

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
SAMSUNG ELECTRONICS CO., LTD., and
SAMSUNG ELECTRONICS AMERICA, INC.,

Petitioners

v.

CERENCE OPERATING COMPANY,

Patent Owner

Case IPR2024-01464

U.S. Patent No. 11,393,461

**PETITION FOR *INTER PARTES* REVIEW OF
U.S. PATENT NO. 11,393,461**

TABLE OF CONTENTS

PETITIONERS' EXHIBIT LIST	viii
LISTING OF CHALLENGED CLAIMS	x
I. INTRODUCTION	1
II. GROUNDS FOR STANDING (37 C.F.R. § 42.104(A))	1
III. STATEMENT OF PRECISE RELIEF REQUESTED FOR EACH CLAIM CHALLENGED	1
IV. OVERVIEW OF THE '461 PATENT	3
A. Specification	3
B. Prosecution History	5
V. OVERVIEW OF THE PRIMARY PRIOR ART REFERENCES	5
A. Sarin (Ex-1005)	6
B. Hwang (Ex-1006)	9
C. Garner (Ex-1007)	11
D. Gruber (Ex-1008)	12
E. Lebeau (Ex-1016)	13
VI. LEVEL OF ORDINARY SKILL IN THE ART	13
VII. CLAIM CONSTRUCTION	14
VIII. DETAILED EXPLANATION OF GROUNDS	14
A. Ground 1: Sarin in Combination with Hwang Renders Obvious Claims 1-4, 6, 7, 11, 14, 16, and 18	14
1. Motivation to Combine Sarin and Hwang	14
2. Independent Claim 1	20

a.	1[pre]: A method of monitoring an acoustic environment of a mobile device for voice commands when the mobile device is operating in a low power mode, the method comprising:	20
b.	1[a]: receiving acoustic input from the environment of the mobile device while the mobile device is operating in idle mode; and.....	22
c.	1[b]: in response to receiving the acoustic input, determining whether the received acoustic input includes a voice command by processing the received acoustic input, ...	24
d.	1[c]: the processing employing at least one contextual cue, the at least one contextual cue comprising information other than content of the acoustic input and comprising a time the acoustic input was received, motion of the mobile device, location of the mobile device, occurrence of an event on the mobile device, recent activity on the mobile device, a stored sample of the acoustic environment of the mobile device and/or a power source of the mobile device,.....	26
e.	1[d]: the processing of the received acoustic input comprising a plurality of processing stages, at least one of the processing stages being performed while the mobile device is operating in the idle mode.....	28
3.	Dependent Claims 2-4, 6, 7, 11, 14, and 16.....	30
a.	Claim 2: The method of claim 1, further comprising transitioning the mobile device from the idle mode to an active mode in response to detecting that the acoustic input includes a voice command.	30
b.	Claim 3: The method of claim 1, wherein the at least one contextual cue includes the time the acoustic input was received, and the method comprises influencing evaluation of at least one of the plurality of processing	

stages based on the time the acoustic input was received.
.....31

- c. Claim 4: The method of claim 1, wherein the at least one contextual cue includes recent activity on the mobile device, the method further comprising evaluation of at least one of the plurality of processing stages based on at least one of how recently the user activated the mobile device, how recently the user issued a voice command and how recently the user performed at least one action on the mobile device.....32
- d. Claim 6: The method of claim 1, wherein the at least one contextual cue includes motion of the mobile device received as an input from an accelerometer of the mobile device, the method comprises influencing evaluation of at least one of the plurality of processing stages based on an assessment of whether the user moved the mobile device proximate in time with receiving the acoustic input based, at least in part, on the input from the accelerometer. 34
- e. Claim 7: The method of claim 1, wherein the at least one contextual cue includes the location of the mobile device, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on determining whether the mobile device is located in at least one of a vehicle, a home or a place of work of the user of the mobile device.....36
- f. Claim 11: The method of claim 1, wherein the mobile device includes a primary processor and a secondary processor, and wherein at least one first processing stage of the plurality of processing stages is performed by the secondary processor prior to engaging the primary processor to process the acoustic input. 36
- g. Claim 14: The method of claim 11, wherein performing the at least one first processing stage comprises performing limited vocabulary automatic speech

	recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger	38
h.	Claim 16: The method of claim 1, wherein the mobile device is capable of connecting to at least one server via a network, and the method comprises transmitting at least a portion of the acoustic input to the at least one server via the network for processing by the at least one server.	39
4.	Independent Claim 18	41
a.	18[pre]: A mobile device comprising:.....	41
b.	18[a]: at least one processor configured to:.....	41
	(i) Elements 18[a(i)]-18[a(iii)]	41
B.	Ground 2: Sarin in Combination with Hwang and Garner Renders Obvious Claim 13.....	42
1.	Motivation to Combine Garner with Sarin-Hwang	42
2.	Dependent Claim 13	43
a.	Claim 13: The method of claim 11, wherein performing the at least one first processing stage includes performing at least one voice activity detection processing stage including performing at least one of spectral analysis on the acoustic input to evaluate whether a spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether signal periodicity of the acoustic input is indicative of voice activity, and phone loop analysis to evaluate whether the acoustic input includes speech.	43
C.	Ground 3: Sarin in Combination With Hwang and Gruber Renders Obvious Claims 3, 5, 7, 9, 10, 12, 15, and 17	46
1.	Motivation to Combine Gruber With Sarin and Hwang.....	46
2.	Dependent Claims 3, 5, 7, 9, 10, 12, 15, and 17	48

a.	Claim 3.....	48
b.	Claim 5: The method of claim 1, wherein the at least one contextual cue includes an occurrence of an event on the mobile device, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on whether the acoustic input is received proximate in time with the occurrence of the event.	50
c.	Claim 7.....	54
d.	Claim 9: The method of claim 1, wherein the at least one contextual cue includes the sample of the acoustic environment obtained by the mobile device by storing acoustic input previously received by the mobile device and determined not to include a voice command, and the method comprises subtracting the sample of the acoustic environment from the acoustic input.....	56
e.	Claim 10: The method of claim 1, wherein a sequence of performing the plurality of processing stages is determined, at least in part, by the at least one contextual cue.....	58
f.	Claim 12: The method of claim 11, wherein at least one second processing stage of the plurality of processing stages is performed by the primary processor only after the at least one first processing stage is performed and neither concludes that the acoustic input includes a voice command nor concludes that the acoustic input is the result of spurious acoustic activity.	61
g.	Claim 15: The method of claim 11, wherein performing the at least one second processing stage comprises performing at least one speech processing stage including at least one of automatic speech recognition on the acoustic input and at least one natural language processing stage to understand the meaning of the voice command.....	62

h.	Claim 17: The method of claim 1, comprising performing one or more default actions when a specific task could not be ascertained from the acoustic input.	65
D.	Ground 4: Sarin in Combination with Hwang and Lebeau Renders Obvious Claim 8.....	67
1.	Motivation to Combine Lebeau with Sarin-Hwang.....	67
2.	Dependent Claim 8	69
a.	Claim 8: The method of claim 1, wherein the at least one contextual cue includes the power source of the mobile device, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on at least one of determining whether the mobile device is being powered using a battery or wall power, and a state of the battery if the mobile device is being powered using the battery.....	69
IX.	THE DISCRETIONARY FACTORS FAVOR INSTITUTING TRIAL	71
A.	35 U.S.C. § 314(a).....	71
1.	Stay.....	71
2.	Trial Date	72
3.	Diligence/Investment	73
4.	Overlap.....	74
5.	Parties.....	74
6.	Other considerations	75
B.	35 U.S.C. § 325(d).....	75
X.	MANDATORY NOTICES UNDER 37 C.F.R. § 42.8.....	77
A.	Real Parties-in-Interest	77
1.	Related Matters	77

B.	Lead and Backup Counsel.....	78
C.	Service Information.....	80
D.	Power of Attorney	80
XI.	FEES	80

PETITIONERS' EXHIBIT LIST

Exhibit No.	DESCRIPTION
1001	U.S. Patent 11,393,461 (“’461 Patent”)
1002	Declaration of Christopher M. Schmandt
1003	Curriculum Vitae of Christopher M. Schmandt
1004	File History of U.S. Patent No. 11,393,461
1005	U.S. Patent No. 9,031,847 (“Sarin”)
1006	U.S. Patent Application Publication No. 2013/0110521 (“Hwang”)
1007	U.S. Patent No. 7,475,012 (“Garner”)
1008	U.S. Patent Application Publication No. 2012/0265528 (“Gruber”)
1009	U.S. Patent No. 8,396,463 (“Marcellino”)
1010	U.S. Patent No. 8,331,315 (“Purkayastha”)
1011	U.S. Patent No. 8,694,316 (“Pickering”)
1012	U.S. Patent Application Publication No. 2013/0173269 (“Adler”)
1013	Intentionally Left Blank
1014	U.S. Patent Application Publication No. 2020/0279576 (“Binder”)
1015	U.S. Patent Application Publication No. 2014/0006825 (“Shenhav”)
1016	U.S. Patent Application Publication No. 2012/0034904 (“LeBeau”)
1017	U.S. Patent Application Publication No. 2010/0129068 (“Binda”)
1018	U.S. Patent Application Publication No. 2012/0296655 (“Kristjansson”)
1019	U.S. Patent Application Publication No. 2014/0163978 (“Basye”)

Exhibit No.	DESCRIPTION
1020	Claim Mapping Table
1021	Plaintiff Cerence Operating Company's Disclosure of Asserted Claims and Infringement Contentions, <i>Cerence Operating Company v. Samsung Electronics Co., Ltd. et al.</i> , No. 2:23-cv-0482-JRG-RSP (E.D. Tex. Mar. 11 2024)
1022	Docket Control Order, <i>Cerence Operating Company v. Samsung Electronics Co., Ltd. et al.</i> , No. 2:23-cv-0482-JRG-RSP (E.D. Tex. Apr. 16, 2014)
1023	Intentionally Left Blank
1024	U.S. Patent Application Publication No. 2013/0332165 ("Beckley")
1025	U.S. Patent Application Publication No. 2014/0350927 ("Yamabe")
1026	U.S. Patent Application Publication No. 2003/0179888 ("Burnett")
1027	Intentionally Left Blank
1028	Intentionally Left Blank
1029	U.S. Patent Application Publication No. 2013/0006638 ("Lindahl")

LISTING OF CHALLENGED CLAIMS

Claim 1	
1[pre]	A method of monitoring an acoustic environment of a mobile device for voice commands when the mobile device is operating in a low power mode, the method comprising:
1[a]	receiving acoustic input from the environment of the mobile device while the mobile device is operating in idle mode; and
1[b]	in response to receiving the acoustic input, determining whether the received acoustic input includes a voice command by processing the received acoustic input,
1[c]	the processing employing at least one contextual cue, the at least one contextual cue comprising information other than content of the acoustic input and comprising a time the acoustic input was received, motion of the mobile device, location of the mobile device, occurrence of an event on the mobile device, recent activity on the mobile device, a stored sample of the acoustic environment of the mobile device and/or a power source of the mobile device,
1[d]	the processing of the received acoustic input comprising a plurality of processing stages, at least one of the processing stages being performed while the mobile device is operating in the idle mode.
Claim 2	
2	The method of claim 1, further comprising transitioning the mobile device from the idle mode to an active mode in response to detecting that the acoustic input includes a voice command.
Claim 3	
3	The method of claim 1, wherein the at least one contextual cue includes the time the acoustic input was received, and the method comprises influencing evaluation of at least one of the plurality of

	processing stages based on the time the acoustic input was received.
Claim 4	
4	The method of claim 1, wherein the at least one contextual cue includes recent activity on the mobile device, the method further comprising evaluation of at least one of the plurality of processing stages based on at least one of how recently the user activated the mobile device, how recently the user issued a voice command and how recently the user performed at least one action on the mobile device.
Claim 5	
5	The method of claim 1, wherein the at least one contextual cue includes an occurrence of an event on the mobile device, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on whether the acoustic input is received proximate in time with the occurrence of the event.
Claim 6	
6	The method of claim 1, wherein the at least one contextual cue includes motion of the mobile device received as an input from an accelerometer of the mobile device, the method comprises influencing evaluation of at least one of the plurality of processing stages based on an assessment of whether the user moved the mobile device proximate in time with receiving the acoustic input based, at least in part, on the input from the accelerometer.
Claim 7	
7	The method of claim 1, wherein the at least one contextual cue includes the location of the mobile device, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on determining whether the mobile device is located in at least one of a vehicle, a home or a place of work of the user of the mobile device.
Claim 8	

8	The method of claim 1, wherein the at least one contextual cue includes the power source of the mobile device, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on at least one of determining whether the mobile device is being powered using a battery or wall power, and a state of the battery if the mobile device is being powered using the battery.
Claim 9	
9	The method of claim 1, wherein the at least one contextual cue includes the sample of the acoustic environment obtained by the mobile device by storing acoustic input previously received by the mobile device and determined not to include a voice command, and the method comprises subtracting the sample of the acoustic environment from the acoustic input.
Claim 10	
10	The method of claim 1, wherein a sequence of performing the plurality of processing stages is determined, at least in part, by the at least one contextual cue.
Claim 11	
11	The method of claim 1, wherein the mobile device includes a primary processor and a secondary processor, and wherein at least one first processing stage of the plurality of processing stages is performed by the secondary processor prior to engaging the primary processor to process the acoustic input.
Claim 12	
12	The method of claim 11, wherein at least one second processing stage of the plurality of processing stages is performed by the primary processor only after the at least one first processing stage is performed and neither concludes that the acoustic input includes a voice command nor concludes that the acoustic input is the result of spurious acoustic activity.

Claim 13	
13	The method of claim 11, wherein performing the at least one first processing stage includes performing at least one voice activity detection processing stage including performing at least one of spectral analysis on the acoustic input to evaluate whether a spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether signal periodicity of the acoustic input is indicative of voice activity, and phone loop analysis to evaluate whether the acoustic input includes speech.
Claim 14	
14	The method of claim 11, wherein performing the at least one first processing stage comprises performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger.
Claim 15	
15	The method of claim 11, wherein performing the at least one second processing stage comprises performing at least one speech processing stage including at least one of automatic speech recognition on the acoustic input and at least one natural language processing stage to understand the meaning of the voice command.
Claim 16	
16	The method of claim 1, wherein the mobile device is capable of connecting to at least one server via a network, and the method comprises transmitting at least a portion of the acoustic input to the at least one server via the network for processing by the at least one server.
Claim 17	
17	The method of claim 1, comprising performing one or more default actions when a specific task could not be ascertained from the acoustic input.

Claim 18	
18[pre]	A mobile device comprising:
18[a]	at least one processor configured to:
18[a(i)]	in response to the mobile device receiving acoustic input from the environment while operating in an idle mode, determine whether the received acoustic input includes a voice command by processing the received acoustic input,
18[a(ii)]	the processing employing at least one contextual cue, the at least one contextual cue comprising information other than content of the received acoustic input and comprising a time the acoustic input was received, motion of the mobile device, location of the mobile device, occurrence of an event on the mobile device, recent activity on the mobile device, a stored sample of the acoustic environment of the mobile device and/or a power source of the mobile device,
18[a(iii)]	the processing of the received acoustic input comprising a plurality of processing stages, at least one of the processing stages being performed while the mobile device is operating in the idle mode.

I. INTRODUCTION

Google LLC (“Google”), Samsung Electronics Co., Ltd. and Samsung Electronics America, Inc. (“Samsung”) (collectively “Petitioners”) request *inter partes* review of Claims 1-18 of U.S. Patent No. 11,393,461 (“the ’461 Patent”).

II. GROUNDS FOR STANDING (37 C.F.R. § 42.104(A))

Petitioners certify that the ’461 Patent is available for *inter partes* review and that Petitioners are not barred or estopped from requesting *inter partes* review to challenge the claims on the grounds herein.

III. STATEMENT OF PRECISE RELIEF REQUESTED FOR EACH CLAIM CHALLENGED

For purposes of this proceeding only, Petitioners apply the earliest alleged priority date of the ’461 Patent: March 12, 2013. Accordingly, pre-AIA 35 U.S.C. §102 applies. Petitioners respectfully request review and cancellation under 35 U.S.C. § 311 of Claims 1-18 (“Challenged Claims”) in view of:¹

Reference	Date(s)	Prior Art At Least Under
US9031847 (“Sarin”) (Ex-1005)	11/15/2011 (filed) 5/16/2013 (published)	§102(e)

¹ Petitioners do not concede that any Challenged Claims satisfy other requirements for patentability that cannot be raised in IPR.

Reference	Date(s)	Prior Art At Least Under
US20130110521(“Hwang”) (Ex-1006)	5/30/2012 (filed) 5/2/2013 (published)	§102(e)
US7475012 (“Garner”) (Ex-1007)	1/6/2009 (issued)	§102(b)
US20120265528 (“Gruber”) (Ex-1007)	9/30/2011 (filed) 10/18/2012 (published)	§§102(a), (e)
US20120034904 (“LeBeau”) (Ex-1016)	08/6/2010 (filed) 02/09/2012 (published)	§§102(a), (b), (e)

Petitioners request review and cancellation of the Challenged Claims as unpatentable based on the following grounds, supported by a declaration from Mr. Christopher Schmandt. Ex-1002 ¶¶1-38, 48-73; Ex-1003; Ex-1020.

Ground	Claims	Basis
1	1-4, 6, 7, 11, 14, 16, 18	§103: Sarin, Hwang
2	13	§103: Sarin, Hwang, Garner
3	3, 5, 7, 9, 10, 12, 15, 17	§103: Sarin, Hwang, Gruber
4	8	§103: Sarin, Hwang, Lebeau

IV. OVERVIEW OF THE '461 PATENT

A. Specification

The '461 Patent is generally directed to techniques used by mobile devices to detect voice commands in received audio in a way that minimizes the device's power consumption. Ex-1001, 6:17-20. To this end, the patent discloses "a hierarchical approach to monitoring and processing acoustic input that applies relatively low power processing to received acoustic information first and generally higher power processing as needed to evaluate the likelihood that the acoustic information includes a voice command." *Id.*, 6:21-26.

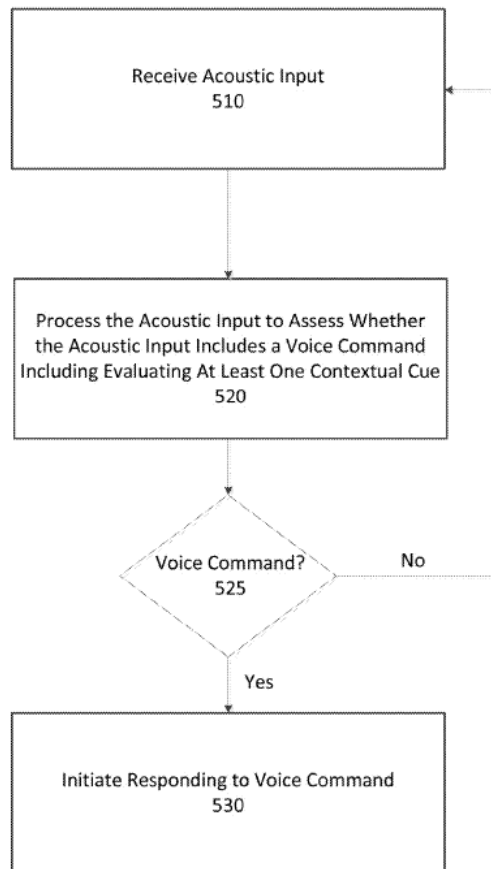


FIG. 5

As shown in Figure 5, “contextual cues” are used to help the mobile device detect whether received acoustic input includes a voice command. *See id.*, 2:29-31, 20:35-44. Example contextual cues include “the time of day, the location of the mobile device, motion of the mobile device, recent activity on the mobile device, biometric information, status of the power source, information about the acoustic environment, etc.” *Id.*, 20:45-49; Ex-1002 ¶¶39-43.

B. Prosecution History

The '461 Patent was filed on March 12, 2013 as U.S. Application 13/795,933 (“the '933 App.”). The '933 App. remained pending for over nine years and the claims were rejected ten times. On October 30, 2015, the applicant amended the independent claims to replace “low power mode” with “idle mode,” and to recite that the at least one contextual cue is used “while the mobile device is operating in the idle mode.” Ex-1004, 478-482. On April 4, 2016, the applicant further amended the independent claims to recite specific types of contextual cues. *Id.*, 535-539.

On March 3, 2022, the applicant overcame a double patenting rejection by filing a terminal disclaimer over U.S. Patent No. 9,361,885, and canceled one of the independent claims to obviate a §101 rejection. *Id.*, 1457-1459. The Office subsequently issued a Notice of Allowance on March 16, 2022. *Id.*, 1463-1471. This Petition presents arguments not considered during prosecution. *See* Ex-1004; Ex-1002 ¶¶44-47.

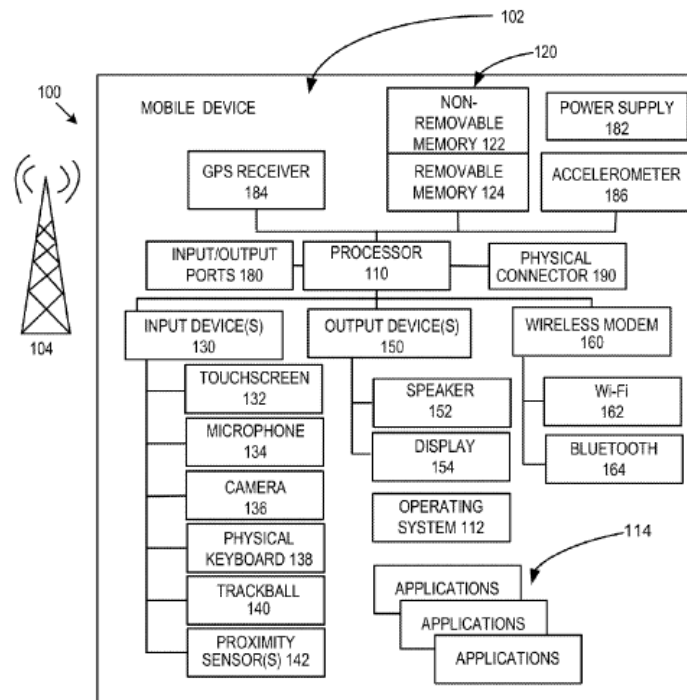
V. OVERVIEW OF THE PRIMARY PRIOR ART REFERENCES

The prior art references analyzed in the grounds below are analogous to the '461 Patent because they are within the same field of endeavor as the '461 Patent and reasonably pertinent to one or more problems addressed by the '461 Patent. Ex-1002 ¶74.

A. Sarin (Ex-1005)

Sarin discloses a mobile device capable of operating in a “low-power” state and an “active” state with a camera that captures images based on voice commands. Ex-1005, 1:43-44; Ex-1002 ¶¶75-82. Sarin’s mobile device receives and recognizes image capture-related voice commands while the device is operating in a low-power state. Ex-1005, 13:38-46; *see also id.*, 6:39-44 (“speech recognition also can be used for other applications and functionality such as voice calls ... and search functions (e.g., to perform searches”). According to Sarin, “in a low-power state, the display is typically dimmed or powered off, and other components (e.g., hardware buttons, keypads, trackballs) can be deactivated, as well.” *Id.*, 9:63-65. If the device determines that a received voice command matches an expected command associated with an image capture application, the device transitions from a low-power state to an active state. *Id.*, 13:53-57, 13:57-62. Once in the active state, the mobile device’s display is typically powered on, and the device’s operating system may be started to activate an image capture application to capture images. *Id.*, 9:46-49, 10:53-55, 13:62-14:6, Fig. 10.

FIG. 1

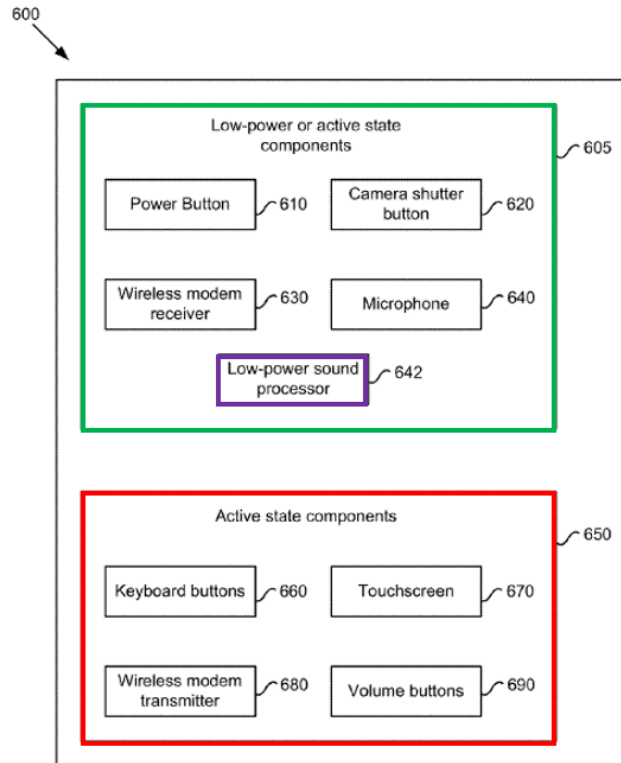


As shown in Figure 1, Sarin’s mobile device includes one or more primary processors 110, which perform “signal coding, data processing, input/output processing, power control, and/or other functions.” Ex-1005, 4:15-20, 6:19-22. As shown below in Figure 6, to enable voice command recognition while in a low-power state, Sarin’s mobile device utilizes low-power sound processor 642 (purple), which is functional when the mobile device is in the low-power state or active state (605, green). *Id.*, 11:10-15, 11:22-25.² Sarin’s mobile device can also recognize voice commands when it is in an active state (650, red) (Ex-1005, 10:3-7, 10:22-26),

² All emphases/annotations/color are added unless noted otherwise.

such that voice command recognition can be performed “at a remote server; on a general purpose processor, or on a dedicated special-purpose processor.” *Id.*, 13:6-8.

FIG. 6



To recognize image capture-related voice commands in a power efficient manner, Sarin, like the '461 Patent, uses contextual cues to disambiguate voice commands that are applicable to different functions which may be invoked with the same word or phrase. Ex-1005, 9:9-13, 11:32-34. Exemplary contextual cues include “which applications or functions are active on a device, time of day, and location.” *Id.*, 9:42-43.

B. Hwang (Ex-1006)

Hwang teaches techniques for processing audio in an electronic device (e.g., wireless telephone) with reduced power consumption. Ex-1006 ¶¶5, 26; Ex-1002 ¶¶83-90. Hwang discloses a low-power system for receiving audio, extracting and analyzing feature data from the received audio, and activating components of the device, such as the application processor, based on the analysis. Ex-1006 ¶¶5, 26. As shown in Hwang’s Figure 1, CODEC 120 receives “raw” audio data and extracts audio features from the audio data to store in buffer 140, which is then analyzed by DSP 150. Ex-1006 ¶¶24, 26.

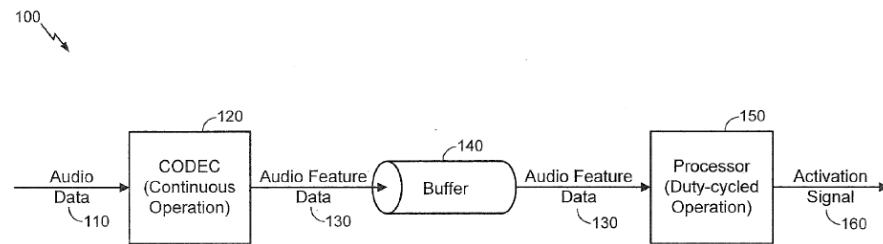


FIG. 1

As shown in Figure 2 below,³ DSP 230 includes sound recognition modules (241-245) that analyze the audio feature data extracted by CODEC 220. Ex-1006 ¶38. For example, continuous keyword detection module 243 may “detect the existence of pre-enrolled (e.g., predetermined) keyword sets” in a relatively low-

³ CODEC 220 and DSP 230 in Figure 2 correspond to CODEC 120 and DSP 150 shown in Figure 1. Ex-1005 ¶33.

power state and “activate predefined applications based on detected keywords.” *Id.*

¶41. Hwang also discloses that its DSP can perform the audio processing functions performed by its CODEC (e.g., audio feature extraction). *Id.* ¶52; Fig. 3.

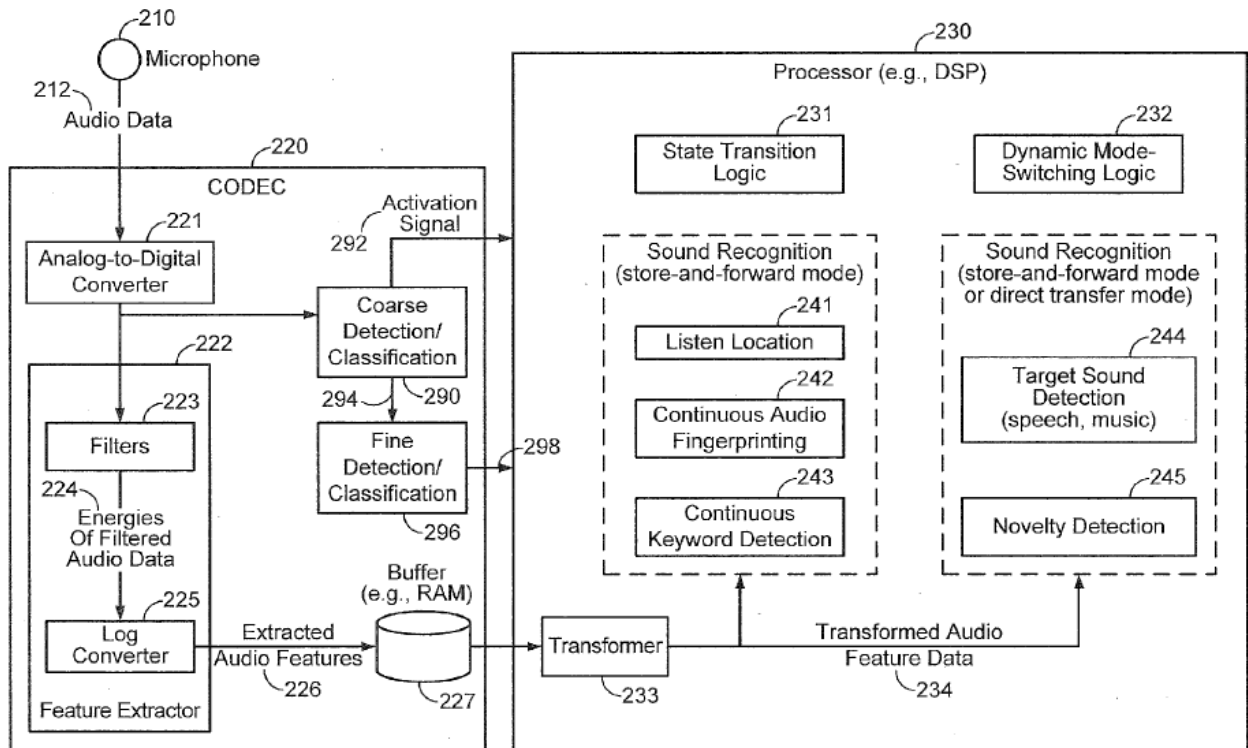


FIG. 2

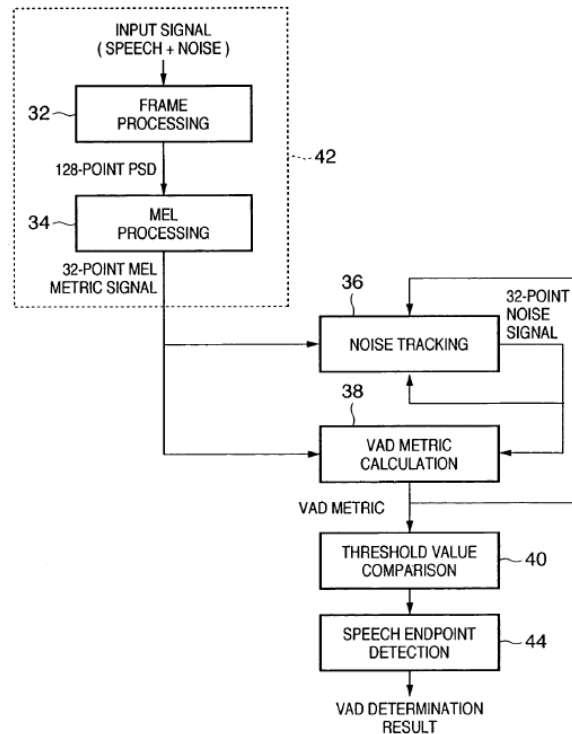
Based on its analysis, the DSP generates activation signal 160 to activate various components of the device, such as “an application processor and/or component(s) of a mobile station modem.” Ex-1006 ¶¶31, 44. For example, DSP 150 may generate the activation signal when “the analysis of the audio feature data 130 identifies a particular voice input command (e.g., ‘wake up’).” *Id.* ¶31. Hwang’s

DSP may also provide “higher-level applications” with the results of its sound recognition operations. *Id.* ¶45.

C. Garner (Ex-1007)

Garner teaches signal detection techniques designed to be robust against background noise for voice activity detection (VAD). Ex-1007, 2:65-67; Ex-1002 ¶¶91-94. As shown below in Figure 4, Garner divides an acoustic signal into frames, performs mel processing on the frames, and calculates a likelihood ratio that a frame is noise or voice. Ex-1007, 5:55-6:46. The likelihood ratio is compared with a predetermined threshold value to generate a value indicating speech or noise (i.e., VAD determination result). *Id.*, 11:24-26.

FIG. 4



D. Gruber (Ex-1008)

Gruber teaches a “virtual assistant” implemented on a smartphone (Ex-1008, ¶74), which uses “context information” to supplement natural language input (spoken and/or text) from a user. Ex-1008, Abstract, ¶11; Ex-1002 ¶¶95-99. The virtual assistant “interprets user inputs, operationalizes the user’s intent into tasks and parameters to those tasks, executes services to support those tasks, and produces output that is intelligible to the user.” Ex-1008 ¶11.

Gruber’s use of context information “helps to clarify the user’s intent and to reduce the number of candidate interpretations of the user’s input, ... reduc[ing] the need for the user to provide excessive clarification input.” *Id.* ¶13. Gruber discloses that “[c]ontext can include any available information that is usable by the assistant to supplement explicit user input to constrain an information-processing problem and/or to personalize results.” *Id.*; *see also id.* ¶178.

Examples of Gruber’s context information used in processing speech input include: applications currently in use on a device (Ex-1008 ¶180); information stored on a device (*id.* ¶193); a user’s dialog history with the virtual assistant (*id.* ¶203); sensor data that reports the user’s current location, the local time, the position, orientation, and motion of the device, and environmental data (*id.* ¶218); and events, such as incoming texts, emails, phone calls, and alerts (*id.* ¶238).

E. Lebeau (Ex-1016)

Lebeau teaches a mobile device that operates in two modes of operation. Ex-1016, Abstract; Ex-1002 ¶100. Lebeau determines to “switch” the device to receive voice input indicating a request to perform an operation. Ex-1016 ¶9. Lebeau’s device attempts to infer whether, based on the current context, monitoring for voice input would provide at least a threshold level of convenience for the user and, if so, monitors for voice input. Ex-1016 ¶23. Lebeau explains that a context includes whether the device’s battery is “running low on stored power (e.g., below 25% charge remaining in the battery)” (Ex-1016 ¶24) or the device is receiving power from a power cord. Ex-1016 ¶30.

VI. LEVEL OF ORDINARY SKILL IN THE ART

A person of ordinary skill in the art (“POSITA”) at the time of the alleged invention of the ’461 Patent, which for purposes of this proceeding is no earlier than March 12, 2013, would have possessed a Bachelor of Science (or equivalent) in electrical engineering or computer science, and two years of work experience in the field of speech and natural language interfaces for interacting with and directing computer systems and software, including speech recognition and processing. A greater amount of education could compensate for less work experience, and vice versa. Ex-1002 ¶¶32-33.

VII. CLAIM CONSTRUCTION

Petitioners are unaware of any “prior claim construction determination” related to the ’461 Patent. *See* 37 C.F.R. §42.100(b). Given the close correlation and substantial identity between the prior art references and the Challenged Claims, Petitioners believe that no express constructions of the claims are necessary to assess whether the prior art reads on the Challenged Claims. Thus, the claims should be given their plain and ordinary meaning.⁴ *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc); Ex-1002 ¶101.

VIII. DETAILED EXPLANATION OF GROUNDS

A. Ground 1: Sarin in Combination with Hwang Renders Obvious Claims 1-4, 6, 7, 11, 14, 16, and 18

1. Motivation to Combine Sarin and Hwang

A POSITA would have been motivated to combine Sarin and Hwang (“Sarin-Hwang”) in the manner discussed in the Grounds presented here, for the reasons discussed in this section, and for the additional reasons specifically discussed below. Ex-1002 ¶¶102-110.

A POSITA also would have been motivated to combine Sarin with Hwang because they are analogous art and teach improvements to the same type of devices.

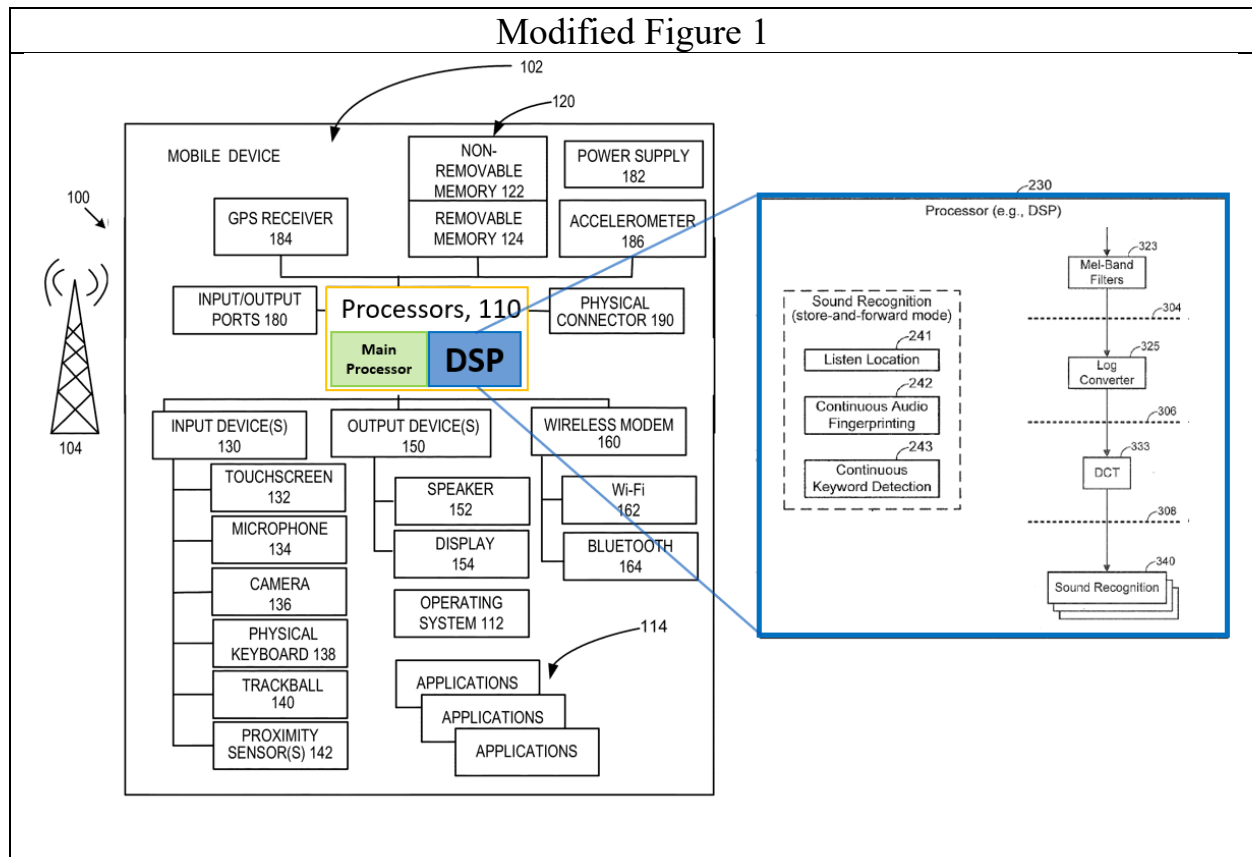
⁴ Petitioners reserve all rights to raise claim construction arguments and other arguments in any parallel or future litigation concerning the ’461 Patent. For example, comparing the claims to the accused products in the litigation may raise controversies that require construction of certain claim terms.

Sarin teaches a mobile device that uses contextual cues to assist in recognizing voice commands (such as image capture commands), which reduces power consumption and can extend battery life. Ex-1005, 11:32-34. Sarin also reduces power consumption through its use of a low-power sound processor. *Id.*, 10:65-11:1. Hwang similarly teaches techniques for processing audio in an electronic device (such as a wireless telephone) with reduced power consumption. Ex-1006 ¶¶5, 26. For example, Hwang's low-power system receives audio, extracts and analyzes feature data from the received audio, and activates components of the electronic device based on the analysis. Ex-1006 ¶5. Specifically, Hwang's feature extractor includes "a plurality of filters 223 that filter the audio data 212 to generate energies 224 (e.g., mel-band energies) of filtered audio data." Ex-1006 ¶35. Therefore, Sarin and Hwang both listen and perform sound recognition with a goal of reducing power consumption. Ex-1002 ¶¶103-105.

Hwang offers additional benefits to Sarin's mobile device. First, while Sarin's sound processor is functional when the mobile device is in a low-power state (Ex-1005, 11:9-19), where the display and other components are deactivated (*id.*, 9:62-65), Sarin leaves the implementation details of its low-power sound processor to a POSITA. A POSITA would have been motivated to look to references that include details for implementing a low-power sound processor that meets Sarin's goal of reducing power consumption, such as Hwang. Ex-1002 ¶105. To address this

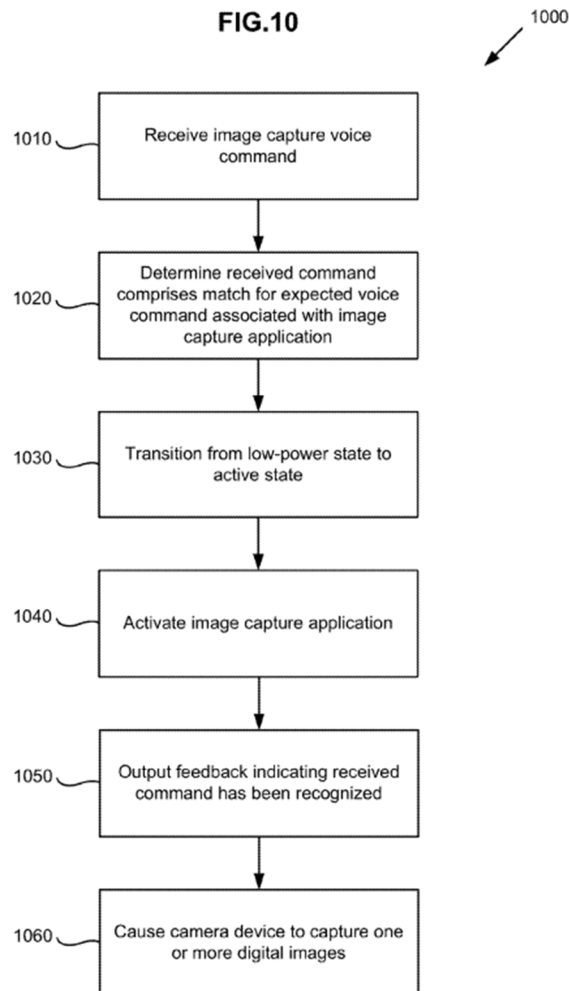
problem, Hwang detects voice commands in audio in a way that reduces power consumption (Ex-1006 ¶5) by using a DSP that receives and processes audio input while other mobile device components (such as an application processor) are not activated. *Id.* ¶26.⁵

A POSITA would have thus been motivated to implement Sarin's low-power sound processor using Hwang's DSP, as shown below in an example modified Figure 1 of Sarin (Ex-1002 ¶106):



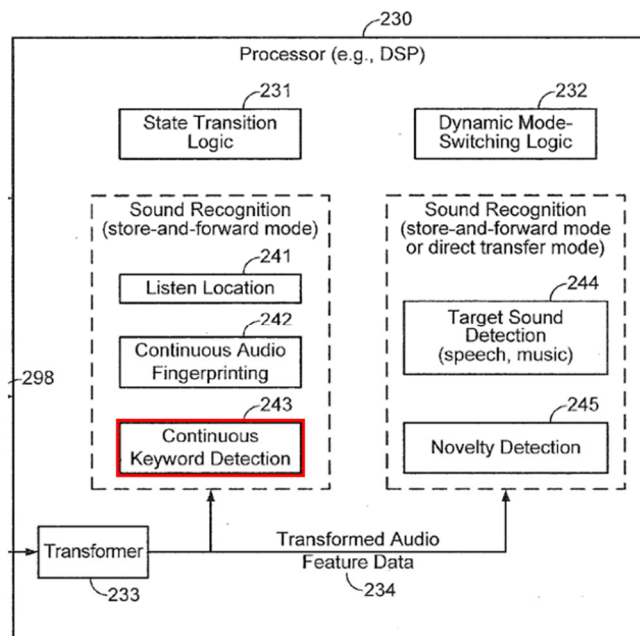
⁵ As discussed in §V.B, Hwang discloses that the audio processing functions of the CODEC and DSP can be consolidated in one DSP. Ex-1006 ¶52, Fig. 3.

A POSITA would have recognized that Hwang's DSP is compatible with and meets the functional requirements of Sarin's low-power sound processor. *Id.* ¶107. As shown in Figure 10, Sarin teaches a low-power state voice command recognition, focusing on voice commands associated with image capture:



A POSITA would have understood that Hwang's DSP is capable of receiving and recognizing the type of image capture-related commands that Sarin's low-power sound processor would need to recognize. Ex-1002 ¶108. For example, as highlighted below, Hwang's DSP includes a "continuous keyword detection

module,” which operates “in a relatively low-power state and may activate predefined applications” based on “predetermined keywords.” Ex-1006 ¶41. A POSITA would have recognized that Hwang’s detected “keyword sets” are similar to Sarin’s image capture-related voice commands and that Sarin’s image capture application is an example of a predefined application that Hwang’s DSP could activate after detecting image capture-related commands. Ex-1002 ¶108; Ex-1005, 13:62-63 (after transitioning to an active state, “the computing device activates the image capture application.”).



Ex-1006, Fig. 2 (excerpt)

Based on Hwang’s teachings, a POSITA would have been motivated to incorporate Hwang’s teachings regarding its DSP into Sarin’s mobile device because

Hwang furthers Sarin's goal of conserving power in a mobile device. Sarin recognizes that, aside from the display, *other components* may be deactivated in a low-power state, and that transitioning out of a low-power state may entail *starting the operating system*. A POSITA would have recognized that the operating system typically runs on the primary processors and, thus, Sarin teaches changing the power state and/or usage of the primary processor. For this additional reason, a POSITA would have been motivated to apply Hwang's teachings and perform low-power voice command detection while Sarin's main processor is deactivated. Ex-1002 ¶109. A POSITA would have recognized that receiving and processing audio using the DSP while the primary processor is inactive is desirable in a mobile device that runs on a battery, as maintaining the primary processor in a deactivated state until the processor is needed would tend to conserve power usage and extend the battery's life (a goal of Sarin). Ex-1005, 11:32-34; Ex-1002 ¶109.

Further, a POSITA would have had a reasonable expectation of success in combining Sarin and Hwang because a POSITA would have recognized that Hwang's DSP meets the functional requirements of Sarin's low-power sound processor and could be readily incorporated into Sarin's mobile device. Ex-1002 ¶110. A POSITA would have understood, therefore, that these references disclose interrelated teachings based on routine technologies that would have been amenable

to various well-understood and predictable combinations to provide quick and efficient speech recognition and processing. *Id.*

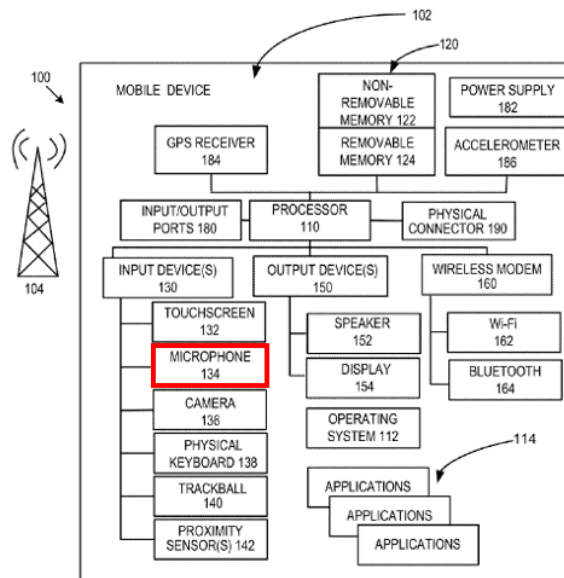
2. Independent Claim 1

- a. 1[pre]: A method of monitoring an acoustic environment of a mobile device for voice commands when the mobile device is operating in a low power mode, the method comprising:

To the extent the preamble is limiting, Sarin teaches this element.⁶ Sarin teaches monitoring an acoustic environment of a mobile device for voice commands when the mobile device is in a low power mode. Ex-1002 ¶¶111-115. As shown in Figure 1 below, Sarin teaches a mobile device with microphone 134 (Ex-1005, 4:10-14, 4:50-54), which allows the user to speak a voice command that cause the mobile device to capture an image. Ex-1005, 7:49-52. Sarin’s mobile device “*receives the image capture voice command while in a low-power state.*” *Id.*, 13:48-53; Ex-1002 ¶113.

⁶ Petitioner uses the term “teaches” as including both express teachings or those fairly suggested to a person of ordinary skill in the art. *In re Baird*, 16 F.3d 380, 383 (Fed. Cir. 1994); *In re Keller*, 642 F.2d 413, 425 (CCPA, 1981) (“The test for obviousness is ... what the combined teachings of the references would have suggested to those of ordinary skill in the art.” (citations omitted)).

FIG. 1



A POSITA would have understood that Sarin’s mobile device’s “low-power state” is similar to the ’461 Patent’s “low power mode.” Ex-1002 ¶114. Sarin describes its “low-power state” as a state where “the display is typically dimmed or powered off, and other components (e.g., hardware buttons, keypads, trackballs) can be deactivated, as well.” Ex-1005, 9:62-65. Similarly, the ’461 Patent teaches that “[e]ntering a low power mode may include turning off the display, shutting down one or more processors (e.g., disabling the clock tree associated with one or more processors) and/or deactivating components that would otherwise draw power if left operational.” Ex-1001, 5:5-9; Ex-1002 ¶114.

Accordingly, Sarin teaches a mobile device that listens for voice commands (“a method of monitoring an acoustic environment of a mobile device for voice

commands”) while the mobile device is operating in a low-power state (“when the mobile device is in a low power mode”). Ex-1002 ¶115.

b. 1[a]: receiving acoustic input from the environment of the mobile device while the mobile device is operating in idle mode; and

Sarin-Hwang teaches this element. Sarin-Hwang’s mobile device receives acoustic input from its environment while operating in idle mode. Ex-1002 ¶¶116-121. As discussed in §VIII.A.1, Sarin-Hwang’s DSP is functional when the primary processor is not activated.⁷ Specifically, Hwang achieves reduced power consumption in a mobile device by incorporating a DSP, which receives and processes audio input while other mobile device components (such as an application processor) are not activated. Ex-1006 ¶26. Hwang’s DSP includes a “continuous keyword detection module,” which, “receive[s] sound input and may detect the existence of *pre-enrolled (e.g., predetermined) keyword sets. Continuous keyword detection may be performed in a relatively low-power state and may activate*

⁷ As explained in §VIII.A.2.a, Sarin’s mobile device receives acoustic input and interprets voice commands (such as commands relating to image capture), while operating in a low-power state in which “the display is typically dimmed or powered off, and other components (e.g., hardware buttons, keypads, trackballs) can be deactivated, as well.” Ex-1005, 9:63-65. Sarin also teaches that, when its mobile device recognizes a voice command, the device can “transition to an active state, start the operating system, display a GUI for the tool, and generate command events to cause one or more digital images to be captured by an image capture device.” *Id.*, 10:52-57. Thus, Sarin’s sound processor is functional when the mobile device is in a low-power state. *Id.*, 11:9-19; Ex-1002 ¶117 n.6.

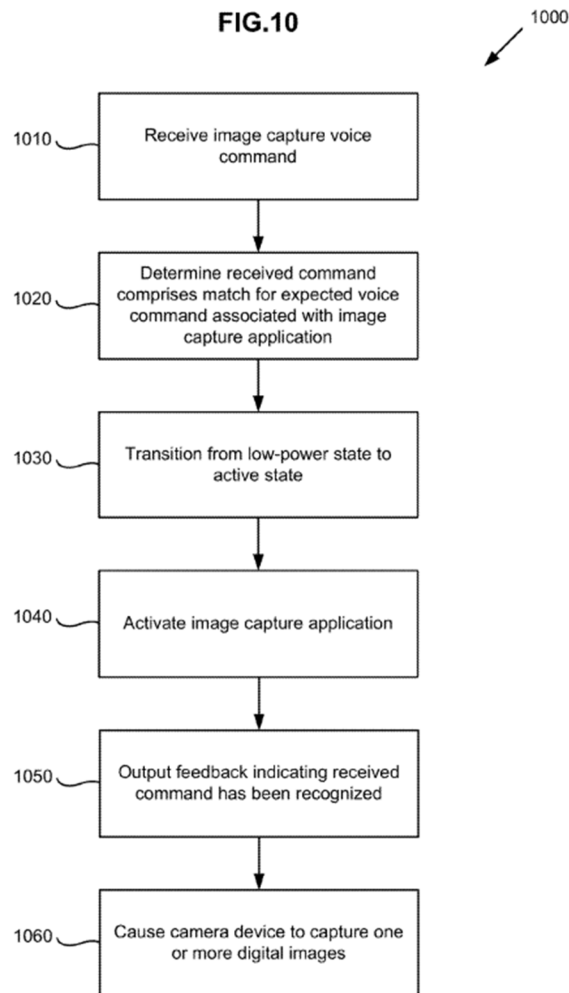
predefined applications based on detected keywords.” Ex-1006 ¶41. Hwang’s DSP further includes sound recognition modules that analyze audio feature data and, based on the analysis, “determine[s] whether to *activate an application processor* and/or component(s) of a mobile station modem.” Ex-1006 ¶44. In mobile devices, an “application processor” typically refers to the device’s primary (or main) processor. Ex-1002 ¶¶117 (discussing Exs-1009, 1010). Accordingly, a POSITA would have understood that Hwang teaches detecting voice commands while a mobile device’s primary processor is not activated to reduce power consumption and extend battery life. Ex-1002 ¶118.

When Sarin-Hwang’s primary processor is not activated, the mobile device consumes less power and is in an idle mode. The ’461 Patent states that, “when a mobile device enters a low power mode (*e.g., powers down into a sleep or idle mode*), *the primary processor may be powered down* to a desired extent.” Ex-1001, 26:46-48. Based on this disclosure, “idle mode,” as used in the ’461 Patent, encompasses using less power, including by powering down a primary processor. Ex-1002 ¶119-120.

Accordingly, Sarin-Hwang teaches a mobile device that receives audio (“receives acoustic input from the environment of the mobile device”) while consuming less power as the primary processor of the mobile device is not activated (“operating in an idle mode”). *Id.* ¶121.

c. 1[b]: in response to receiving the acoustic input, determining whether the received acoustic input includes a voice command by processing the received acoustic input,

Sarin-Hwang teaches this element. Sarin-Hwang teaches to determine whether the received acoustic input includes a voice command. Ex-1002 ¶¶122-125. As shown in Figure 10 below, Sarin “receives an image capture voice command” at step 1010, and in response, at step 1020, “determines that the received command comprises a match for an expected command associated with an image capture application.” Ex-1005, 13:47-55; Ex-1002 ¶123.



Sarin-Hwang’s mobile device “process[es] the received acoustic input.” Ex-1002 ¶124. For example, Hwang’s DSP includes a “continuous keyword detection module 243” that “may receive sound input and may detect the existence of pre-enrolled (e.g., predetermined) keyword sets.” Ex-1006 ¶41. As discussed in §VIII.A.1, a POSITA would have understood that Sarin’s image capture-related voice commands would be included among the “keyword sets” recognized by Hwang’s DSP. Ex-1002 ¶124. Detecting keywords in audio input by the continuous keyword detection module is consistent with the “processing” described in the ’461

Patent. *See, e.g.*, Ex-1001, 30:56-62 (“speech processing stages may include ASR ... acoustic input 705 may undergo limited vocabulary ASR to perform ***keyword spotting***”).

Accordingly, Sarin-Hwang’s mobile device recognizes predetermined keywords corresponding to image capture-related voice commands (“determining whether the received acoustic input includes a voice command”) in received audio input (“in response to receiving the acoustic input”), using a continuous keyword detection module (“by processing the received acoustic input”). Ex-1002 ¶125.

d. 1[c]: the processing employing at least one contextual cue, the at least one contextual cue comprising information other than content of the acoustic input and comprising a time the acoustic input was received, motion of the mobile device, location of the mobile device, occurrence of an event on the mobile device, recent activity on the mobile device, a stored sample of the acoustic environment of the mobile device and/or a power source of the mobile device,

Sarin-Hwang teaches this element. Sarin teaches several “contextual cue[s] comprising information other than content of the acoustic input” that may assist in recognizing voice commands. Ex-1002 ¶¶126-129. As explained in §VIII.A.2.b, Sarin-Hwang’s low-power DSP processes the acoustic input while the device is in an idle mode. Sarin further teaches that “voice-controlled applications” deployed on its mobile device “can use ***contextual cues*** to assist in recognizing commands and

using resources (e.g., battery life) efficiently.” Ex-1005, 11:32-34; *see also id.*, 9:9-12 (contextual cues used to disambiguate voice commands). For example, Sarin teaches contextual cues comprising: a “time the acoustic input was received” (Ex-1005, 11:42-46), “motion of the mobile device” (*id.*, 12:19-24 (a gyroscope or an accelerometer for additional contextual cues)), “location of the mobile device” (*id.*, 11:42-46), “occurrence of an event on the mobile device” (*id.*, 11:47-52 (if “a voice call is active on a smart phone” or “whenever the camera is not active”)), and “recent activity on the mobile device” (*id.*).

A POSITA would have understood that Sarin’s contextual cues are similar to the contextual cues described in the ’461 Patent, including information other than content of the acoustic input in the same way. *See also* Ex-1001, 21:29-33 (time of day), 21:5-10 (mobile device’s location in the home or office), 22:41-47 (response from an accelerometer); Ex-1002 ¶128.

Accordingly, Sarin-Hwang’s mobile device processes audio input using a contextual cue (“employing at least one contextual cue”) comprising a number of the specific contextual cues required by the claim. Moreover, as the claim is drafted, only one contextual clue is required, yet Sarin discloses several. Ex-1002 ¶129.

- e. 1[d]: the processing of the received acoustic input comprising a plurality of processing stages, at least one of the processing stages being performed while the mobile device is operating in the idle mode.**

Sarin-Hwang teaches this element. Sarin-Hwang teaches at least one processing stage being performed while the mobile device is in the idle mode. Ex-1002 ¶¶130-133. As shown in Figure 3, Sarin-Hwang's DSP "may perform feature extraction (e.g., via mel-band filters 323 and a log converter 325), data transformation (e.g., via a DCT transformer 333), and sound recognition (e.g., via sound recognition modules 340)." Ex-1006 ¶52. Each of these are "processing stages" because they are stages used in processing the received acoustic input to extract identifiable features. Ex-1002 ¶131.

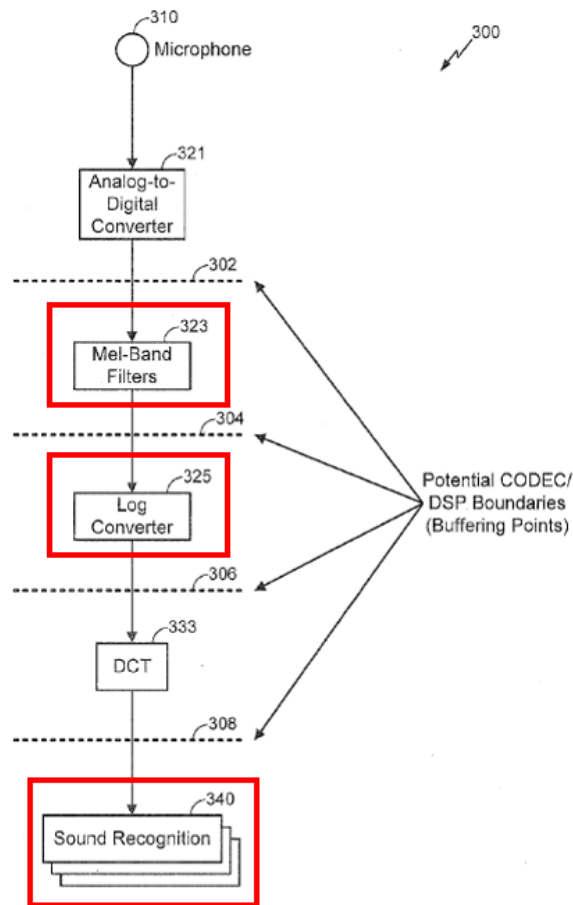


FIG. 3

At least one of the above processing stages is performed when the mobile device is operating in idle mode. Ex-1002 ¶132. As discussed in § VIII.A.2.b, Sarin-Hwang’s DSP receives and processes audio while the mobile device’s primary processor is not activated and the device is in an idle mode as it consumes less power. Specifically, both feature extraction and sound recognition are processing stages performed prior to the DSP activating the mobile device’s primary processor. Ex-1006 ¶¶44 (“processor 230 may analyze the transformed audio feature data 234...,

and may determine whether to activate an application processor”), 53-55 (“Data transformation and sound recognition may be performed by the DSP.”).

Accordingly, Sarin-Hwang’s mobile device extracts audio features from audio input and performs keyword detection on the extracted audio features (“processing of the received acoustic input comprising a plurality of processing stages”) while the device’s primary processor is not activated and consuming less power (“at least one of the processing stages being performed while the mobile device is operating in the idle mode”). Ex-1002 ¶133.

3. Dependent Claims 2-4, 6, 7, 11, 14, and 16

a. **Claim 2: The method of claim 1, further comprising transitioning the mobile device from the idle mode to an active mode in response to detecting that the acoustic input includes a voice command.**

Sarin-Hwang teaches this claim. Sarin-Hwang transitions the mobile device from idle to an active mode in response to a detected voice command. Ex-1002 ¶¶134-136. Specifically, Hwang’s DSP analyzes audio feature data 130 and generates “activation signal 160 to activate various components of an electronic device” when the audio feature data 130 “identifies a particular *voice input command* (e.g., ‘*wake up*’).” Ex-1006 ¶31; *see also id.* ¶¶41 (continuous keyword detection module enabling “voice-activation commands without the use of a dedicated voice command button or non-verbal user input.”), 44 (DSP “may analyze

the transformed audio feature data ... via one or more of the sound recognition modules 241-245, and *may determine whether to activate an application processor*").

Accordingly, Sarin-Hwang's mobile device activates its primary processor ("transitions the mobile device from the idle mode to an active mode") when it recognizes a "wake up" voice command ("in response to detecting that the acoustic input includes a voice command"). Ex-1002 ¶136.

b. Claim 3: The method of claim 1, wherein the at least one contextual cue includes the time the acoustic input was received, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on the time the acoustic input was received.

Sarin-Hwang teaches this claim. Sarin's contextual cues include the time the acoustic input was received and that time influences the evaluation of a processing stage. Ex-1002 ¶¶137-139. First, Sarin teaches that "voice-controlled applications" deployed on its mobile device "can use *contextual cues* to assist in recognizing commands and using resources (e.g., battery life) efficiently." Ex-1005, 11:32-34. Second, Sarin teaches different "contextual cues" that may assist in recognizing voice commands corresponding to image capture functionality, such as the time the acoustic input was received. *Id.*, 11:42-46 ("Other contextual clues include ... time of day."). A POSITA would have understood that Sarin's consideration of the "time

of day,” is a contextual cue that influences the processing stage when used to recognize voice commands corresponding to image capture functionality. Ex-1002 ¶138. For example, Sarin explains that the “time of day” cue can be used to influence the evaluation of the processing stage by, for example, turning off image capture-related voice recognition. Ex-1005, 11:44-46 (“if the time of day and location indicate that the user is at work, voice recognition relating to image capture operations can be turned off.”).

Accordingly, Sarin-Hwang’s mobile device uses time of day (“contextual cue includes the time the acoustic input was received”) to recognize voice commands corresponding to image capture functionality (“influencing evaluation ... based on the time the acoustic input was received.”). Ex-1002 ¶139.

c. Claim 4: The method of claim 1, wherein the at least one contextual cue includes recent activity on the mobile device, the method further comprising evaluation of at least one of the plurality of processing stages based on at least one of how recently the user activated the mobile device, how recently the user issued a voice command and how recently the user performed at least one action on the mobile device.

Sarin-Hwang teaches this claim. Sarin teaches that the contextual cues include recent activity on the mobile device. Ex-1002 ¶¶140-144. As explained in §VIII.A.2.d, Sarin teaches using contextual cues to assist in recognizing voice commands. Ex-1005, 11:32-34. Sarin’s contextual cues include “which applications

or functions are active on a device.” Ex-1005, 11:42-43. Sarin thus teaches a contextual cue of the recent activity on the mobile device. Ex-1002 ¶141; Ex-1001, 23:67-24:5 (“if a user opens a text or email ... [o]pening a calendar or scheduling application”).

Sarin teaches to base the evaluation on how recently the user performed at least one action on the mobile device. Sarin provides that if a “voice call is active on a smart phone, voice recognition relating to image capture operations can be turned off” to avoid “unintended camera operation.” Ex-1005, 11:47-53. Sarin’s consideration of whether a “voice call is active” is a contextual cue that influences the processing stage when used to recognize voice commands corresponding to image capture functionality. Ex-1002 ¶142. In particular, Sarin’s system considers whether a voice call is currently active to influence the evaluation of the processing stage by, for example, turning off the processing of image capture-related voice commands, until the call is finished. *Id.* A POSITA would have understood that the recently performed action of an active voice call is an example of considering “how recently the user performed at least one action on the mobile device.” *Id.* ¶¶142-143 (citing Ex-1014). Indeed, a user engaging in a currently active voice call is as recent as a user action can be.

Accordingly, Sarin-Hwang’s mobile device considers whether a voice call is currently active (“contextual cue includes recent activity on the mobile device”) to

turn off the processing of image capture-related voice commands until the call is finished (“evaluation of ... how recently the user performed at least one action on the mobile device”). *Id.* ¶144.

d. Claim 6: The method of claim 1, wherein the at least one contextual cue includes motion of the mobile device received as an input from an accelerometer of the mobile device, the method comprises influencing evaluation of at least one of the plurality of processing stages based on an assessment of whether the user moved the mobile device proximate in time with receiving the acoustic input based, at least in part, on the input from the accelerometer.

Sarin-Hwang teaches this claim. Sarin’s contextual cues include the mobile device’s motion received as an input from the mobile device’s accelerometer. Ex-1002 ¶¶145-152. Sarin provides an example of the mobile device using an accelerometer and/or gyroscope to detect “the orientation of the device as being ready perform an operation such as capturing an image or recording video.” Ex-1005, 12:19-24. A POSITA would have understood that Sarin thus teaches that the contextual cue includes “motion of the mobile device received as an input from an accelerometer” because Sarin receives input from other device components, such as accelerometers, which sense motion. Ex-1002 ¶146.

Sarin teaches that the processing stage is influenced by the accelerometer in assessing whether the user moved the mobile device (e.g., into a specific orientation)

proximate in time with receiving the acoustic input. Sarin's accelerometer and gyroscope detect the "orientation of the device," which is then used to determine that the device is "ready [to] perform an operation such as capturing an image or recording video." Ex-1005, 12:19-24. A POSITA would have understood that the device is "ready" to take a picture near in time to the end of the motion because a movement would have brought the device into the specific orientation. Ex-1002 ¶¶147-148 (citing Ex-1017 ¶¶22, 31). A POSITA would have thus understood that moving the device into a particular orientation results in motion that would influence the evaluation of the processing stage by ensuring that the device is "ready" to capture the picture near in time to when the motion ends, the device is oriented correctly, and the user issues the command to capture an image. Ex-1002 ¶¶148-151.

Accordingly, Sarin-Hwang's mobile device uses motion information from an accelerometer to determine the device's orientation ("contextual cue includes motion of the mobile device received as an input from an accelerometer") to determine whether the device is "ready" to take a picture near in time to the end of the motion because a movement would have brought the device into the specific orientation ("influencing evaluation ... based on an assessment of whether the user moved the mobile device proximate in time with receiving the acoustic input based... on the input from the accelerometer"). Ex-1002 ¶152.

- e. Claim 7: The method of claim 1, wherein the at least one contextual cue includes the location of the mobile device, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on determining whether the mobile device is located in at least one of a vehicle, a home or a place of work of the user of the mobile device.**

Sarin-Hwang teaches this claim. Sarin teaches that the user's work location is a contextual cue that influences the processing stage. Ex-1002 ¶¶153-155. Sarin teaches different "contextual cues" that may assist in recognizing voice commands corresponding to image capture functionality, such as when the "location indicate[s] that the user is at *work*." Ex-1005, 11:42-46. This "work" location is a cue used to influence the evaluation of the processing stage by, for example, turning off image capture-related voice recognition. *Id.* ("image capture operations can be turned off.").

- f. Claim 11: The method of claim 1, wherein the mobile device includes a primary processor and a secondary processor, and wherein at least one first processing stage of the plurality of processing stages is performed by the secondary processor prior to engaging the primary processor to process the acoustic input.**

Sarin-Hwang teaches this claim. Sarin teaches that its mobile device includes a primary processor. Ex-1002 ¶¶156-160. Sarin's mobile device includes "controller

or processor 110” for “signal coding, data processing, input/output processing, power control, and/or other functions.” Ex-1005, 4:15-20.

Sarin-Hwang teaches that the mobile device has a secondary processor. Ex-1002 ¶158. Sarin teaches that its mobile device may have more than one main processor. Ex-1005, 6:19-22. Moreover, as discussed in §VIII.A.1, Hwang teaches a device that includes a DSP, which processes and extracts features from audio input. *See, e.g.*, Ex-1006 ¶52. A POSITA would have understood that Hwang’s DSP is a distinct processor from the first processor because Hwang describes that the DSP may determine whether to activate a different application processor. Ex-1006 ¶44; Ex-1002 ¶158.

Sarin-Hwang’s secondary processor performs a first processing stage before engaging the primary processor to process the acoustic input. Sarin teaches to perform speech recognition to listen for and interpret voice input, on the local device. Ex-1005, 6:29-49, 13:15-20. Moreover, Sarin-Hwang’s DSP performs feature extraction, data transformation, and sound recognition *before* engaging a primary processor to further process audio. For example, the DSP “may analyze ... transformed audio feature data 234 via one or more of the sound recognition modules 241-245, and *may determine whether to activate an application processor ... based on the analysis.*” Ex-1006 ¶44. Because the DSP analyzes with sound recognition

modules before activating the application processor, the application processor is not engaged to process the acoustic input. Ex-1002 ¶159.

Accordingly, Sarin-Hwang's mobile device includes audio feature extraction, data transformation, and sound recognition ("at least one first processing stage of the plurality of processing stages") performed by a DSP ("performed by the secondary processor") and, based on the results of the sound recognition, may activate an application processor ("prior to engaging the primary processor to process the acoustic input"). Ex-1002 ¶160.

g. Claim 14: The method of claim 11, wherein performing the at least one first processing stage comprises performing limited vocabulary automatic speech recognition on the acoustic input to detect whether the acoustic input includes an explicit voice trigger

Sarin-Hwang teaches this claim. Sarin-Hwang teaches performing a limited vocabulary automatic speech recognition, as part of the first processing stage. Ex-1002 ¶¶161-164. As explained in §VIII.A.3.f (Claim 11), the first processing stage is performed by Sarin-Hwang's DSP. Sarin-Hwang's DSP includes a "continuous keyword detection module 243," which "receive[s] sound input and may detect the existence of pre-enrolled (e.g., predetermined) keyword sets." Ex-1006 ¶41. Sarin-Hwang's continuous keyword detection using "pre-enrolled (e.g., predetermined) keyword sets," is the claimed "limited vocabulary automatic speech recognition"

because it is a restricted set of predefined words, just like the '461 Patent. Ex-1001, 18:14-16 (“using a vocabulary restricted to the words in the explicit voice trigger phrase (which may include as few as a single word)”).

Sarin-Hwang teaches using limited vocabulary automatic speech recognition to detect whether the acoustic input includes an explicit voice trigger. For example, a keyword set recognized by the DSP may include a specific voice trigger such as “wake up,” which may cause the DSP to “generate [an] activation signal ... to activate various components of an electronic device.” Ex-1006 ¶31.

Accordingly, Sarin-Hwang’s mobile device includes a DSP with a continuous keyword detection module (“performing limited vocabulary automatic speech recognition on the acoustic input”) that recognizes keyword sets, such as “wake up” (“to detect whether the acoustic input includes an explicit voice trigger”). Ex-1002 ¶164.

h. Claim 16: The method of claim 1, wherein the mobile device is capable of connecting to at least one server via a network, and the method comprises transmitting at least a portion of the acoustic input to the at least one server via the network for processing by the at least one server.

Sarin-Hwang teaches this claim. Sarin-Hwang teaches that the mobile device connects to a server via a network. Ex-1002 ¶¶165-168. For example, Sarin teaches that the mobile device has a “wireless modem” which is coupled to an antenna and

supports “*two-way communications between the processor 110 and external devices.*” Ex-1005, 4:60-63. The wireless modem may be “a cellular modem for communicating with [a] *mobile communication network.*” Ex-1005, 4:63-65. Sarin also teaches that the mobile device’s speech recognition components can also be “implemented in whole or in part *on remote servers.*” *Id.*, 6:15-18.

Sarin-Hwang teaches transmitting a portion of the acoustic input for the server to process. For example, in performing voice command recognition for image-capture functionality, the mobile device or remote server (Ex-1005, 13:4-8) “compares ... received audio input with an expected audio command (e.g., a word or phrase) associated with an image capture operation.” *Id.*, 12:32-35. Sarin teaches that the decision “on where to perform speech recognition tasks can be based on,” among other things, “resources available at local devices and remote servers (e.g., processing power, memory or storage capacity).” *Id.*, 13:8-12. Thus, for recognizing more general voice commands, e.g. beyond those directed to image capture, a POSITA would have been motivated to transmit audio that includes such general voice commands to a remote server with sufficient processing power to process those commands. Ex-1002 ¶167; Ex-1005, 6:44-49 (“generalized voice input modes”).

Accordingly, Sarin-Hwang’s mobile device “is capable of connecting to at least one server via a network” because it has a wireless cellular modem to transmit received audio input over a network to a remote server for speech recognition

(“transmitting at least a portion of the acoustic input to the at least one server via the network for processing by the at least one server”). Ex-1002 ¶168.

4. Independent Claim 18

a. 18[pre]: A mobile device comprising:

Sarin teaches this element for the reasons discussed in §VIII.A.2.a. Ex-1002 ¶169.

b. 18[a]: at least one processor configured to:

Sarin-Hwang teaches this element for the reasons discussed in §VIII.A.2.b. Ex-1002 ¶170. In particular, as discussed in §VIII.A.1, Sarin-Hwang’s mobile device includes a low power DSP and an application processor. While Claim 18 does not recite the specific processor employing the following processing steps, Sarin-Hwang’s low-power DSP (identified as the “processor” for Claim 1) is being applied as Claim 18’s “at least one processor” in substantially the same way it is applied with respect to Claim 1.

(i) Elements 18[a(i)]-18[a(iii)]

Sarin-Hwang teaches Elements 18[a(i)]-18[a(iii)] for the reasons discussed in §§VIII.A.2.c-VIII.A.2.e, according to the table below. Ex-1002 ¶¶171-173.

Element	Section
18[a(i)]	VIII.A.2.c.
18[a(ii)]	VIII.A.2.d.

18[a(iii)]	VIII.A.2.e.
------------	-------------

B. Ground 2: Sarin in Combination with Hwang and Garner Renders Obvious Claim 13

1. Motivation to Combine Garner with Sarin-Hwang

A POSITA would have been motivated to combine the teachings of Sarin, Hwang, and Garner (“Sarin-Hwang-Garner”) and would have had a reasonable expectation of success in doing so because each relates to the same well-known technologies, such that they are analogous art and teach improvements to the same type of devices. Ex-1002 ¶¶174-178. Each teach using speech processing. Ex-1005, 6:29-49, 13:15-20; Ex-1006 ¶¶52 (“feature extraction..., data transformation..., and sound recognition”), 41 (“continuous key word detection”); Ex-1007, 1:45-51.

Garner teaches that voice activity detection (“VAD”) prevents a speech recognition process in an ASR unit “from recognizing background noise as speech” (Ex-1007, 1:45-47), which “prevent[s] an error of converting noise into a word.” *Id.*, 1:47-49. VAD “more skillfully manage[s] the throughput of the entire [automatic speech recognition] system ... that utilizes many computer resources.” *Id.*, 1:49-51. Thus, a POSITA would have understood that Garner’s VAD allows for more efficient use of device resources because it allows the device “to concentrate on other functions while speech recognition is not in progress and concentrate on ASR while the user utters.” *Id.*, 1:56-58; Ex-1002 ¶176. Indeed, a POSITA would have

understood that incorporating Garner's VAD as a lead-in to Sarin-Hwang's continuous keyword detection would have furthered Sarin-Hwang's goal of conserving device power by allowing keyword detection to focus on speech instead of noise. Ex-1002 ¶176.

A POSITA would have understood that Garner's VAD is compatible with, and could successfully be implemented on, Hwang's DSP. *Id.* ¶¶177-178. In particular, Hwang's feature extractor includes a plurality of mel-band filters that generate mel-band energies from input audio data, which, similar to Garner, is divided into frames. Ex-1006 ¶¶35-36. A POSITA would have understood that the mel-band energies produced in Hwang are comparable to the mel metric signals produced in Garner. Ex-1002 ¶178.

2. Dependent Claim 13

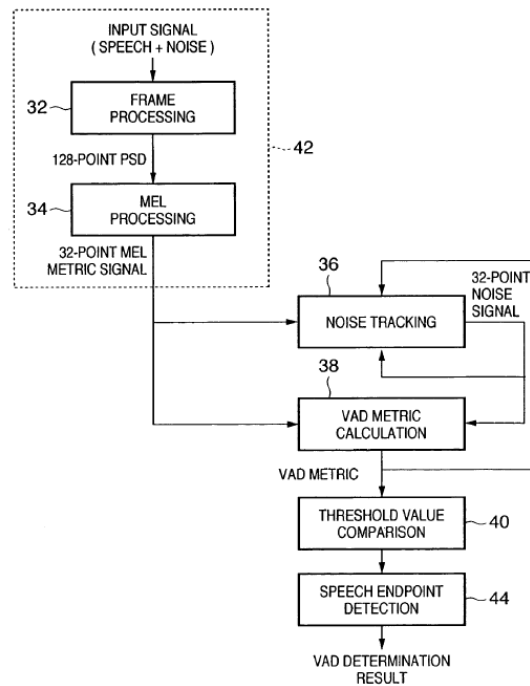
- a. Claim 13: The method of claim 11, wherein performing the at least one first processing stage includes performing at least one voice activity detection processing stage including performing at least one of spectral analysis on the acoustic input to evaluate whether a spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether signal periodicity of the acoustic input is indicative of voice activity, and phone loop analysis to evaluate whether the acoustic input includes speech.**

Sarin-Hwang-Garner teaches this claim. Sarin-Hwang-Garner teaches that its first processing stage includes performing voice activity detection including spectral

analysis on the acoustic input to evaluate whether the acoustic input is indicative of voice activity. Ex-1002 ¶¶179-193. As explained in §VIII.A.2.e (Element 1[d]), Sarin-Hwang's DSP "may perform feature extraction (e.g., via mel-band filters 323 and a log converter 325), data transformation (e.g., via a DCT transformer 333), and sound recognition (e.g., via sound recognition modules 340)." Ex-1006 ¶52. Moreover, Hwang's "target sound detection module 244" "detect[s] speech." Ex-1006 ¶42.

A POSITA would have understood that the detection of speech in input audio may be an example of VAD, as taught by Garner. Ex-1002 ¶180. As shown in Figure 4 below, Garner teaches a "voice activity detection processing stage" through its signal detection technique that uses VAD to determine whether a signal includes speech. *Id.*

FIG. 4



As shown in Figure 4, Garner processes an acoustic input into frames before mel processing module 34 generates mel metric signals, which are further processed to generate a “VAD metric.” Ex-1007, 5:55-56, 6:3-12, 6:25-26. The VAD metric indicates the likelihood that an acoustic frame is speech as opposed to noise. *Id.*, 6:31-46. In describing the VAD metric, Garner teaches that the speech variance calculation in a given audio frame uses a “spectral subtraction method.” Ex-1007, 6:60-67. Garner also discloses using “spectral magnitude” or “spectral power” in computing an estimate of speech variance in the frame. *Id.*, 6:57-60. Garner’s VAD technique thus is spectral analysis for evaluating whether the spectrum of the acoustic input is indicative of voice activity. Ex-1002 ¶¶181-183.

As discussed in §VIII.B.1, a POSITA would have been motivated to incorporate Garner’s VAD technique as a precursor to Sarin-Hwang’s keyword detection in order to conserve device power. Ex-1002 ¶¶184-192. As such, Garner’s VAD processing stage runs on Sarin-Hwang’s DSP (i.e., the “secondary processor”) and is therefore a “first processing stage.” *See* §VIII.A.3.f (Claim 11).

Accordingly, Sarin-Hwang-Garner’s mobile device includes a DSP that performs VAD to determine that audio input more likely comprises speech rather than noise before performing keyword detection on the audio input (“performing...voice activity detection processing stage”) that uses spectral subtraction, spectral magnitude, and spectral power (“performing spectral analysis on the acoustic input) to compute a likelihood that indicates whether audio input includes speech (“to evaluate whether a spectrum of the acoustic input is indicative of voice activity”). Ex-1002 ¶193.

C. Ground 3: Sarin in Combination With Hwang and Gruber Renders Obvious Claims 3, 5, 7, 9, 10, 12, 15, and 17

1. Motivation to Combine Gruber With Sarin and Hwang

A POSITA would have found it obvious to combine Gruber with Sarin and Hwang (“Sarin-Hwang-Gruber”) and would have had a reasonable expectation of success in doing so because each relates to the same well-known technologies such that they are analogous art and teach improvements to the same type of devices. Ex-

1002 ¶¶194-199. Each teach using speech processing to determine a voice command. Ex-1005, 6:29-49, 13:15-20; Ex-1006 ¶¶52, 41; Ex-1008 ¶11 (“interprets user inputs, operationalizes the user’s intent into tasks and parameters to those tasks”). Relatedly, Sarin and Gruber both teach using contextual cues. Ex-1005, 11:37-41; Ex-1008 ¶¶131, 354.

Gruber offers additional benefits to Sarin-Hwang’s mobile device, such as employing contextual cues when a general voice command is ambiguous. Ex-1002 ¶¶196-197. Particularly, Gruber receives user input, such as spoken language, processes the input, and generates output to the user and/or performs actions on behalf of the user. Ex-1008 ¶178. Gruber’s processing includes “speech recognition, natural language processing, task flow processing, and dialog generation.” *Id.* ¶14. Like Sarin, Gruber recognizes that contextual cues are advantageous because they assist in disambiguating voice commands. *Compare* Ex-1005, 9:9-12 (“desired functionality ... based on contextual cues”) *with* Ex-1008 ¶15 (“use context when ... disambiguating among competing interpretations”). Gruber additionally teaches “prompt[ing] the user for confirmation and/or request additional context information from any suitable source before calling a service ... to perform a function.” Ex-1008 ¶88. Additionally, like Sarin, Gruber considers “the acoustic environments in which the speech is entered.” Ex-1008 ¶250. Gruber stores “noise profiles” of the user’s

location since “noise profiles of a quiet office are different from those of automobiles or public places.” *Id.*

Gruber, therefore, addresses a stated goal of Sarin (disambiguation), and does so for voice commands beyond image capture functionality. Ex-1002 ¶¶197-198. A POSITA would have understood that implementing Gruber’s teachings on Sarin’s mobile device would result in a device that is better equipped to disambiguate and recognize general voice commands by “reduc[ing] the need for the user to provide excessive clarification input.” Ex-1008, Abstract; Ex-1002 ¶198.

A POSITA would also have understood that Gruber’s teachings are compatible with, and can be implemented on Sarin-Hwang’s combined system. *Id.* ¶199. Gruber explains that its software can be implemented on “CPU 62,” which may include one or more processors, such as microprocessor(s). Ex-1008 ¶77; *see also id.* ¶¶75-76, Fig. 28. A POSITA would have understood that Sarin-Hwang’s application processor is similar to the processor disclosed in Gruber, and, therefore, a POSITA would have understood that Sarin-Hwang would have had a reasonable expectation of success of implementing Gruber’s teachings. Ex-1002 ¶199.

2. Dependent Claims 3, 5, 7, 9, 10, 12, 15, and 17

a. Claim 3

Sarin-Hwang-Gruber teaches this claim. Sarin-Hwang-Gruber teaches that the contextual cue includes the time the acoustic input was received and that time

influences the evaluation of a processing stage. Ex-1002 ¶¶200-205. As explained above in §VIII.A.3.b (Claim 3), Sarin’s consideration of the “time of day,” is a contextual cue that influences the processing stage when used to recognize image-capture related voice commands.

Further, a POSITA would have been motivated to apply Gruber’s teachings to Sarin-Hwang to recognize more general voice commands based on the time acoustic input was received in order to disambiguate those commands. Ex-1002 ¶202. For example, Gruber teaches that “[t]ime information from clocks on client devices” can be used as contextual information for interpreting phrases within voice commands. Ex-1008 ¶132. Phrases such as “in an hour” and “tonight,” may be spoken in the context of user requests, and for which time information can be used to ascertain the meaning of the request. *Id.*; Ex-1002 ¶202.

Additionally, a POSITA would have understood, or at least found it obvious, that in a conversational system such as the virtual assistant implemented in Sarin-Hwang-Gruber’s mobile device (*see* Ex-1008 ¶17), which a user interacts with in real time, that phrases such as “in an hour” or “tonight” appearing in voice commands would be interpreted based on the time the voice command was received by the device. Ex-1002 ¶¶203-204.

Accordingly, Sarin-Hwang-Gruber’s mobile device includes a conversational virtual assistant that uses time (“contextual cue includes the time the acoustic input

was received”) to interpret certain phrases within received voice commands (“influencing evaluation...based on the time the acoustic input was received.”). *Id.* ¶205.

b. Claim 5: The method of claim 1, wherein the at least one contextual cue includes an occurrence of an event on the mobile device, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on whether the acoustic input is received proximate in time with the occurrence of the event.

Sarin-Hwang-Gruber teaches this claim. Ex-1002 ¶¶206-217. As explained above in §VIII.A.2.d, Sarin teaches that an occurrence of an event on the mobile device is a contextual cue. *See, e.g.*, Ex-1005, 11:47-52 (“if a voice call is active on a smart phone” or “whenever the camera is not active”).

A POSITA would have been motivated to use Sarin’s contextual cue and influence evaluation of at least one of the plurality of processing stages based on whether the acoustic input is received proximate in time with the occurrence of the event, in view of Gruber. Ex-1002 ¶¶208-213.

First, Gruber’s “virtual assistant 1002 is able to use context associated with *asynchronous events* that happen independently of the user’s interaction.” Ex-1008 ¶238. Exemplary asynchronous events include “incoming text messages or pages; incoming email messages; incoming phone calls; reminder notifications or task alerts; calendar alerts; alarm clock, timers, or other time-based alerts; notifications

of scores or other events from games; notifications of financial events such as stock price alerts; news flashes or other broadcast notifications; [and] push notifications from any application.” *Id.* ¶¶240-249.

Second, Gruber considers whether the acoustic input is received proximate in time with the occurrence of various events. For example, as shown in Figure 23, an incoming text message is an event that is used by the system (such as the message being received at 6:23 PM). Ex-1008 ¶238.



FIG. 23

Next, as shown below in Figure 24, an end user has spoken the command “call him.” *Id.* Gruber then “use[s] the event context to disambiguate the command by interpreting ‘him’ to mean the person who sent the incoming text message.” *Id.* Based on the time displayed (6:34 PM), a POSITA would have understood that the device received the voice command (“call him”) proximate in time with the event (the receipt of the 6:23 PM text message) and would use this proximity in time to determine to whom the command is directed to (i.e., to discern who “him” refers to in the command “call him”), as Gruber suggests. Ex-1002 ¶¶211-212. Specifically, Gruber teaches to “interpret instructions such as ‘send him a text message’, wherein the ‘him’ is interpreted according to context information derived from a *current* phone call.” Ex-1008 ¶18.



FIG. 24

Thus, a POSITA, in order to disambiguate the pronoun “him” in the voice command “call him,” would have been motivated to implement Gruber’s teachings to consider whether receipt of an event (e.g., a text message) was received contemporaneously with the voice command. Ex-1002 ¶¶213-215. This is just like the ’461 Patent, which provides an event example of the mobile device providing a “reminder to call John,” so that “the user may instruct the mobile device to ‘Call John Doe,’ to initiate a telephone call.” Ex-1001, 23:46-64 (providing certain events that trigger a voice response). The ’461 Patent further explains that “various system resources may be activated to, for example, check for recent activity, check for an

occurrence of an event, check the battery power, etc. to incrementally wake-up portions of the mobile device on an as-needed basis.” *Id.*, 16:26-30. A POSITA would have understood that, just like the ’461 Patent wakes-up portions of the device to check for an occurrence of an event, Gruber’s considers the event and timing of the event. Ex-1002 ¶¶215-216.

A POSITA would have been motivated to implement Gruber’s teachings of considering whether the asynchronous events occurred proximate in time with the receipt of audio input, in Sarin-Hwang’s mobile device because doing so would better address events that happen independently of the user’s interaction with the mobile device, and better evaluate the acoustic input for a precise, executable statement. Ex-1008, Abstract, ¶238; Ex-1002 ¶216.

Accordingly, Sarin-Hwang-Gruber’s mobile device uses events, such as a received text message (“contextual cue includes an occurrence of an event on the mobile device”) and considers whether the reception of the text message (i.e. the event) was close in time with the receipt of audio containing a voice command (“influencing evaluation ... based on whether the acoustic input is received proximate in time with the occurrence of the event”). Ex-1002 ¶217.

c. Claim 7

Sarin-Hwang-Gruber teaches this claim. Sarin-Hwang-Gruber teaches that the contextual cue includes the location of the mobile device and that location influences

the evaluation of a processing stage. Ex-1002 ¶¶218-224. As explained above in §VIII.A.3.e (Claim 7), Sarin’s consideration of the “location of the mobile device,” is a contextual cue that influences the processing stage of image-capture related voice commands.

Further, a POSITA would have been motivated to apply the teachings of Gruber to Sarin-Hwang to recognize more general voice commands based on the location of the mobile device in order to better understand these commands. Ex-1002 ¶¶220-223. For example, the virtual assistant implemented in Sarin-Hwang-Gruber’s mobile device uses “[l]ocation information coming from sensors or location-based systems” as contextual information. Ex-1008 ¶131 (using Global Positioning System). Virtual assistant combines location information with the explicit user input “to *detect when a user is at home*, based on known address information and current location determination.” *Id.* Thus, “*certain inferences may be made about the type of information the user might be interested in when at home* as opposed to outside the home, as well as the type of services and actions that should be invoked on behalf of the user depending on whether or not he or she is at home.” *Id.*

Accordingly, Sarin-Hwang-Gruber’s mobile device uses its GPS location data to assist in interpreting voice commands, (“contextual cue includes the location of the mobile device”), and considers whether a user is at home in making certain

inferences while interpreting the voice command (“influencing evaluation...based on determining whether the mobile device is located in ... a home”). Ex-1002 ¶224.

d. Claim 9: The method of claim 1, wherein the at least one contextual cue includes the sample of the acoustic environment obtained by the mobile device by storing acoustic input previously received by the mobile device and determined not to include a voice command, and the method comprises subtracting the sample of the acoustic environment from the acoustic input.

Sarin-Hwang-Gruber teaches this claim. Sarin-Hwang teaches a contextual cue that includes the sample of the mobile device’s acoustic environment and subtracting the sample from the acoustic input. Ex-1002 ¶¶225-232. Sarin determines whether audio input is a sufficient match to a command depending on “environmental factors,” such as the background noise. Ex-1005, 8:21-25. For example, Sarin’s voice recognition engine listens for specific commands by “quickly dismiss[ing] some sounds (such as background noise or conversation) and focus[ing] on a command grammar and vocabulary that is relevant to the current context (e.g., words or phrases that correspond to image capture commands).” Ex-1005, 11:37-41. A POSITA would have understood that an obvious implementation of Sarin’s sampling and dismissal of background noise would have been to subtract a sample of the acoustic environment from the received acoustic input. Ex-1002 ¶226.

Further, a POSITA would have been motivated to store the acoustic input previously received by the mobile device and determined not to include a voice command, as taught by Gruber. Ex-1002 ¶227. Gruber stores acoustic profile data, which can be used as contextual information to “aid in recognition and interpretation” of speech. Ex-1008 ¶250. Specifically, Gruber takes in to “account the acoustic environments in which the speech is entered,” by storing “noise profiles” of the user’s location. *Id.* For example, “noise profiles of a quiet office are different from those of automobiles or public places.” *Id.* The noise profiles are identified and stored as the acoustic environment, which includes sounds but not a voice command. Ex-1002 ¶227.

A POSITA would have been motivated to remove the background noise using Gruber’s stored acoustic environment (noise profiles) to “aid in recognition and interpretation of input” by being used as a contextual cue. Ex-1008 ¶250; Ex-1002 ¶¶228-231. Additionally, Sarin-Hwang’s recognition would have been improved by using Gruber’s saved noise profiles of the user’s location so that when Sarin considers the user’s location and compares audio input with other audio data to determine the image capture-related command, it can estimate noise models and dismiss the background noise commonly associated with the user’s location more efficiently and accurately. Ex-1005, 12:38-41; Ex-1002 ¶¶229-231 (citing Ex-1018 ¶¶42, 57).

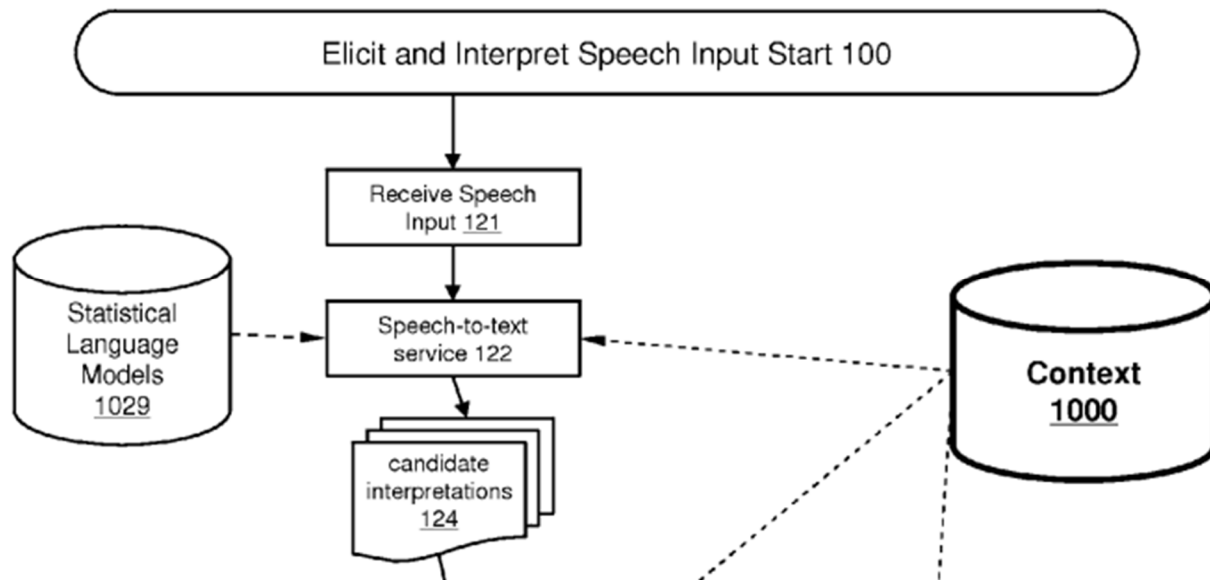
Accordingly, Sarin-Hwang-Gruber stores noise profiles of the user's location(s) which include sound but not a voice command ("contextual cue includes the sample of the acoustic environment obtained by the mobile device by storing acoustic input previously received by the mobile device and determined not to include a voice command") to estimate noise models and dismiss the background noise commonly associated with the user's location ("subtracting the sample of the acoustic environment from the acoustic input"). Ex-1002 ¶232.

e. Claim 10: The method of claim 1, wherein a sequence of performing the plurality of processing stages is determined, at least in part, by the at least one contextual cue.

Sarin-Hwang-Gruber teaches this claim. Ex-1002 ¶¶233-239. As explained in §VIII.A.2.d (Element 1[c]), Sarin-Hwang uses "contextual cues" in processing image capture-related voice commands. Moreover, Sarin-Hwang performs a plurality of processing stages. *See* §VIII.A.2.e (Element 1[d]).

A POSITA would have been motivated to perform Sarin-Hwang's plurality of processing stages in a sequence based on a contextual cue, as taught by Gruber. Ex-1002 ¶235. Gruber's virtual assistant employs a number of processing stages, such as using "statistical language models" ("SLMs") to generate candidate text interpretations of input speech. Ex-1008 ¶317. As shown in excerpted Figure 3 below, contextual information (context 1000) is used by speech-to-text service 122

to select appropriate SLMs 1029 (i.e. processing stages) and generate candidate interpretations 124 of received input speech 121.



Specifically, Gruber’s “Personal Acoustic Context Data 1080” is “used to *select from possible SLMs* 1029 or otherwise tune them to optimize for recognized acoustical contexts.” *Id.* ¶332. Gruber’s personal acoustic context data includes the acoustic environments, which can be used to “aid in recognition and interpretation of input.” *Id.* ¶250. Thus, Gruber uses context data, such as the user’s location, to determine the SLM (that is, to select the appropriate processing stage for a given context) for further processing the voice input. Ex-1002 ¶235.

Gruber’s selection of SLMs are “processing stages” that use contextual cues, consistent with the ’461 Patent because the selection of appropriate SLMs based on contextual information determines the sequence that speech-to-text processing

stages are performed. Ex-1002 ¶236. For example, after interpreting speech input, Gruber teaches to produce candidate interpretations based on one or more SLMs to the speech input. Ex-1008 ¶¶307, 317, Figure 2 (element 190). Further, Gruber teaches that “process steps, method steps, algorithms or the like may be described in a *sequential* order.” *Id.* ¶63. Thus, selecting additional SLMs based on contextual information impacts the sequence of speech recognition processing performed by the virtual assistant. Ex-1002 ¶236; Ex-1008 ¶321 (noting that an SLM can be *added* and tuned based on contextual information).

A POSITA would have been motivated to implement Gruber’s teaching to apply processing sequences based on at least one contextual cue into Sarin-Hwang’s processing stages to improve the speech recognition processing to further personalize results. Ex-1002 ¶¶237-238. For example, Gruber teaches that one benefit of its approach is to provide adaptive processing based on user or “environmental measures” (Ex-1008 ¶¶218, 222, 250), such that the contextual cue will be used to evaluate the processing steps appropriate based on the user’s location or schedule. Ex-1002 ¶¶237-238 (discussing Ex-1005, 11:42-50; Ex-1016 ¶¶3, 22, 43).

Accordingly, Sarin-Hwang-Gruber selects the speech recognition processing sequence (“a sequence of performing the plurality of processing stages”) based on

contextual information (“at least in part, by the at least one contextual cue”). *Id.* ¶239.

- f. Claim 12: The method of claim 11, wherein at least one second processing stage of the plurality of processing stages is performed by the primary processor only after the at least one first processing stage is performed and neither concludes that the acoustic input includes a voice command nor concludes that the acoustic input is the result of spurious acoustic activity.**

Sarin-Hwang-Gruber teaches this claim. Sarin-Hwang-Gruber teaches a second processing stage performed by the primary processor only after the first processing stage and neither concludes that the acoustic input includes a voice command nor concludes that the acoustic input is the result of spurious acoustic activity. Ex-1002 ¶¶240-245. Gruber teaches that when a user request is ambiguous, Gruber’s virtual assistant may undertake a number of actions, “including active elicitation, paraphrasing, suggestions, and the like, and including obtaining context information, so that the correct services ... are called and the intended action taken.” Ex-1008 ¶88. Gruber “may *prompt the user for confirmation and/or request additional context information* from any suitable source before calling a service ... to perform a function.” *Id.*

A POSITA would have been motivated to implement Gruber’s second stage of processing in Sarin-Hwang’s mobile device when the first stage of processing is

not sufficient to specifically identify the voice command to request additional information and confirm that the acoustic input indeed includes a voice command. Ex-1002 ¶¶242-244. A POSITA would have been motivated to modify Sarin-Hwang to include a second processing stage performed by the primary processor, as taught by Gruber, only after the first processing stage is performed in cases where the first stage is not sufficient to definitively identify the command, because doing so would have improved the Sarin-Hwang mobile device's ability to disambiguate voice commands. Ex-1002 ¶244.

Accordingly, Sarin-Hwang-Gruber's second processing stage determines that an audio input includes a voice command ("second processing stage...is performed by the primary processor") after a first processing stage is unable to definitively identify the command ("only after the...first processing stage is performed and neither concludes that the acoustic input includes a voice command nor concludes that the acoustic input is the result of spurious acoustic activity"). *Id.* ¶245.

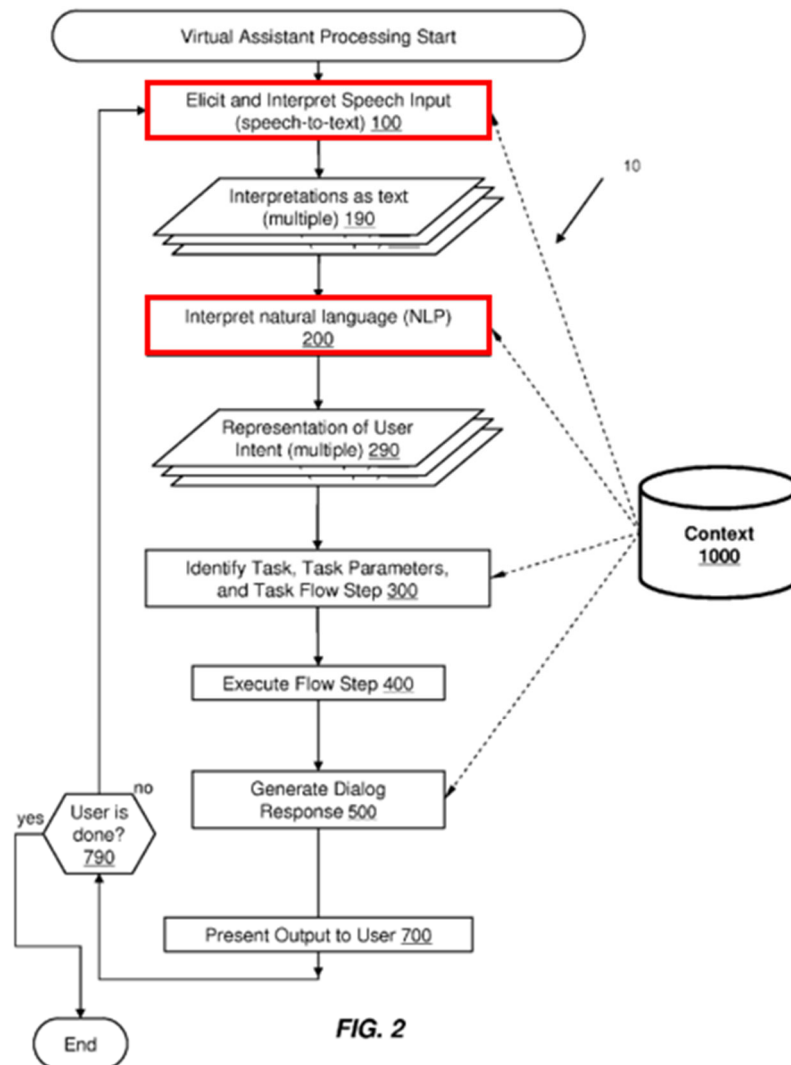
g. Claim 15: The method of claim 11, wherein performing the at least one second processing stage comprises performing at least one speech processing stage including at least one of automatic speech recognition on the acoustic input and at least one natural language processing stage to understand the meaning of the voice command.

Sarin-Hwang-Gruber teaches this claim. Ex-1002 ¶¶246-251. As discussed in §VIII.A.2.e, Sarin-Hwang teaches performing at least one processing stage,

including automatic speech recognition on the acoustic input to detect keywords and image capture-related voice commands.

Further, a POSITA would have been motivated to apply the teachings of Gruber to modify Sarin-Hwang's mobile device to perform at least one second processing stage of automatic speech recognition and natural language processing to understand the meaning of more general voice commands utilizing Sarin-Hwang's primary processor. Ex-1002 ¶248.

As shown in Gruber's Figure 2 below, speech input is received as an auditory signal at step 100 and "*speech-to-text service 122 or processor generates a set of candidate text interpretations 124 of the auditory signal.*" Ex-1008 ¶316. A POSITA would have understood that the speech-to-text processing performed by the virtual assistant on the primary processor entails automatic speech recognition. Ex-1002 ¶¶248-249 (discussing Ex-1012).



At step 200, “[t]he set of candidate interpretations” produced at step 100 is processed by a natural language processor, “which parses the text input and generates a set of possible interpretations of the user’s intent.” Ex-1008 ¶308.

A POSITA would have been motivated to implement Gruber’s automatic speech recognition and natural language processing in Sarin-Hwang’s mobile device because Gruber’s technique enables more reliable interpretation of general voice commands by using speech models that better parse acoustic input and generate a

set of possible interpretations of the user's intent. Ex-1002 ¶250. Moreover, a POSITA would have understood the advantages of implementing Gruber's virtual assistant on Sarin-Hwang's more powerful primary processor because a POSITA would have understood that Sarin-Hwang's low power DSP does not have sufficient processing power to implement Gruber's speech models. *Id.*

Accordingly, Sarin-Hwang-Gruber's second processing stage includes a speech-to-text processing step ("performing at least one speech processing stage including at least one of automatic speech recognition") and a natural language processing step ("and at least one natural language processing stage to understand the meaning of the voice command"). Ex-1002 ¶251.

h. Claim 17: The method of claim 1, comprising performing one or more default actions when a specific task could not be ascertained from the acoustic input.

Sarin-Hwang-Gruber teaches this claim. Sarin-Hwang-Gruber teaches performing a default action when a specific task could not be ascertained from the acoustic input. Ex-1002 ¶¶252-256. Specifically, when a user request is ambiguous or needs further clarification, Gruber teaches to undertake a number of actions, "including active elicitation, paraphrasing, suggestions, and the like, and including obtaining context information, so that the correct services ... are called and the intended action taken." Ex-1008 ¶88. For example, Gruber "may *prompt the user*

for confirmation and/or request additional context information from any suitable source before calling a service ... to perform a function.” *Id.*

The above actions are consistent with “default actions” disclosed in the ’461 Patent that are performed when a detected voice command “cannot be suitably mapped to a specific task or set of tasks.” Ex-1001, 37:27-28. For example, “the default action may include *formulating a question to be rendered to the user to obtain further information* and/or clarification regarding the voice command.” *Id.*, 37:30-33; Ex-1002 ¶254.

A POSITA would have been motivated to implement a default action when a specific task could not be ascertained from the acoustic input, as taught by Gruber, in Sarin-Hwang’s mobile device in order to request additional information in so that the task can be performed. Ex-1002 ¶255.

Accordingly, Sarin-Hwang-Gruber’s mobile device prompts the user to confirm the request and/or requests additional contextual information (“performs one or more default actions”) when a user request is found to be ambiguous, (“when a specific task could not be ascertained from the acoustic input”). Ex-1002 ¶256.

**D. Ground 4: Sarin in Combination with Hwang and Lebeau
Renders Obvious Claim 8**

1. Motivation to Combine Lebeau with Sarin-Hwang

A POSITA would have been motivated to combine the teachings of Sarin, Hwang, and Lebeau (“Sarin-Hwang-Lebeau”) and would have had a reasonable expectation of success in doing so because they are analogous art and teach improvements to the same type of devices. Ex-1002 ¶¶257-261. Each teach an electronic device in a first mode of operation which then switches to a second mode of operation. Ex-1005, 13:53-62 (transitions from a low-power state to an active state); Ex-1006 ¶¶5 (low-power system for receiving audio), ¶31 (based on analysis of the audio feature data, activating various components); Ex-1016, Abstract, ¶¶7-9 (“switch to the second mode of operation”). Each teach using speech processing to determine a voice command. Ex-1005, 6:29-49, 13:15-20; Ex-1006 ¶¶52; Ex-1016 ¶¶7-9 (“monitors ambient sounds for voice input that indicates a request to perform an operation”).

Lebeau offers additional benefits to Sarin-Hwang’s mobile device. Lebeau teaches that the mobile device will detect its condition in order to determine whether to switch the device to monitor for voice input indicating a request. Ex-1016 ¶9. Lebeau’s device attempts to infer whether, based on the current context, monitoring for voice input would provide at least a threshold level of convenience for the user

and, if so, monitor for voice input. Ex-1016 ¶23. Lebeau explains that a context includes whether the mobile device's battery is "running low on stored power (e.g., below 25% charge remaining in the battery)" (Ex-1016 ¶24) or the mobile device is powered from a power cord. Ex-1016 ¶30. Lebeau's device determines whether the "potential convenience to the user of monitoring for voice input is outweighed by the inconvenience to the mobile computing device...of potentially draining the battery in a relatively short period of time (short when compared to standby operation)." *Id.* ¶31.

A POSITA would have recognized that considering the mobile device's battery when processing a voice input is desirable in a mobile device that runs on a battery, as maintaining the primary processor in a deactivated state until the processor is needed would tend to conserve power usage and extend the battery's life, which is a goal of Sarin-Hwang. Ex-1005, 11:32-34; Ex-1002 ¶¶259-260.

A POSITA would have had a reasonable expectation of success in combining Lebeau with Sarin-Hwang. Ex-1002 ¶261. The electronic components of Lebeau and Sarin-Hwang's devices are similar and compatible. A POSITA would have understood that Sarin-Hwang's and Lebeau's electronic components are similar and are interchangeable, therefore, disclose interrelated teachings based on routine technologies that would have been amenable to various well-understood combinations to provide quick and efficient speech recognition. *Id.*

2. Dependent Claim 8

- a. Claim 8: The method of claim 1, wherein the at least one contextual cue includes the power source of the mobile device, and the method comprises influencing evaluation of at least one of the plurality of processing stages based on at least one of determining whether the mobile device is being powered using a battery or wall power, and a state of the battery if the mobile device is being powered using the battery.**

Sarin-Hwang-Lebeau teaches Claim 8. Sarin-Hwang-Lebeau teaches that the mobile device's power source is a contextual cue that influences the processing stage's evaluation based on determining that the device is being powered using the battery. Ex-1002 ¶¶262-266.

A POSITA would have been motivated for Sarin-Hwang's contextual cue (*see* §VIII.A.2.d) that influences evaluation of at least one processing stage to be based on whether the mobile device is being powered using a battery or wall power, and a state of the battery if the mobile device is being powered using the battery, in view of Lebeau. Ex-1002 ¶¶263-264. Lebeau considers “a battery for the mobile computing device running low on stored power (e.g., below 25% charge remaining in the battery)” when determining whether “to switch the mobile computing device from a current mode of operation to a second mode of operation” based on a voice input indicating a request. Ex-1016, Abstract, ¶24. Lebeau also teaches that the mobile device can be powered from a power cord plugged into an outlet. Ex-1016

¶30. Therefore, Lebeau considers the mobile device's power source and allows the battery's status to influence the evaluation performed in the processing stage by not performing voice recognition if the battery is low but performing voice recognition when the device is powered using wall power. Ex-1002 ¶264.

A POSITA would have been motivated to consider the mobile device's battery status, as taught by Lebeau, as one of Sarin-Hwang's contextual cues used when processing the acoustic input because doing so would allow the device to determine whether the "potential convenience to the user of monitoring for voice input is outweighed by the inconvenience of the mobile computing device ... potentially draining the battery in a relatively short period of time." Ex-1016 ¶31; Ex-1002 ¶265.

Accordingly, Sarin-Hwang-Lebeau considers whether the device is being powered by a battery or wall power ("contextual cue includes the power source of the mobile device") and not performing voice recognition if the device is being powered by the battery and it is running low, but performing voice recognition when the device is powered using wall power ("influencing evaluation ... based on at least one of determining whether the mobile device is being powered using a battery or wall power, and a state of the battery if the mobile device is being powered using the battery"). Ex-1002 ¶266.

IX. THE DISCRETIONARY FACTORS FAVOR INSTITUTING TRIAL

A. 35 U.S.C. § 314(a)

Patent Owner (“PO”) may ask the Board to exercise its discretion to deny institution despite the strong invalidity showing on the merits. The Board should decline to do so because the weight of the factors articulated in *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 11 at 6 (PTAB Mar. 20, 2020) (precedential) favors institution.

As an initial matter, there is no litigation currently pending against Co-petitioner Google. Accordingly, none of the *Fintiv* factors that depend on actions or activities of a petitioner in a co-pending litigation apply in any respect to Google. When each of these factors is weighed against Co-petitioner Google, they strongly weigh against denial. Accordingly, the Board should decline to exercise its discretion under 35 U.S.C. §314(a).

1. Stay

This factor is neutral. Petitioners have not requested a stay but intend to do so. The PTAB has explained that it will not speculate on how any such motion would be resolved, before one is filed. *Google LLC v. Parus Holdings, Inc.*, IPR202000847, Paper 9 at 12 (PTAB Oct. 21, 2020). Additionally, the PTAB has explained that this factor is neutral where no such stay motion has yet been filed. *See, e.g., Hulu LLC*

v. SITO Mobile R&D IP, LLC, IPR2021-00298, Paper 11 at 10-11 (PTAB May 19, 2021) (concluding that without evidence of a requested stay or consideration by the district court, this factor does not significantly impact the Board’s exercising discretion to deny institution of the IPR).

2. Trial Date

The district court in the parallel litigation against Co-petitioner Samsung held a scheduling conference on March 25, 2024, and set jury selection for August 4, 2025, approximately nine months away. Ex-1022. Any comparison of a projected FWD date against the scheduled trial date is speculative. *See Dish Network v. Broadband iTV*, IPR2020-01280, Paper 17 at 16 (PTAB Feb. 4, 2021) (“We cannot ignore the fact that the currently scheduled trial date is more than nine months away and much can change during this time”).

Setting that aside, as this Petition is filed on October 16, 2024, a FWD would be expected by April 2026. Although this is approximately eight months after the currently scheduled (speculative) trial date, this factor is not determinative or considered in isolation. *Facebook, Inc. v. USC IP P’Ship*, IPR2021-00034, Paper 13 at 11 (PTAB April 13, 2021) (“[T]his factor is not considered in isolation, but holistically along with other factors”) (citation omitted). Moreover, the Board has instituted IPR and found, on similar facts, that this factor weighs only minimally in favor of denial. *See, e.g., Google LLC, et al. v. Multimodal Media LLC*, IPR2024-

00056, Paper 9 at 8 (PTAB Apr. 12, 2024) (six months between trial and the expected FWD “weighs only marginally in favor” of denial); *NetNut Ltd. v. Bright Data Ltd.*, IPR2021-01492, Paper 12 at 9-16 (PTAB Mar. 21, 2022) (instituting IPR without stipulation and copending trial date six months before FWD); *Facebook*, IPR2021-00034, Paper 13 at 11 (five months between trial and expected FWD “slightly favors denial”); *Coolit Systems, Inc. v. Asetek Danmark A/S*, IPR2021-01195, Paper 10 at 11 (PTAB Dec. 28, 2021) (five months between trial and expected FWD “weighs slightly in favor” of denial); *Equipmentsshare.com Inc. v. Ahern Rentals, Inc.*, IPR2021-00834, Paper 19 at (PTAB Nov. 16, 2021) (seven months between trial and expected FWD “weighs somewhat in favor” of denial).

3. Diligence/Investment

This factor weighs strongly against discretionary denial. No substantive orders have been issued by the court in the underlying litigation, and investment in the parallel litigation against Co-petitioner Samsung will have remained low at the time of institution. *See, e.g., Hulu LLC v. SITO Mobile R&D IP, LLC*, IPR2021-00298, Paper 11 at 12-14 (PTAB May 19, 2021) (holding that this factor supports instituting IPR given the early stage of the district court proceedings, the lack of substantial discovery related to invalidity claims, and the petitioner’s diligent filing of the petition after receiving preliminary infringement contentions). Assuming that a Decision on Institution is issued by April 2025, much work in district court will

remain. *Id.* The opening claim construction brief is currently due January 10, 2025, the *Markman* hearing is scheduled for February 21, 2025, opening expert reports are due March 17, 2025, and dispositive motions are due April 28, 2025. *Id.*

4. Overlap

Factor 4 weighs against discretionary denial. As noted above, the case is in an early stage and there has been very little activity in the case. At present, there is little overlap in issues between the litigation and any IPR resulting from this petition. It would be premature to conclude that the issues raised in this proceeding will overlap with those in the District Court. Instituting this proceeding will allow the Board to address the asserted prior art, narrowing the issues in the co-pending litigations due to the estoppel provisions of 35 U.S.C. §315(e)(2).

Moreover, as noted above, Co-Petitioner Google is not a party to the copending litigation. This IPR is thus the only forum in which Google can currently present its invalidity challenge.

5. Parties

The Board should give little weight to the fact that Co-petitioner Samsung and Patent Owner are also parties to the parallel litigation because Co-petitioner Google and the Samsung defendants are not the same party.

6. Other considerations

Even if the *Fintiv* factors favor discretionary denial (which they do not), the merits are compelling here. The Office did not consider a wealth of prior art during prosecution that discloses and/or suggests the alleged invention. Petitioners therefore respectfully submit that the *Fintiv* factors favor institution and that discretionary denial of this Petition would be neither appropriate nor equitable.

B. 35 U.S.C. § 325(d)

In considering its discretion under 35 U.S.C. §325(d), “the Board uses the following two-part framework: (1) whether the same or substantially the same art previously was presented to the Office or whether the same or substantially the same arguments previously were presented to the Office; and (2) if either condition of first part of the framework is satisfied, whether the petitioner has demonstrated that the Office erred in a manner material to the patentability of challenged claims.” *Advanced Bionics, LLC v. MED-EL Elektromedizinische Geräte GmbH*, IPR2019-01469, Paper 6 at 8 (PTAB Feb. 13, 2020) (precedential).

Here, the first *Advanced Bionics* prong is not met. Four of the references relied upon in this petition (Sarin, Hwang, Garner, and Gruber) were not cited or discussed by the Examiner. While the Lebeau reference was cited substantively in rejections by the Examiner, this is not sufficient to satisfy the first *Advanced Bionics* prong because Lebeau is used in combination with Sarin and Hwang (Ground 4), both not

of record. *See Godbersen-Smith Constr. Co. d/b/a Gomaco Corp. v. Guntert & Zimmerman Const. Div., Inc.*, IPR2021-00050, Paper 24 at 19-21 (PTAB May 7, 2021) (where “neither [the combination references] nor similar prior art was before the examiner ... the prosecution history does not include any overlapping arguments directed to this prior art”); *K/S HIMPP v. Bragi GmbH*, IPR2023-01205, Paper 8 at 20 (PTAB Mar. 19, 2024) (“Although one reference ... was previously presented to the Office, in each asserted ground, Petitioner relies on either Hain or Hain in combination with Sørensen. Because neither of those references was before the office, Patent Owner has not satisfied the first prong of the *Advanced Bionics* test.”).

Moreover, under the second prong, which need not be reached, it was a material error for the Examiner not to have considered the disclosures of Lebeau, as applied here. More particularly, the Office erred in not recognizing that Lebeau teaches to consider the device’s battery power when monitoring for voice input. Ex-1016 ¶¶9, 23-24, 30. During prosecution, the applicant argued that Lebeau uses contextual cues to determine whether or not to enable itself to receive acoustic input so that “Lebeau says nothing about using at least one contextual cue to assist in detecting whether acoustic input which has already been received includes a voice command.” Ex-1004, 1162. But using contextual cues to determine whether the acoustic input includes a voice command is obvious in view of other prior art that the Examiner failed to meaningfully consider. *See* §§VIII.A-C. Because the record

is silent as to the Examiner's understanding of Lebeau in combination with other prior art's teaching of using contextual cues to determine whether the acoustic input includes a voice command, the Examiner must have overlooked it because Lebeau, in combination with Sarin and Hwang teach the claim limitations. *See K/S HIMPP v. Bragi GmbH*, IPR2023-01205, Paper 8 at 20-21 (PTAB Mar. 19, 2024). In view of the foregoing, the Board should not exercise its discretion under §325(d) to deny institution because Sarin, Hwang, Garner, and Gruber were not before the Office and together render all but one of the claims unpatentable.

X. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8

A. Real Parties-in-Interest

Petitioners identify themselves as the real parties-in-interest.⁸

1. Related Matters

To the best of Petitioners' knowledge, the '461 Patent has only been involved in the following district court litigation: *Cerence Operating Co. v. Samsung Electronics Co., Ltd. et al.*, 2:23-cv-00482-JRG-RSP (E.D. Tex.), filed October 13, 2023. Ex-1021.

⁸ Google LLC is a subsidiary of XXVI Holdings Inc., which is a subsidiary of Alphabet Inc. XXVI Holdings Inc. and Alphabet Inc. are not real parties-in-interest to this proceeding.

To the best of Petitioners' knowledge, the '461 Patent has not been challenged in any other *inter partes* review or post-grant review prior to this proceeding.

Petitioners also plan to file IPR2024-01465 directed to U.S. 11,087,750, which is a continuation of the '461 Patent.

B. Lead and Backup Counsel

Lead Counsel	Back-Up Counsel
Benjamin M. Haber (Reg. No. 67,129) bhaber@omm.com O'Melveny & Myers LLP 400 South Hope Street, 19th Floor Los Angeles, CA 90071 Tel: 213-430-6000 Fax: 213-430-6407	Jeff Baxter (Reg. No. 45,560) jbaxter@omm.com O'Melveny & Myers LLP 2801 North Harwood Street, Suite 1700 Dallas, TX 75201 Tel: 972-360-1900 Fax: 972-360-1901 Argirenia Zervos (pro hac vice anticipated) rzervos@omm.com O'Melveny & Myers LLP 1625 Eye Street, NW Washington, D.C. 20006 Tel: 202-383-5300 Fax: 202-383-5414 Kyla Butler (Reg. No.80,082) kbutler@omm.com O'Melveny & Myers LLP 1625 Eye Street, NW Washington, D.C. 20006 Tel: 202-383-5300 Fax: 202-383-5414 Ali R. Sharifahmadian (Reg. No. 48,202)

Lead Counsel	Back-Up Counsel
	ali.sharifahmadian@arnoldporter.com Arnold & Porter Kaye Scholer LLP 601 Massachusetts Ave., NW Washington, DC 20001-3743 Tel: 202-942-5000 Fax: 202-942-5999 Patrick Reidy (Reg. No. 72,148) patrick.reidy@arnoldporter.com Arnold & Porter Kaye Scholer LLP 70 West Madison Street, Suite 4200 Chicago, IL 60602-4231 Tel: 312-583-2300 Fax: 312-583-2360 Albert J. Boardman (Reg. No. 70,601) albert.boardman@arnoldporter.com Arnold & Porter Kaye Scholer LLP 250 West 55th Street New York, NY 10019-9710 Tel: 212-836-8135 Fax: 212-836-8689 Douglas L. Clark (Reg. No. 68,443) douglas.clark@arnoldporter.com Arnold & Porter Kaye Scholer LLP 777 South Figueroa St., 44th Floor Los Angeles, CA 900017-5844 Tel: 213-243-4008 Fax: 213-243-4199

C. Service Information

Please address all correspondence to lead and back-up counsel at the addresses shown above. Petitioners consent to electronic service by e-mail at the following addresses:

bhaber@omm.com
jbaxter@omm.com
rzervos@omm.com
kbutler@omm.com
omm-cerence-samsung@omm.com
ali.sharifahmadian@arnoldporter.com
patrick.reidy@arnoldporter.com
albert.boardman@arnoldporter.com
douglas.clark@arnoldporter.com
xSamsungCerenceAP@arnoldporter.com

D. Power of Attorney

A power of attorney is filed herewith according to 37 C.F.R. §42.10(b).

XI. FEES

Petitioners concurrently electronically submit the required fees for this Petition. The PTO is authorized to charge any fees due during this proceeding to Deposit Account No. LA500639.

Date: October 16, 2024

Respectfully submitted,

/ Benjamin M. Haber /
Benjamin M. Haber (Reg. No. 67,129)
Counsel for Petitioner

CERTIFICATE OF COMPLIANCE

The undersigned hereby certifies that the foregoing Petition for *Inter Partes* Review contains 13,936 words, excluding those portions identified in 37 C.F.R. § 42.24(a), as measured by the word-processing system used to prepare this paper.

/ Benjamin M. Haber /
Benjamin M. Haber (Reg. No. 67,129)
Counsel for Petitioner

CERTIFICATE OF SERVICE

I certify that on October 16, 2024, I caused a true and correct copy of the foregoing Petition for *Inter Partes* Review of U.S. Patent No. 11,393,461 and supporting exhibits to be served via overnight delivery on the Patent Owner at the following correspondence address of record as listed on PAIR:

Brooks Kushman
150 W 2nd St.
Suite 400N
Royal Oak, MI 48067

A courtesy copy was also sent via electronic mail to Patent Owner's litigation counsel listed below:

Andrew D. Weiss
Benjamin T. Wang
Daniel Benjamin Kolko
Qi Tong
Marc A. Fenster
Russ August & Kabat
12424 Wilshire Blvd
Los Angeles, CA 90025
Email: aweiss@raklaw.com
bwang@raklaw.com
dkolko@raklaw.com
ptong@raklaw.com
mafenster@raklaw.com

James Shrin Tsuei
Jonathan D. Link
Paul Anthony Kroeger
Shani M. Tutt
Russ August & Kabat
12424 Wilshire Blvd
Los Angeles, CA 90025
Email: jtsuei@raklaw.com
jlink@raklaw.com
pkroeger@raklaw.com
swilliams@raklaw.com

/ Benjamin M. Haber /
Benjamin M. Haber (Reg. No. 67,129)
Counsel for Petitioner