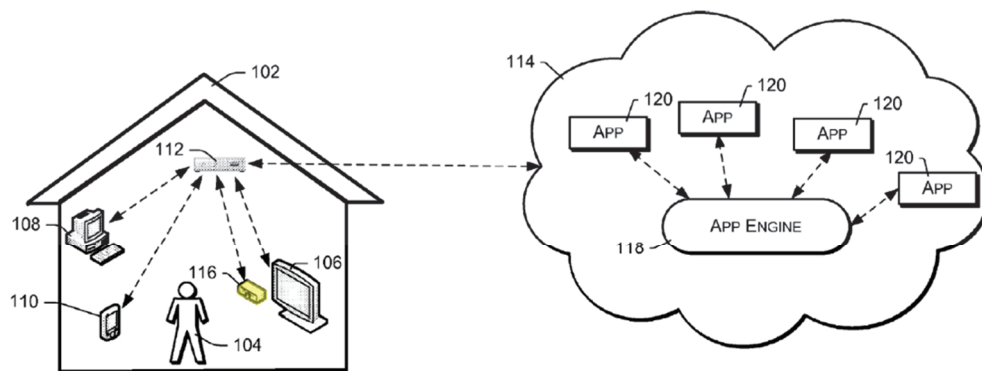
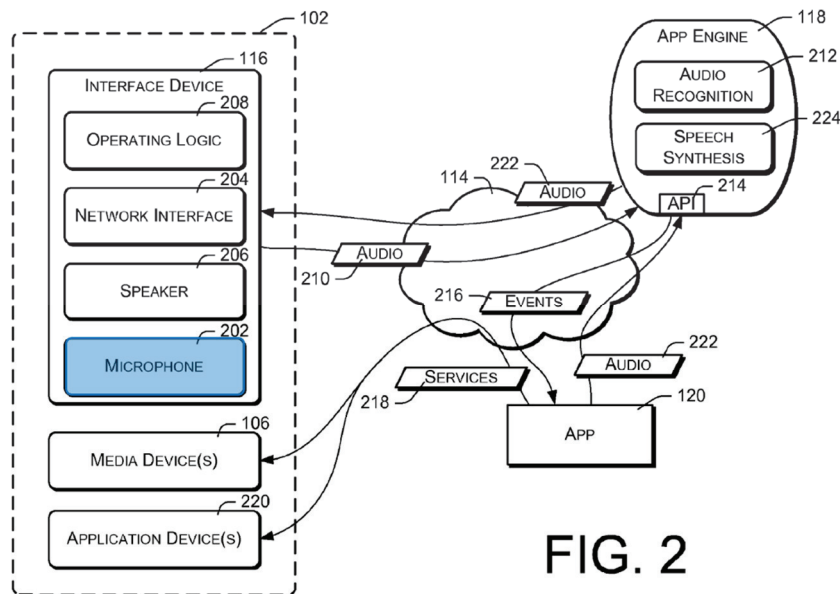


FIG. 1

b. 1[A]: “multiple microphones for capturing first human voice data;”

Williams-Lim-Benesty discloses this element. Williams discloses a device comprising at least one *microphone* for capturing *first human voice data*.

Referring to FIG. 2 below, Williams disclose that “[a]n individual interface device 116 may include a *microphone* 202 that is configured to detect ambient noise, sounds, and *speech*.” (EX1006, [0023], [0016], [0019], [0038], [0039], Claim 1-3 and 21.) The speech data is captured by the microphone and sent to the cloud for processing. (*Id.*, [0022], [0024]-[0025]; EX1002, ¶¶348-351.) Williams discloses many such captured voice commands, including: “*What’s the recipe for deviled eggs?*”, an *inquiry “about a flight or other event”*, etc. (EX1006, [0048], [0050].) Indeed, Williams’ system “provide[s] audio answers to *spoken questions*.” (EX1006, [0048].) It would have been obvious that these voice commands are captured using the microphone. (EX1002, ¶349) Lim also discloses this element. (*See* EX1005, [0023], [0019], [0025], [0013], [0027]-[0030], FIG. 4; EX1002, ¶¶352-354.)



As discussed, Benesty discloses that “microphone arrays have the ability to offer directional gains that can significantly improve the quality of signal pickup in reverberant and noisy environments.” (EX1010, 1021.) It would have been obvious to a POSITA to incorporate one or more additional microphones into Williams-Lim to improve the quality of the captured audio data. (*Supra* §IV.H.2; EX1002, ¶¶355-358.)

c. 1[B]: “a Wireless Local Area Network (WLAN) transceiver for transmitting digital data to, and for receiving digital data from, a WLAN;”

Williams-Lim-Benesty discloses this. Both Williams and Lim disclose a *WLAN transceiver* for transmitting digital data to, and for receiving digital data from, a WLAN. For example, Williams’ interface device communicates over *WLAN*. A WLAN network is characterized by a local access point (*e.g.* router)

that broadcast wireless signals to networked devices. (EX1007, 11, 100; EX1002, ¶¶359-360.) This is consistent with the '720 patent, which explains that a device can “communicate (such as send sensor info or receive actuator commands) with the *router* (or gateway) or the control server using the same or different *WANs used by the router*[.]” (EX1001, 30:13-18; EX1002, ¶361.)

A POSITA would understand that the in-home wireless network in Williams is a WLAN network. (EX1007, 11, 100; EX1002, ¶¶363.) Referring to FIG. 1, “the in-home devices connect to *a router or access point 112*, which is in turn connected to a public network 114. Connection *between in-home devices and the router 112 may be through wired or wireless technologies.*” (EX1006, [0011]-[0013], [0016], [0020], [0023].) Thus, interface device 116 communicates over a wireless network through wireless router/access point 112, which is a WLAN. (EX1002, ¶364.)

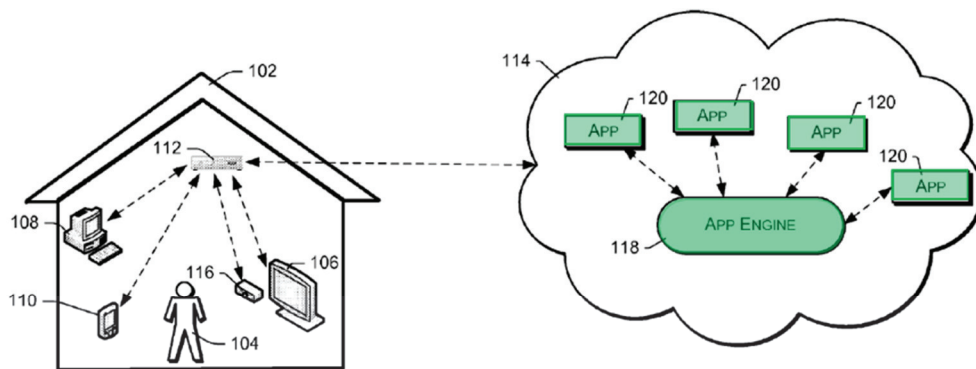


FIG. 1

Further, Williams' device is also wirelessly connected to the cloud network 114, which itself includes "*local-area networks*, wide-area networks". (EX1006, [0013]-[0015]; EX1002, ¶362.) Again, Williams' device connects to a WLAN network and thus communicates over a WLAN. (EX1006, [0012]; EX1002, ¶365.)

A POSITA would understand that Williams' device would have the necessary components to allow WLAN connection, which includes a WLAN transceiver. (EX1035, 3:17-20, 3:38-40; EX1036, [0013]-[0015].)

Indeed, FIG. 2 below shows Williams' interface device has a "network interface", which a POSITA would understand to require a transceiver, consistent with the description of the '720 patent. (EX1001, 28:63-67; EX1006, [0004], [0023], [0007], [0012], [0016], [0020]; EX1007, 99-100; EX1002, ¶364.) As discussed below, Williams' device transmits digital data (*e.g.*, voice commands) to and receives digital data (*e.g.*, audio 222) from a remote server over a WLAN. (*Id.*; *infra* §§IV.H.3.h-IV.H.3.j.) Lim also discloses this element. (*See* EX1005, [0019], [0023], [0027]; EX1021, [0067]; EX1037, 3; EX1035, 1:20-23; EX1002, ¶¶366-368.)

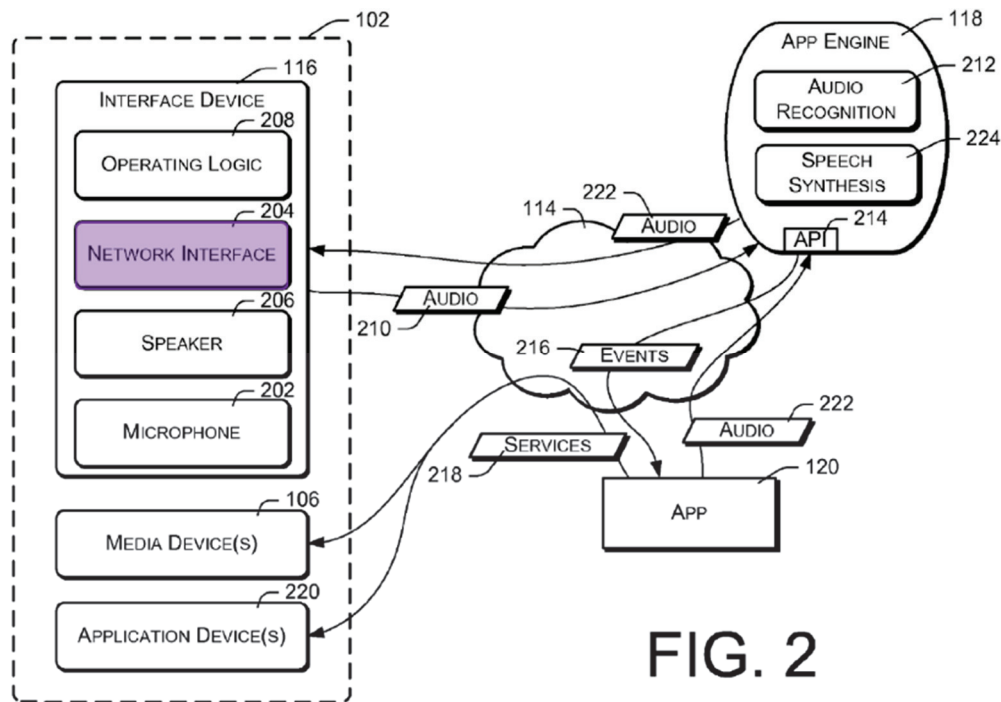


FIG. 2

d. 1[C]: “a speaker for sounding audio data;”

Williams-Lim-Benesty discloses this. Both Williams and Lim disclose a speaker for sounding audio data. For example, Williams discloses that “the interface device may also have a *speaker* or *other form of audio output*” and “the interface device may also include a *speaker 206*[.]” (EX1006, [0016], [0023], FIG. 2; EX1002, ¶¶369-370.) Lim also discloses this element. (See EX1005, [0023], [0019], [0027]; EX1002, ¶¶371-372.)

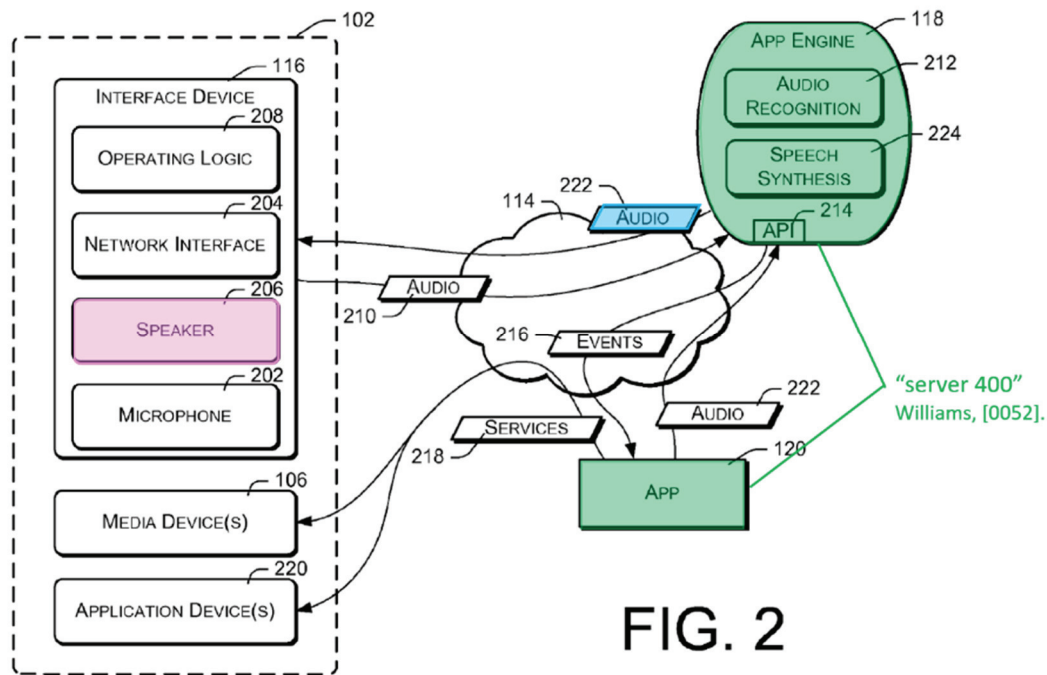


FIG. 2

e. **1[D]: “an electric light source for emitting light for emitting light;”**

Williams-Lim-Benesty discloses this. As discussed, the '720 patent explains that displays and LEDs are electric light sources. (*Supra* §IV.A.4.e; EX1001, 39:9-21.) Lim discloses that the controlling device has **LEDs** and a **display**. For example, “a controlling device 100 . . . may include . . . a means 202 to provide feedback to the user (e.g., one or more **LEDs**, illuminable keys, **display**[.])” (EX1005, [0019].) “A representative **smart device**, illustrated in FIG. 3 in the form of a smart phone, may include . . . an **LCD display 306**; a keypad 308 which may comprise physical keys, touch keys overlaying LCD 306, or a combination thereof[.]” (EX1005, [0023].) It would have been obvious that smart device 110 is similarly equipped with **LEDs**. (EX1002, ¶¶373-374.)

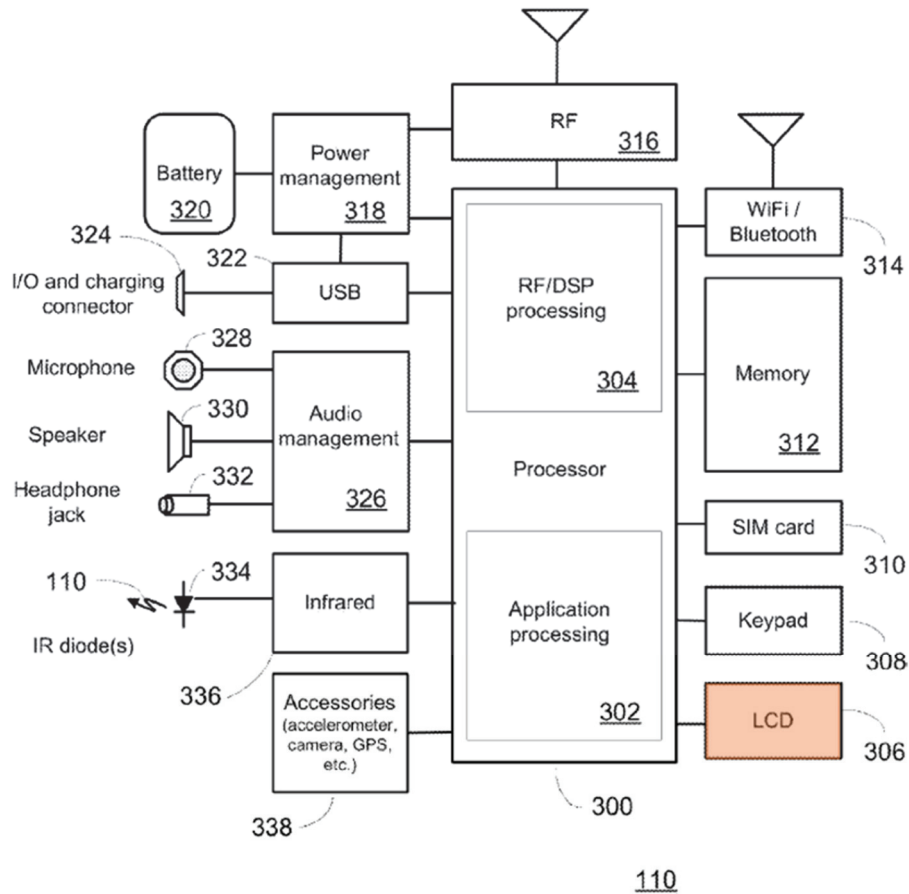


Figure 3

Lim's electric light sources (e.g., its *display* and *LEDs*) are for emitting light, consistent with the '720 patent. (EX1001, 39:19-26, 109:45-50, 110:64-111:3, 111:11-23.) "This and subsequent prompts for voice input may take the form of an *illuminated LED* and/or '*speak*' *symbol*, a *displayed text or image*." (EX1005, [0027]; EX1002, ¶375.)

It would have been obvious to a POSITA that the Williams-Lim-Benesty combination discloses this element. (*Supra* §IV.H.2; EX1002, ¶376.)

- f. 1[E]: “an enclosure that houses the multiple microphones, the WLAN transceiver, the speaker, and the electric light source;”

Williams-Lim-Benesty discloses this element. Williams discloses an *enclosure* that houses the WLAN transceiver (e.g., *network interface 204*), speaker (e.g., *speaker 206*), microphone (e.g., *microphone 202*). (*Supra* §§IV.H.3.b-IV.H.3.d; EX1006, [0023], FIG. 2; EX1002, ¶[377].)

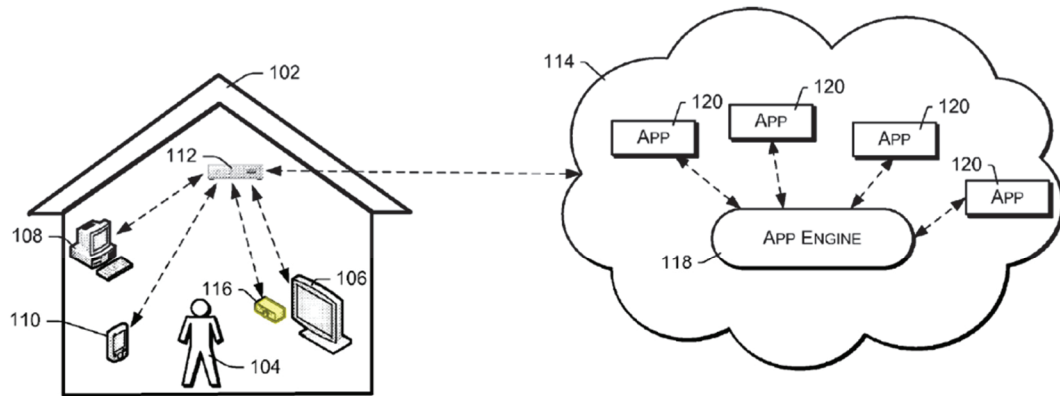


FIG. 1

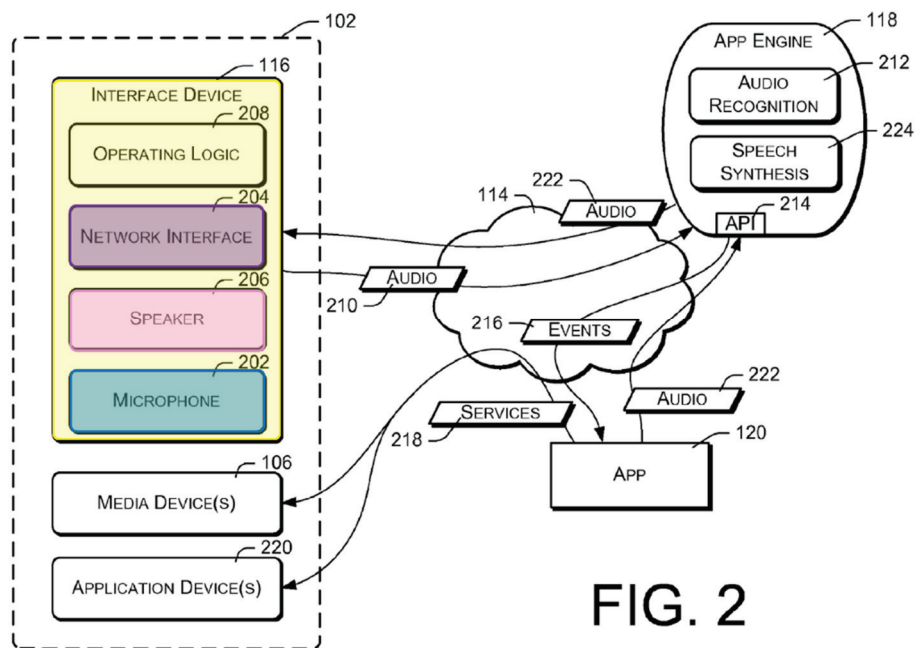


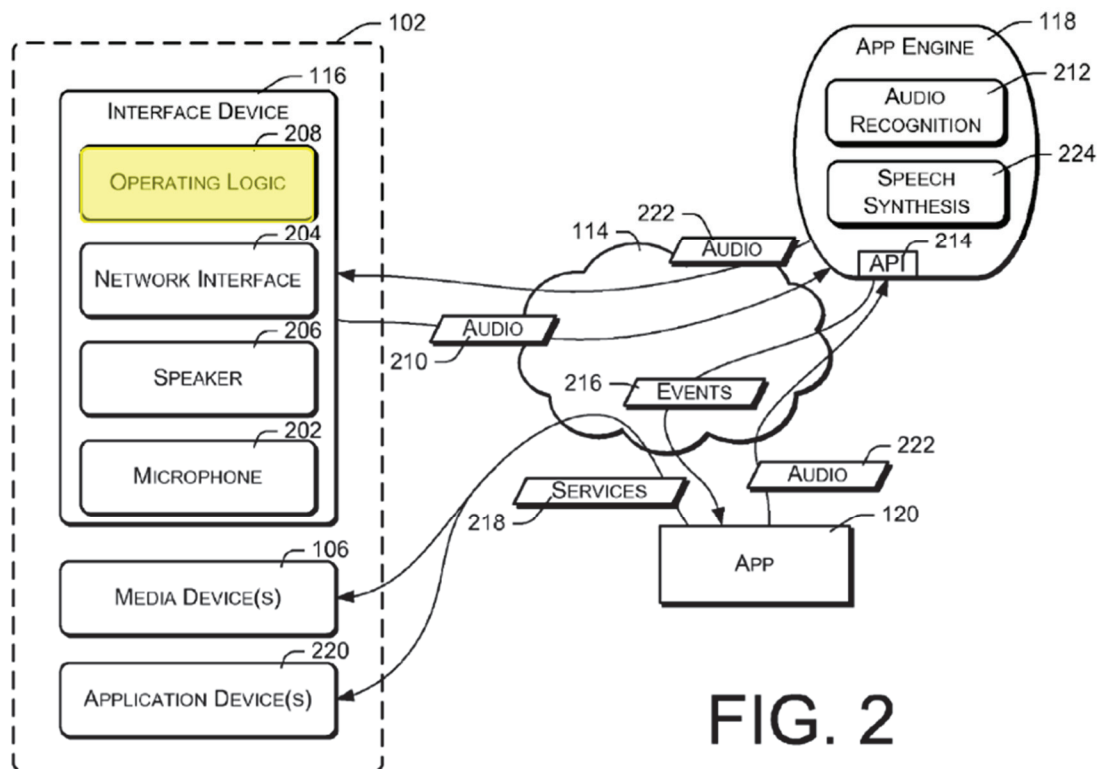
FIG. 2

It would have been obvious for a POSITA to combine Williams, Lim, and Benesty, for example, resulting in a system where Benesty's microphone array and Lim's display or LEDs are incorporated in Williams. (*Supra* §IV.H.2.) It would have been obvious that these components would all be housed in an enclosure (*i.e.*, Williams' interface device). (*Supra* §§IV.H.3.b-IV.H.3.e; EX1002, ¶¶378-382; *see* EX1005, FIGs. 1b, 3.)

- g. 1[F]: “and one or more processors programmed with computer program instructions that, when executed, cause the device to:”**

Williams-Lim-Benesty discloses this. Both Williams and Lim disclose one or more processors programmed with computer program instructions for performing the recited operations. (*Infra* §IV.H.3.h-IV.H.3.j; EX1005, [0019],

[0020], [0024]; EX1002, ¶¶383-386.) For example, referring to FIG. 2, Williams’ “operating logic 208 may be implemented as *one or more programmable processing units*, associated memory, and appropriate *instructions* stored in the memory and *executed by the processing unit*. Other types of logic may also be used to configure the interface device 116 to *perform the functions described here.*” (*Id.*; EX1006, [0023]; *infra* §§IV.H.3.h-IV.H.3.j.)



- h. 1[G]: “send, to a server device over the Internet via the WLAN by the WLAN transceiver, at least part of the first human voice data captured by the multiple microphones;”

Williams-Lim-Benesty discloses this. Both Williams and Lim disclose this element. For example, Williams discloses a *server device*. Williams’ “network 114 may thus include . . . *the Internet*. . .” (EX1006, [0015], [0001].) The network 114 includes servers, such as “a *server 400* that may be used to implement the functionality of the application engine 118 and/or any of the applications 120.” (EX1006, [0052], [0014], [0021].) “[T]he *server 400* may also comprise a *web service component 410*.” (*Id.*, [0055]; EX1002, ¶¶387-389.) Lim also discloses this element. (See EX1005, [0027]; EX1007, 11, 100; EX1002, ¶399.)

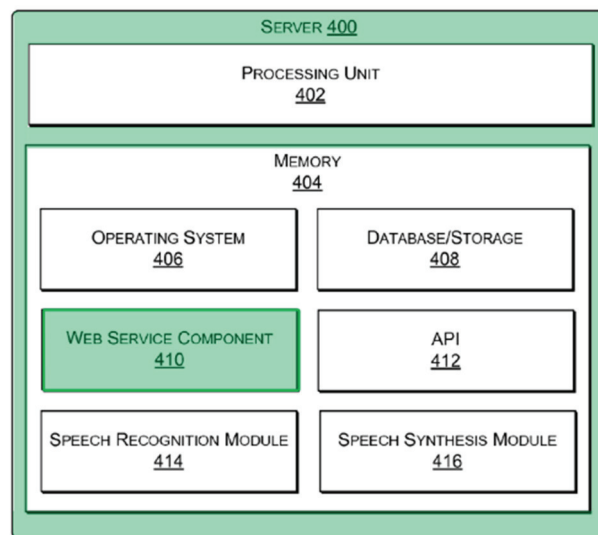


FIG. 4

Given FIG. 1, a POSITA would understand that **server 400** (running 118 and 120) is an Internet-connected **server device** in the cloud, which is connected to the in-home WLAN (e.g. via router 112). (EX1002, ¶¶393-398.) This is consistent with the '721 patent, which explains that “[t]he system may comprise . . . an Internet-connected control server (referred herein also as ‘**cloud server**’ and ‘gateway server’) external to the building . . . and communicatively coupled to the router over the Internet via the external network.” (EX1001, 32:25-45.)

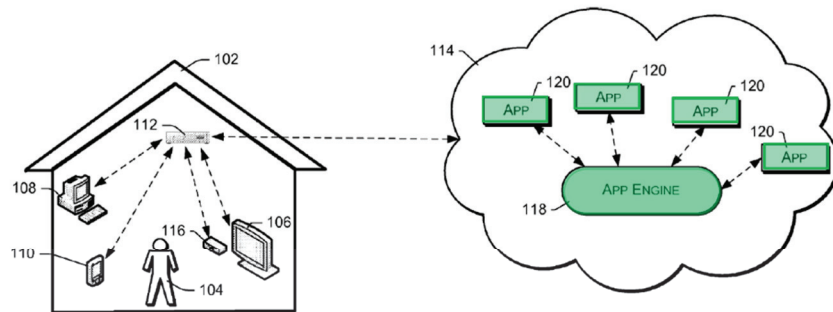


FIG. 1

Williams discloses that the device sends at least part of the **first human voice data** to the **server** over the **Internet via the WLAN**. Referring to FIG. 2, interface device 116 sends **audio 210** (captured by microphone 202) to **server 400**. (EX1006, [0023].) “[T]he application engine 118 [of **server 400**] receives **audio-based information** from the interface devices 116.” (EX1006, [0022], [0020],

[0024]-[0028], [0033]-[0034], [0038]-[0039]; EX1002, ¶¶390-393.) Lim also discloses this element. (See EX1005, [0027]-[0030]; EX1002, ¶¶403-405.)

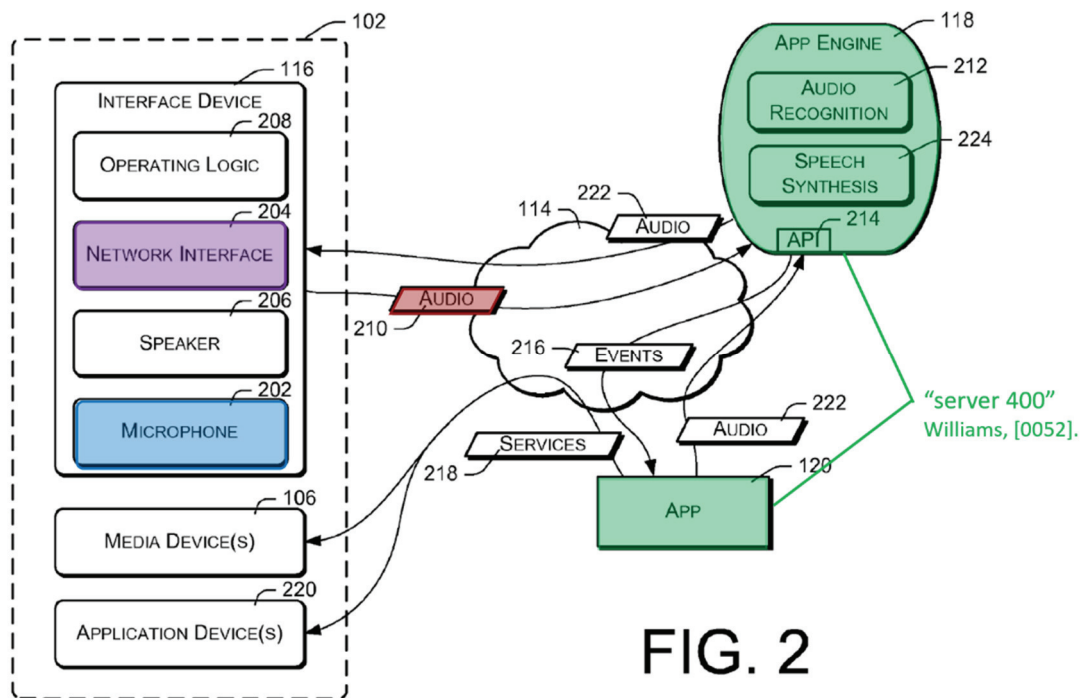


FIG. 2

Referring to FIG. 1, it would have been obvious that the *first human voice data* is sent by interface device 116 to the router 112 over the in-home WLAN and then to the *server* over the Internet (network 114). Williams discloses “the interface devices 116 communicate with and transmit environmental information to a cloud-based application service or engine 118 [of the *server*], *via the router 112 and the network 114* [i.e., the Internet]”. (EX1006, [0001], [0007], [0011]-[0015], [0020], [0038], [0051]; *supra* §IV.H.3.c.) It would be obvious that the voice data is sent by the *WLAN transceiver*. (*Supra* §IV.H.3.c; EX1006, [0023], [0007], [0012], [0016], [0020]; EX1002, ¶¶393-396.) Lim also discloses

this element. (See EX1005, [0019], [0023], [0027]-[0030], FIG. 4; EX1021, [0067]; EX1035, 3:17-20, 3:38-40; EX1036, [0013]-[0015]; EX1002, ¶¶400-402.)

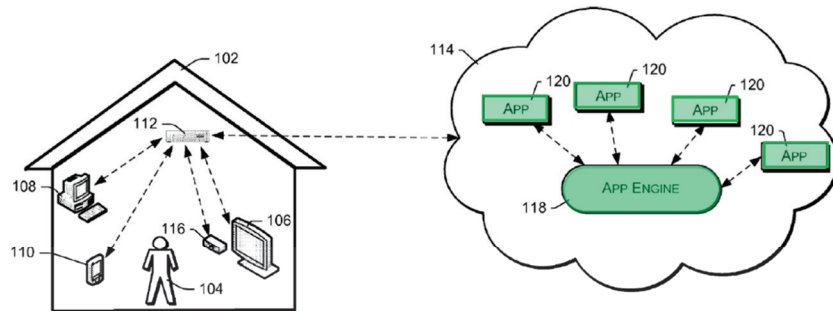


FIG. 1

- i. **1[H]: “play, by the speaker, an audio data received from the server device over the Internet via the WLAN by the WLAN transceiver;”**

Williams-Lim-Benesty discloses this. (*Supra* §IV.A.4.i; EX1002, ¶406.)

- j. **1[I]: “and emit, by the electric light source, in response to data received from the server over the Internet via the WLAN by the WLAN transceiver;”**

Williams-Lim-Benesty discloses this.

First, Lim discloses that the controlling device receives **data** from the **server** over the Internet via the WLAN by the WLAN transceiver. (EX1002, ¶407.) As discussed, Lim’s controlling device sends the captured human voice data to the **server** for speech processing. (*Supra* §IV.H.3.h.) For example, at step 402, a

user's voice input specifying "*brand name of the appliance*" captured by the controlling device may be "presented to a speech recognition means [*remote server 500*] for processing." (EX1005, [0027].) It would have been obvious that *speech processing results* are sent from the *server* back to the controlling device.

(EX1002, ¶¶408-409.) For example, a POSITA would find it obvious that *speech processing result(s)* (*i.e.*, first message) is sent back to the controlling device such that "[a]t step 404, if the speech recognition means is unable to uniquely identify the desired brand [based on the *speech processing result(s)*], step 402 may be repeated in order to solicit additional identifying information from the user."

(EX1005, [0028]; EX1002, ¶408.) Indeed, at the time of the invention, this was a routine operation by remote speech processing servers. (*E.g.*, EX1009; EX1038, 307; EX1039, 2367.) A POSITA would have understood that the *speech processing result(s)* may be the recognized speech (*e.g.* text strings), a result of analyzing the recognized speech (*e.g.* semantic processing to determine meaning and adequacy of the speech), and/or information associated with the next prompt (*e.g.* text, audio, information to display, information triggering a visual output such as flashing a light indicator). (EX1002, ¶411; EX1009; EX1038, 307; EX1039, 2367.)

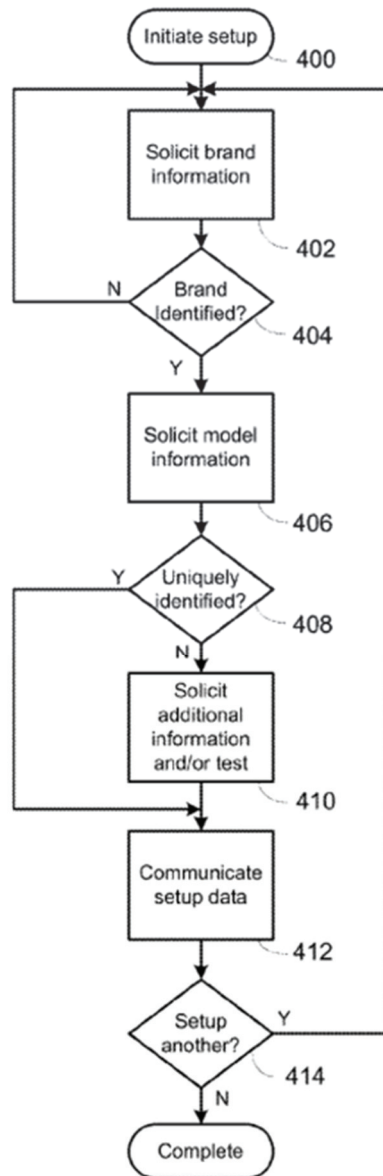


Figure 4 “Set up and configuration of controlling device”
Lim, [0025].

Lim discloses other instances of *speech processing result(s)* received by the controlling device, for example, to determine “[i]f the initially provided model information is insufficient to uniquely identify the appliance [based on the *speech processing result(s)*]” in block 408 or “if [the user] wish to configure additional

appliances [based on the *speech processing result(s)*]” in block 414. (EX1005, [0029]-[0030]; EX1002, ¶¶409-410.)

Lim discloses its controlling device communicates with a *server device via the WLAN* over the *Internet* by the WLAN transceiver (e.g., “via a local network connection 508, such as *WiFi* or Bluetooth, through a gateway device such as a router”). (EX1005, [0027].) It thus follows that Lim’s controlling device similarly receives the recited “*data*” from the server *via the WLAN* over the *Internet* by the *WLAN transceiver*. (EX1005, [0019], [0023]; EX1002, ¶412.)

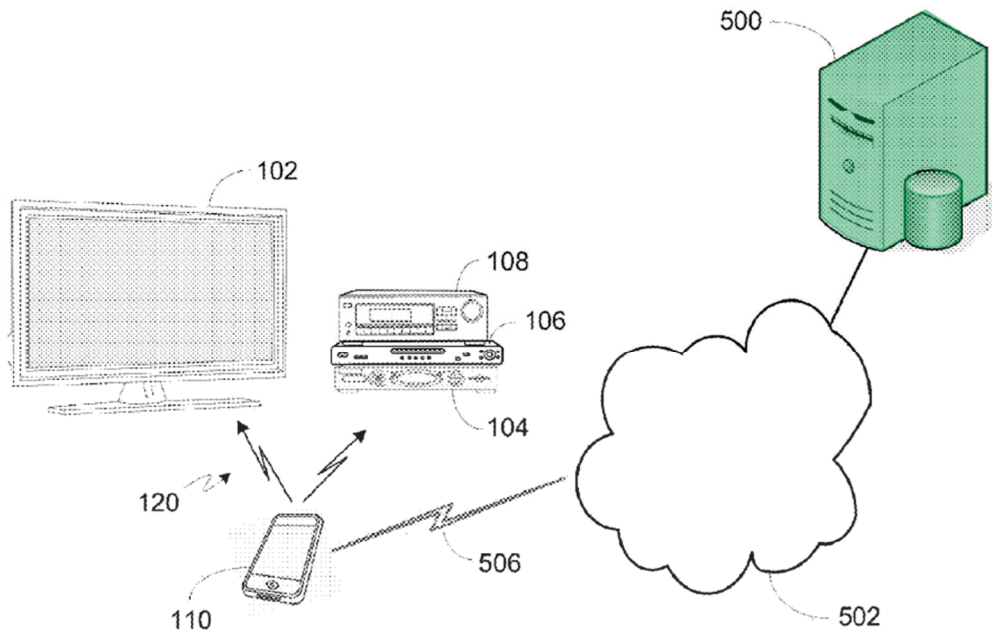


Figure 5b

Lim discloses emitting by the light source in response to the *data* received from the server. Lim’s *display* and *LEDs* emit in the device to provide a visual prompt. For example, “prompts for voice input may take the form of an

illuminated LED and/or ‘*speak*’ *symbol*, a *displayed text or image*, ... etc.”

(EX1005, [0027].) The prompt is provided in response to data (*i.e.*, *speech processing result(s)*) received from the speech recognition *server 500*. For example, Lim discloses that after step 402, “[a]t step 404, if the speech recognition means is unable to uniquely identify the desired brand [based on *speech processing result(s)*], step 402 may be repeated in order to *solicit* additional identifying information from the user.” (*Id.*, [0028], [0029], [0030], [0036]; EX1002, ¶¶413-415.)

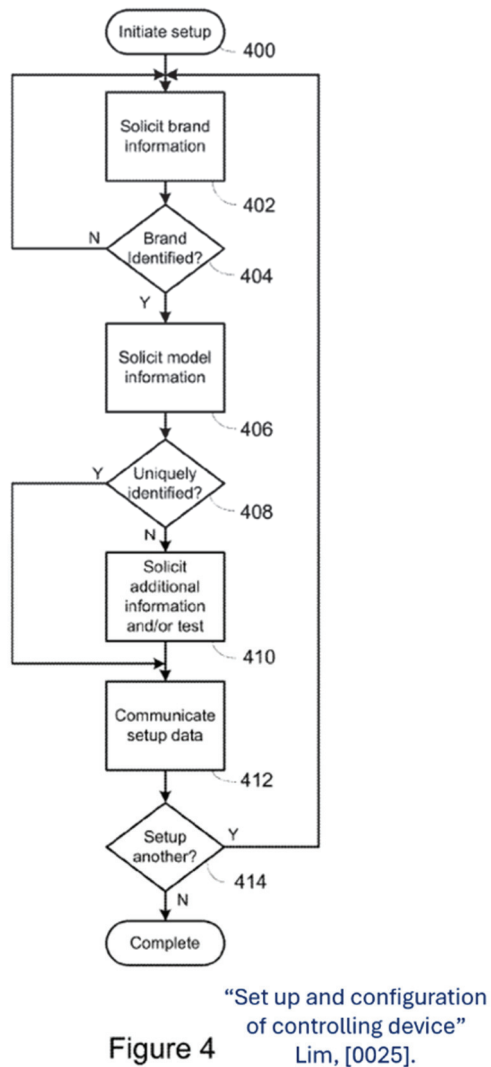


Figure 4

- k. **1[J]: “wherein the device is addressed in the WLAN using an Internet Protocol (IP) address that uniquely identifies the device in the WLAN or in the Internet.”**

Williams-Lim-Benesty discloses this. Both Williams and Lim disclose or at least suggest this. As discussed, Williams’ device communicates with the remote server in the cloud via WLAN (*i.e.*, over the Internet via a wireless router). (*Supra* §§IV.H.3.c, IV.H.3.h.) A POSITA would have understood that implementing this network involves associating each device with an IP address. (EX1002, ¶¶416-

418; *see also*, EX1005, [0038].) At the time of the invention, it was well-known that networked devices are associated with IP addresses. (EX1007, 99.)

A POSITA would have understood that the IP address may be an assigned IP address for the device (*e.g.*, by the router), the router's IP address, a combination of the two, or a globally unique IP address. (EX1002, ¶416.) "[E]ach device (computer, laptop, and so on) in a home network needs to have an IP address." (EX1007, 99, 108, 210, 175, 340.)

4. Claims 4/5/6

Williams and Lim's devices are addressed in the WLAN using IP addresses. (*Supra* §IV.H.3.k.) For example, Lim's controlling device has an "IP address associated with [it]". (EX1005, [0038].) For similar reasons discussed above, a POSITA would have understood that Williams-Lim-Benesty discloses claims 4-5. (*Supra* §IV.A.6; EX1007, 99, 108, 210, 175, 277; EX1029, 127; EX1023, 11, 13; EX1024, 2:60-62; EX1002, ¶¶419-425.) For claim 6, Williams' device includes "operating logic 208" with "associated memory" and Lim's controlling device includes "ROM," "RAM," and "a non-volatile read/write memory", which a POSITA would understand to locally store the IP addresses to allow network connectivity. (*Id.*; EX1006, [0023]; EX1005, [0019]-[0021], [0023]; EX1007, 211, Figure 10-44, 265; EX1002, ¶¶426-431.)

5. Claims 7/8/11/12

Williams' device comprises standard microphone(s) and speaker(s). (*Id.*; EX1006, [0016], [0007], [0019], [0023], [0037], [0048], claims 1-3 and 21.) Lim and Benesty also disclose standard microphones and speakers. (*Supra* §§IV.H.3.b, IV.H.3.d; EX1005, [0019], [0023], [0027]-[0029], [0037]; EX1010, 1021, 980.) A POSITA would understand that Williams-Lim-Benesty's microphones and speakers operate as claimed. (EX1014, 757-758, 309, 763, 765; EX1018, 122-123; EX1002, ¶¶432-433, 438-439.)

6. Claims 9/10

As discussed, Benesty discloses claims 9-10. (*Supra* §IV.E.2) It would be obvious to combine Williams-Lim with Benesty. (*Supra* §IV.H.2; EX1002, ¶¶434-437.)

7. Claims 13/14/15

Williams-Lim-Benesty discloses these claims. Lim's controlling device 100 "may include . . . a means 202 to provide feedback to the user (e.g., *one or more LEDs*, illuminable keys, *display* [.]" (EX1005, [0019], [0023].) Lim's smart device includes "an *LCD display 306*" and it also would have been obvious that it would include LEDs. *Id.* As discussed, display screens and LEDs are electric light sources emitting visible light for illumination or indication and the '720 patent confirms the same. (*Supra* §§IV.A.4.e, IV.A.8; EX1001, 89:50-57, 109:45-50;

110:64-111:3, 111:11-23; EX1002, ¶¶440-441, 445-446.) As discussed, LCD displays and LEDs emit non-visible light. (*Supra* §IV.A.8.) Further, Lim’s smart device “may include . . . **IR output diode(s) 334.**” (EX1005, [0023], [0017].) These diodes output non-visible light that is infrared. (EX1002, ¶¶442-444.) It would have been obvious to a POSITA that Lim’s LEDs and other light outputting features would be incorporated in the combination to enhance functionality and improve usability. *Id.*

8. Claims 16/17/18/19

Williams-Lim-Benesty discloses these claims. Both Williams and Lim disclose these. For example, regarding claim 16, Williams discloses “provid[ing] audio 222 for eventual transmission to the interface device 116, to be played on the **speaker 206.**” (*Supra* §§IV.H.3.d, IV.A.4.i; EX1006, [0037].) A POSITA would understand (and it would be obvious) that Williams’ device stores the audio 222 in its local memory upon receipt from the server and Williams’ “programmable processing units” are programmed with instructions that, when executed, cause the device to play the stored digital audio content by its speaker. (EX1006, [0023], [0037]; EX1002, ¶¶447-448, 450-451, 454-455, 459.) Further, Williams discloses claims 17-19. (*Supra* §IV.A.9)

Additionally or alternatively, Lim’s device plays a digital audio content that is synthesized human speech, which comprises words, phrases, and sentences.

(EX1002, ¶¶449, 452-453, 456-458, 460-461.) “This and subsequent prompts [by the device] for voice input may take the form of . . . a *synthesized spoken message*” (*Id.*; EX1005, [0027].) A POSITA would understand (and it would be obvious) that Lim’s device stores the synthesized spoken messages in its “memory means 312 [] compris[ing] ROM, RAM” and its “processing means 300” is programmed with instructions that plays the messages when executed. (EX1005, [0023]-[0024], [0019]-[0020]; EX1002, ¶449.) It would have been obvious that Williams’ audio 222 would include synthesized spoken message as taught by Lim (*e.g.*, generated on the server) to provide speech output. *Id.*

9. Claims 20/21

Both Williams and Lim disclose these. For example, Williams discloses usage of “cellular or other wireless data communications technologies,” thus suggesting communications via known 802.11b WLAN (ISM Band II frequencies) technologies. (*Supra* §§IV.H.3.c, IV.H.3.h; EX1006, Figs. 1-2, [0012]-[0013], [0020], [0038]; EX1042, 8-21, 22-23; EX1002, ¶462.) Further, Lim discloses that the controlling device communicates with the *remote server 500* over the *Internet* (*i.e.*, network 502) and *via WiFi*. (EX1005, [0027], Figure 5b.) As discussed, WiFi is a form of WLAN based on the IEEE 802.11 standard group comprising 802.11a/b/g standard, consistent with the ’720 patent description. (*Supra* §§IV.A.4.c, IV.H.3.c; EX1042, 600-601; EX1013, 7:48-51; EX1001, 8:25-32;

EX1002, ¶463.) Further, 802.11b WLANs are ISM Band II frequencies.

(EX1042, 22-23; EX1002, ¶464.) It would have been obvious to incorporate Lim's networking features into the combination to improve networking capabilities. *Id.*

10. Claim 22

Williams-Lim-Benesty discloses this claim. Lim's controlling device can be "in the form of a smart phone". As discussed, the smart phone housing is an handheld enclosure that houses all the claimed components. (*Supra* §IV.A.11; EX1002, ¶¶465-466.) It would have been obvious that the Williams-Lim-Benesty device can be implemented in the form of Lim's smart phone to improve portability and usability. *Id.*

11. Claims 23/24/25/26/29

Benesty discloses a "loudspeaker array" that "create suitable spatial cues at the ears of a listener." (EX1010, 1100, 1103.) The loudspeaker array has "two or more loudspeakers." (*Id.*; EX1002, ¶¶467-469.)

It would have been obvious to a POSITA to apply Benesty's loudspeaker array in Williams-Lim to improve audio playback quality. (EX1002, ¶¶471-474.) Adding a loudspeaker array is a conventional and well-established design choice to improve playback quality. *Id.* Regarding claims 23 and 29, the speakers in the loudspeaker array would be in the enclosure and they would be coupled to the WLAN transceiver for being activated or controlled as claimed. (*Supra* §§

IV.H.3.i, IV.A.12, IV.A.4.i.) These speakers are configured to directly or indirectly affect, change, produce, or create audio/sound. (*Supra* §IV.A.12; EX1018, 122-123; EX1014, 763; EX1001, 105:51-106:7; EX1002, ¶¶482-484.)

Regarding claims 24-25, a POSITA would understand that the speakers in the array are actuators (*e.g.*, causing “linear motion” and operating using the moving electromagnetic coil principle). (*Supra* §IV.A.12; EX1002, ¶¶475-480.)

Regarding claim 26, it was well-known (and obvious) that a speaker system comprises a DAC—an electrical signal generator actuator—for converting received digital audio into analog signals for playback. (EX1014, 714, 522; EX1018, 170; EX1001, 34:25-28, 107:13-20; EX1002, ¶¶470, 481.)

I. Ground 9: Williams, Lim, Benesty and iPhone User Guide Render Obvious Claim 2

iPhone User Guide discloses claim 2. (*Supra* §IV.B.2; EX1002, ¶¶485-486.)

It would have been obvious to a POSITA to combine the teachings of Williams-Lim-Benesty and iPhone User Guide. (EX1002, ¶¶487-490.) At the time of invention, iPhone was a prominent type of smart phone. *Id.* Lim itself discloses a smart phone device and thus suggests a motivation to combine. (EX1005, [0007].) A POSITA would be motivated to look to iPhone User Guide for implementing Williams-Lim-Benesty and incorporate iPhone User Guide’s rechargeable battery in Williams-Lim-Benesty to improve usability of a portable

device intended for frequent repeated use, aligning with user expectations. (EX1002, ¶¶489.) It would have been obvious to try, and a POSITA would reasonably expect a rechargeable battery to serve as a straightforward substitution or addition to disposable cells or external power supplies. *Id.*

J. Ground 10: Williams, Lim, Benesty and Doherty Render Obvious Claims 4-5

Doherty discloses claims 4-5. (*Supra* §IV.C.2; EX1002, ¶¶491-494.)

It would have been obvious for a POSITA to look to Doherty to implement conventional wireless networks in Williams-Lim-Benesty and implement the combination. (*Supra* §IV.H.2; EX1002, ¶¶495-498.) MAC and IP addressing are fundamental networking features and routine design choices with strong market incentives, predictable product design requirements, and reasonable expectation of successes for their implementation. (EX1002, ¶497.)

K. Ground 11: Williams, Lim, Benesty and Rossing Render Obvious Claims 7-8, 11-12

Rossing discloses claims 7-8 and 11-12. (*Supra* §§IV.D.2-IV.D.3; EX1002, ¶¶501-502.) Williams, Lim, and Benesty disclose standard microphones and speakers. (*Supra* §IV.H.5.) A POSITA would have been motivated to look to Rossing for implementation details of microphones and speakers for implementing Williams-Lim-Benesty. (EX1002, ¶¶502-519.) Both Williams and Lim disclose standard speakers and microphones and thus suggests a motivation to combine. *Id.*

L. Ground 12: Williams, Lim, Benesty and Goldsmith Render Obvious Claims 20-21

Goldsmith discloses claim 20-21. (*Supra* §§IV.F.2-IV.F.3; EX1002, ¶¶520-528.)

It would have been obvious for a POSITA to look to Goldsmith for conventional wireless networking techniques and protocols that can be applied to Williams-Lim-Benesty. (*Supra* §IV.H.2; EX1002, ¶¶523-526.) Williams expressly discloses usage of “cellular or other wireless data communications technologies”, and Lim discloses usage of “local network connection” and “networked protocol 120” such as WiFi that implements 802.11 WLAN. (EX1002, ¶523; EX1006, Figure 1, Figure 2, [0012], [0013], [0020], [0038]; EX1005, [0027], Figure 5b; EX1042, 599-600; EX1013, 7:48-51.) Lim itself discloses WiFi and thus suggests a motivation to combine. *Id.* WLAN over IEEE 802.11 a/b/g is a known (and popular) wireless communication technique and its application yields predictable results with a reasonable expectation of success. (EX1002, ¶¶523-526.) Customer expectations, strong market incentives, and product design requirements would prompt its application. *Id.*

M. Ground 13: Williams, Lim, Benesty, and Slaby Render Obvious Claims 14, 23, 28-29.

Slaby discloses claims 14, 23, and 28-29. (*Supra* §IV.G.2; EX1002, ¶¶529-540.)

It would have been obvious to combine Williams-Lim-Benesty with Slaby. (*Supra* §IV.H.2; EX1002, ¶¶533-536.) The combination would result in a system where Slaby's thermal energy emitter/detector devices or touchscreen are incorporated in Williams-Lim-Benesty, thereby resulting in a temperature-sensitive touchscreen with additional enhanced functionalities. (EX1002, ¶535.) Lim itself discloses a touch screen and thus suggests a motivation to combine. (EX1005, [0019], [0023]; EX1020, 5:47-50, 5:65-6:2.) It would have been obvious to try, and a POSITA would reasonably expect it to be a solution to improve touchscreen detection and ensure proper touchscreen operation across operating conditions. *Id.*

N. Ground 14: Lindahl and Evermann Render Obvious Claims 1-2, 4-8, 11-26, 29.

1. Summary of Evermann

Evermann was filed on February 12, 2007, and published on June 26, 2008, and is prior art under §102(a), (b), and (e). (EX1012, 1.) Evermann discloses a “mobile device 102” as shown in FIG. 1b. (EX1012, [0020].) Such mobile device “receiv[es] utterance that includes a search request from a user of the device” send the search request to a “remote server over a wireless data connection.” (EX1012, Abstract; EX1002, ¶¶541-542.)

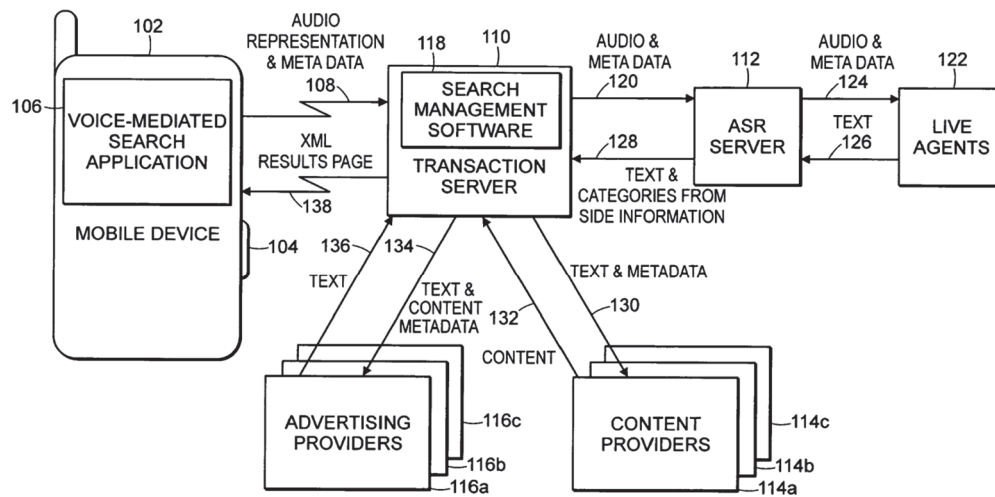


FIG. 1

Evermann's mobile device has a microphone, a display, a speaker, WLAN transceiver, battery, processor and memory. (EX1012, [0085]-[0086], [0020], FIG. 7; EX1002, ¶543.)

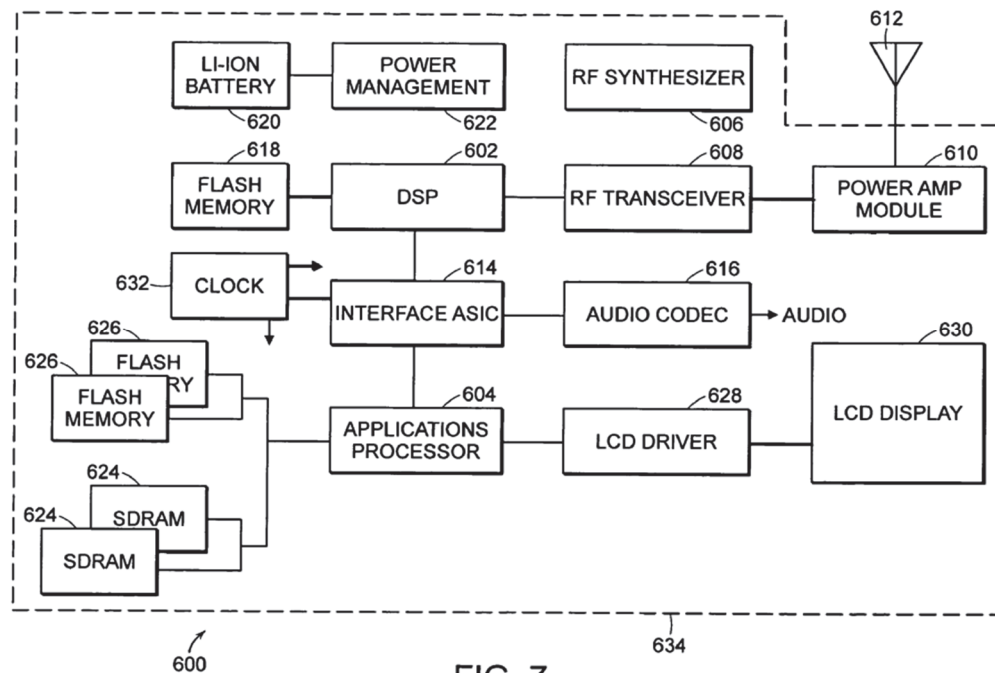


FIG. 7

Evermann's mobile device sends human voice data to a server device and receives multimodal results and advertisements from the server for playback and display. Evermann's device "recognizes the word 'search' and sends the *utterance that follows* to one or more *remote servers* for speech recognition and further handing[.]" (EX1012, [0030]-[0031], [0045].) "*Transaction server 110* transmits the combined *search results and advertisements* to the phone over *connection 138*," and such "results can be multimodal, *i.e.*, include text, *graphics audio*, and *video*." (EX1012, [0069], [0038].) Evermann discloses that "the voice search results and associated advertising content" may be presented to users "as an *audio* or a *video message*." (EX1012, [0038].) Further, a user can say "search ringtones" and "the *servers* return the results to device 102, which then *displays one or more screens of ringtone choices*." (EX1012, [0026]; EX1002, ¶544.)

2. A POSITA would have been motivated to combine Lindahl and Evermann

It would have been obvious for a POSITA to combine Lindahl and Evermann. They are analogous to the '720 patent and both relate to processing/responding to human speech commands via a remote server. (EX1011, [0073], [0043]; EX1012, [0018], [0026], [0038], abstract; EX1001, 32:60-64; EX1002, ¶545.)

A POSITA would have been motivated to combine Lindahl and Evermann, resulting in a system where Lindahl's server (*e.g.*, computing equipment 14/18) or Evermann's server transmits search results and associated advertising content, which include audio data, to the user device as taught by Evermann. This is an advantageous way to surface relevant advertisement. Further, transmitting audio or video data from the server (*e.g.*, rather than pure text data) allows rich, multimodal feedback and ensures that the data is consistent, high-quality, and free from manipulation, degradation, and variation. (*Supra* §IV.A.3.) These benefits align with common design incentives and market forces in consumer electronics. (EX1002, ¶¶545-546.)

3. Claim 1

a. 1[PR]: "A device comprising:"

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.a.) Evermann discloses a "mobile device 102" as shown in FIG. 2. (EX1012, [0020]; EX1002, ¶547.)

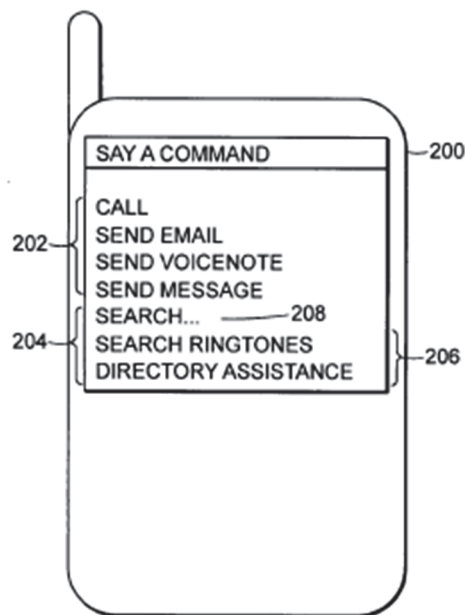


FIG. 2

b. 1[A]: “multiple microphones for capturing first human voice data;”

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.b.) Evermann also discloses “a *microphone*” for capturing “*utterance by the user* of the mobile device.” (EX1012, [0086], [0007], [0020], [0025], abstract, claim 7; EX1002, ¶548.)

c. 1[B]: “a Wireless Local Area Network (WLAN) transceiver for transmitting digital data to, and for receiving digital data from, a WLAN;”

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.c.) Evermann’s mobile device connects with a server through “Wi-Fi connections” via a “*RF radio transceiver 608*.” (EX1012, [0072], [0086], [0020]; EX1002, ¶549.) It was common knowledge that WiFi is a type of WLAN. (EX1042, 600-601; EX1013,

7:48-51; EX1001, 8:25-32.) As discussed below, Evermann's device transmits digital data (*e.g.*, utterances) to and receives digital data (*e.g.*, search results and advertisements) from a remote server over a WLAN. (*Id.*; *infra* §§IV.N.3.h-IV.N.3.j.)

d. 1[C]: “a speaker for sounding audio data;”

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.d.) Evermann's mobile device includes a “*speaker*” for sounding audio. (EX1012, [0086]; EX1002, ¶550.)

e. 1[D]: “an electric light source for emitting light for emitting light;”

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.e; EX1001, 39:9-21.) Evermann's mobile device includes a “*display for displaying text and graphics.*” (EX1012, [0020], [0086], [0004], [0008]; EX1002, ¶551.)

f. 1[E]: “an enclosure that houses the multiple microphones, the WLAN transceiver, the speaker, and the electric light source;”

Lindahl-Evermann discloses this. (*Supra* §IV.N.3.f.) As shown in FIG. 7 below, Evermann's mobile device has a “*designed housing 634*” that comprises all claimed components. (EX1012, [0086], [0089], FIG. 7; EX1002, ¶¶552-553.)

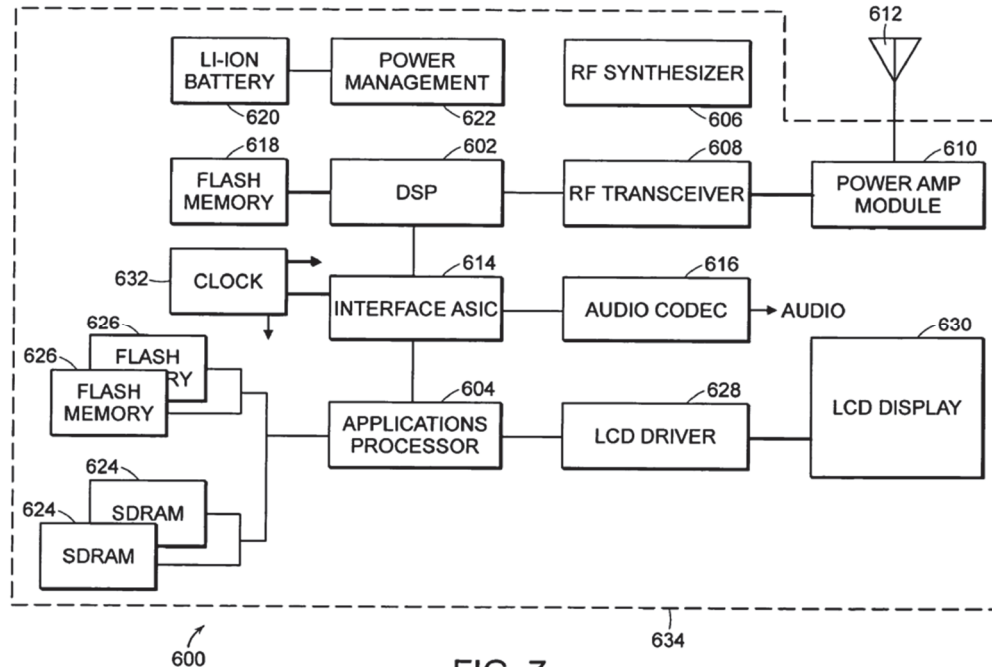


FIG. 7

- g. **1[F]: “and one or more processors programmed with computer program instructions that, when executed, cause the device to:”**

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.g.) Evermann’s mobile device includes “a *processor system* and *memory storing code*” which when executed, cause the device to perform the recited operations. (*Infra* §IV.N.3.h-IV.N.3.j; EX1012, [0009]-[0010], [0085], [0087]; EX1002, ¶554.)

- h. **1[G]: “send, to a server device over the Internet via the WLAN by the WLAN transceiver, at least part of the first human voice data captured by the multiple microphones;”**

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.h.) Evermann also discloses *server devices*, such as “[t]ransaction server 110” and “speech

recognition servers 112.” (EX1012, [0039]-[0040]. [0091].) Evermann’s device “recognizes the word ‘search’ and sends the *utterance that follows* to one or more *remote servers* for speech recognition and further handing[.]” (EX1012, [0030]-[0031], [0045]; EX1002, ¶¶555-558.)

- i. **1[H]: “play, by the speaker, an audio data received from the server device over the Internet via the WLAN by the WLAN transceiver;”**

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.i; EX1002, ¶¶559-562.) Evermann’s mobile device “receiv[es] *information* over the wireless data connection that is responsive to the search request.” (EX1012, abstract.) For example, “[t]ransaction server 110 transmits the *combined search results and advertisements* to the phone over *connection 138*,” and such “results can be multimodal, *i.e.*, include text, *graphics audio*, and *video*.” (EX1012, [0069], [0038].) Evermann discloses that “the voice search results and associated advertising content” may be presented to users “as an *audio* or a *video message*.” (EX1012, [0038].) Thus, it would be obvious that such “*graphics audio* and *video*” received from the *server* are played on Evermann’s *speaker*. (EX1012, [0086].) Further, in or before 2012, advertisements were known to be accompanied with music. (EX1022, 434.)

Evermann discloses that such connection 138 “is a *wireless connection*, ”

“connection between mobile device 102 and a remote resource” can be via “Wi-Fi connections” and that the mobile device has access to information stored on any server via “the Internet.” (EX1012, [0041], [0072], [0018].) It would have been obvious that the receipt of audio data is by the *WLAN transceiver*. (EX1012, [0072], [0086], [0020]; EX1002, ¶561.)

A POSITA would have been motivated to combine Lindahl and Evermann, resulting in a system where Lindahl’s server (*e.g.*, computing equipment 14/18) or Evermann’s server transmits results/advertisement in the form of audio data (*e.g.*, audio or video) to the user device for playback as taught by Evermann. (*Supra* §IV.N.2; EX1002, ¶562.) Multi-modal feedback improves the usability of the device.

- j. **1[I]: “and emit, by the electric light source, in response to data received from the server over the Internet via the WLAN by the WLAN transceiver,”**

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.j.)

Evermann discloses that “[w]hen *transaction server 110* returns the *voice search results and associated advertising content* to mobile device 102, VMSA 106 . . . *presents* [the information] to the user as *text and graphics* on the device’s *display*.” (EX1012, [0038], [0018], abstract; EX1002, ¶¶563-564.) For example, a user can say “search ringtones” and “the *servers* return the results to device 102,

which then *displays one or more screens of ringtone choices.*” (EX1012, [0026].)

Thus, Evermann’s display emit in the device to present the server-received results/information to user. (EX1001, 39:19-26, 109:45-50, 110:64-111:3, 111:11-23; EX1002, ¶564.)

As discussed above, data is sent to the mobile device 102 from the transaction server 110 over the Internet via the WLAN and by the *WLAN transceiver*. (*Supra* §§IV.N.3.c, IV.N.3.i; EX1002, ¶565.)

- k. 1[J]: “wherein the device is addressed in the WLAN using an Internet Protocol (IP) address that uniquely identifies the device in the WLAN or in the Internet.”**

Lindahl-Evermann discloses this. (*Supra* §IV.A.4.k.) As discussed, at the time of the invention, it was well-known that networked devices are associated with IP addresses. (*Id.*; EX1007, 99, 108, 175, 210, 340; EX1002, ¶566.)

4. Claim 2

Lindahl-Evermann discloses this. (*Supra* §IV.A.5.) Evermann’s mobile device comprises “a Li-Ion (lithium-ion) battery 620 [that] powers the phone.” (EX1012, [0087].) It would be obvious that Evermann’s lithium battery is rechargeable to improve usability of the portable mobile phone. (EX1002, ¶567.) It would have been obvious to incorporate Evermann’s rechargeable battery into Lindahl-Evermann to improve portability. *Id.*

5. Claims 4/5/6

Lindahl-Evermann discloses these claims. (*Supra* §IV.A.6.) For similar reasons as discussed above, a POSITA would have understood that Lindahl-Evermann discloses claims 4-5. (*Supra* §IV.A.6; EX1007, 99, 108, 210, 175, 277; EX1029, 127; EX1023, 11, 13; EX1024, 2:60-62; EX1002, ¶¶568-571; *see also* EX1012, [0087]- [0088], EX1007, 211, Figure 10-44, 265.) A POSITA would understand (and find it obvious that) Evermann’s memory such as one of “flash memory 618,” “SDRAM 624 and flash memory 626” would locally store the IP address. (EX1002, ¶572.)

6. Claims 7/8/11/12

Lindahl-Evermann discloses these claims. (*Supra* §IV.A.7.) Evermann also discloses standard microphones and speakers that operate as claimed. (*Supra* §§IV.N.3.b, IV.N.3.d; EX1012, [0086], [0007], [0020], [0025]; EX1014, 757-758, 309, 763, 765; EX1018, 122-123; EX1002, ¶¶573-576.)

7. Claims 13/14/15

Lindahl-Evermann discloses these claims. (*Supra* §IV.A.8.) Evermann’s mobile device includes a “*display for displaying text and graphics.*” (EX1012, [0020], [0086], [0004], [0008].) Such display may be “a *liquid crystal (LCD) display*” and it would be obvious that it would include LEDs. (EX1012, [0086];

supra §§IV.A.4.e, IV.A.8; EX1002, ¶¶577-580.) As discussed, LCD displays and LEDs emit non-visible light. (*Supra* §IV.A.8; EX1002, ¶579.)

8. Claims 16/17/18/19

Lindahl-Evermann discloses these claims. Evermann’s mobile device “displays the results to the user as text, graphics, and video and, optionally as audio output of *synthesized speech, sounds, or music*.” (EX1012, [0018].) A POSITA would understand (and it would be obvious) that Evermann’s device stores the synthesized speech in its “SDRMA 624 or flash memory 626” and its “processor system and memory storing code” are programmed with instructions that plays the speech when executed. (EX1012, [0009], [0088]; EX1002, ¶¶581-585) Further, Evermann expressly provides that its “SDRMA 624 or flash memory 626” stores “advertisements” to be played. *Id.* It would have been obvious to incorporate Evermann’s synthesized speech output into the combination to expand functionality. (EX1002, ¶582.)

9. Claims 20/21

Lindahl-Evermann discloses these claims. (*Supra* §§IV.A.10, IV.N.3.c, IV.N.3.h; EX1042, 600-601, 22-23; EX1002, ¶¶586-587.)

10. Claim 22

Lindahl-Evermann discloses this claim. (*Supra* §IV.A.11; *see also* EX1002, ¶588.)

11. Claims 23/24/25/26/29

Lindhahl-Evermann discloses these claims. (*Supra* §IV.A.12; EX1002, ¶¶589-593.)

V. NOTICES AND STATEMENTS

Pursuant to 37 C.F.R. § 42.8(b)(1), Amazon.com Services LLC (“Amazon”) and Amazon.com, Inc. are the real parties-in-interest for Petitioner.

Pursuant to 37 C.F.R. § 42.8(b)(2), Petitioner identifies the following related matters.

- *Smart Speaker LLC v. Amazon.com Services LLC*, No. 2-25-cv-00707 (E.D. Tex.)

- IPR2026-00145, *Amazon.com Services LLC v. Smart Speaker LLC*, Patent No. 11,128,710;

- IPR2026-00146, *Amazon.com Services LLC v. Smart Speaker LLC*, Patent No. 12,316,706;

- IPR2026-00147, *Amazon.com Services LLC v. Smart Speaker LLC*, Patent No. 12,010,174;

- IPR2026-00148, *Amazon.com Services LLC v. Smart Speaker LLC*, Patent No. 11,190,590; and

- IPR2026-00276, *Amazon.com Services LLC v. Smart Speaker LLC*, Patent No. 12,401,721.

Pursuant to 37 C.F.R. § 42.8(b)(3), Petitioner identifies the following counsel (and a power of attorney accompanies this Petition).

Lead Counsel	Backup Counsel
Alex S. Yap Email: AYap@mofo.com Registration No. 60,609 MORRISON & FOERSTER LLP 707 Wilshire Blvd., Suite 6000 Los Angeles, CA 90017 Tel: 213-892-5200 Fax: 213-892-5454	Mehran Arjomand Email: MArjomand@mofo.com Registration No.: 48,231 MORRISON & FOERSTER LLP 707 Wilshire Blvd., Suite 6000 Los Angeles, CA 90017 Tel: (213) 892-5200 Fax: (213) 892-5454 Yunan Yuan Email: AYuan@mofo.com Registration No.: 73,637 MORRISON & FOERSTER LLP 755 Page Mill Road Palo Alto, CA 94304-1018 Tel: (650) 813-5600 Fax: (650) 494-0792

Pursuant to 37 C.F.R. § 42.8(b)(4), service information for lead and back-up counsel is provided above. Petitioner consents to electronic service by email to Amazon_SmartSpeaker_IPRs@mofo.com.

Pursuant to 37 C.F.R. § 42.104(a), Petitioner certifies that the '720 patent is available for *inter partes* review and that Petitioner is not barred or estopped from requesting an *inter partes* review challenging the patent claims on the grounds identified in this Petition.

The USPTO is authorized to charge any required fees, including the fee as set forth in 37 C.F.R. § 42.15(a) and any excess claim fees, to Deposit Account No. **03-1952**, referencing Docket No. **41951-00008.26**.

Dated: March 23, 2026

Respectfully submitted,

By /s/ Alex S. Yap
Alex S. Yap
Reg. No. 60,609
MORRISON & FOERSTER LLP
707 Wilshire Blvd., Suite 6000
Los Angeles, CA 90017
Email: AYap@mofo.com
Tel: 213-892-5200

Counsel for Petitioner

Certification of Word Count (37 C.F.R. § 42.24)

I hereby certify that this Petition for *Inter Partes* Review has 13,929 words (as counted by the “Word Count” feature of the Microsoft Word™ word-processing system), exclusive of “a table of contents, a table of authorities, mandatory notices under § 42.8, a certificate of service or word count, or appendix of exhibits or claim listing.”

Dated: March 23, 2026

By /s/ Alex S. Yap

Alex S. Yap

Certificate of Service (37 C.F.R. § 42.6(e)(4) and 42.205(a))

I hereby certify that the attached Petition for *Inter Partes* Review and supporting materials was filed via the P-TACTS system on March 23, 2026, and served on March 24, 2026, by Priority Mail Express®, on the Patent Owner at the correspondence address indicated for U.S. Patent No. 12,401,720.

May Patents Ltd.
c/o Dorit Shem-Tov
P.O. Box 7230
Ramat-Gan, 5217102
Israel

A courtesy copy of the attached Petition for *Inter Partes* Review and supporting materials were also served via email on the following litigation counsel.

Peter Lambrianakos
Email: plambrianakos@fabricantllp.com
Clark Theodore Gordon
Email: cgordon@fabricantllp.com
Justin Kurt Truelove
Email: kurt@truelovelawfirm.com
Ryan F Oliver
Email: roliver@fabricantllp.com
Vincent J Rubino, III
Email: vrubino@fabricantllp.com
Email: smartspeaker@fabricantllp.com

Dated: March 23, 2026

/s/ Alex S. Yap
Alex S. Yap