

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 IPR2024-01465

U.S. Patent No. 11,087,750

**PETITION FOR *INTER PARTES* REVIEW OF
U.S. PATENT NO. 11,087,750**

TABLE OF CONTENTS

PETITIONERS' EXHIBIT LIST	vii
LISTING OF CHALLENGED CLAIMS	ix
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 '750 Patent.....	2
A. Specification	2
B. Prosecution History	5
V. LEVEL OF ORDINARY SKILL IN THE ART	5
VI. CLAIM CONSTRUCTION	6
VII. OVERVIEW OF THE PRIMARY PRIOR ART REFERENCES.....	6
A. Basye (Ex-1005).....	7
B. Shenhav (Ex-1006).....	10
C. Parker (Ex-1007)	12
D. Lebeau (Ex-1014).....	13
E. Conroy (Ex-1017).....	14
VIII. THE BOARD SHOULD NOT EXERCISE ITS DISCRETION UNDER §325(D).....	14
IX. DETAILED EXPLANATION OF GROUNDS.....	17
A. Ground 1: Basye in Combination With Shenhav Renders Obvious Claims 1-8, 11, 13-15, and 19	17
1. Motivation to Combine Basye and Shenhav.....	17

2.	Independent Claim 1	19
a.	1[pre]: A device configured to operate in a lower power mode and a higher power mode, the device comprising:	19
b.	1[a]: at least one processor configured to perform at least one processing stage on an acoustic input received from an environment of the device while the device is operating in the lower power mode, performing the at least one processing stage comprising:.....	21
	(i) 1[a(i)]: determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input, wherein the determining comprises:	25
	(ii) 1[a(ii)]: in response to determining that the acoustic input includes the specific word or phrase, requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command;	30
	(iii) 1[a(iii)]: in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode; and.....	35
	(iv) 1[a(iv)]: responding to the at least one voice command, in either the lower power mode or the higher power mode, depending on the selecting.	37
3.	Dependent Claims 2-8 and 11-13	38
a.	Claim 2: The device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second	

	processor configured to perform at least one second processing stage.	38
b.	Claim 3: The device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.	39
c.	Claim 4: The device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.	42
d.	Claim 5: The device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.	42
e.	Claim 6: The device of claim 4, wherein the first processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.	43
f.	Claim 7: The device of claim 4, wherein the second processor is configured to perform the ASR on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.	46
g.	Claim 8: The device of claim 7, wherein the second processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether	

	the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.	47
h.	Claim 11: The device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.	49
i.	Claim 13: The device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.	50
4.	Independent Claim 14	52
a.	14[pre]: At least one non-transitory computer readable medium storing instructions that, when executed by at least one processor, cause the at least one processor to perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, the method comprising:	52
b.	14[a] and 14[a(i)]	54
c.	14[a(ii)]: determining, while in the low power mode, whether the recognized at least one word or phrase matches a trigger word or phrase;	54
d.	14[a(iii)]	55
e.	14[a(iv)]	55
f.	14[a(v)]	56
5.	Dependent Claim 15	56
a.	Claim 15	56

6.	Independent Claim 19	57
a.	19[pre]-19[a(iv)]	57
b.	19[a(v)]	57
B.	Ground 2: Basye in Combination With Shenhav and Parker Renders Obvious Claims 9 and 16	58
1.	Motivation to Combine Basye, Shenhav, and Parker	58
2.	Dependent Claims 9 and 16	60
a.	Claim 9: ...wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed.	60
	Claim 16: ...wherein the method comprises providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.	60
C.	Ground 3: Basye in Combination With Shenhav and LeBeau Renders Obvious Claims 10 and 18	62
1.	Motivation to Combine Basye, Shenhav, and Lebeau.....	62
2.	Dependent Claims 10 and 18	64
a.	Claims 10/18: The [device of claim 3]/[at least one non-transitory computer readable medium of claim 14], wherein [performing the at least one first processing stage comprises evaluating]/[the method further comprises evaluating, while in the low power mode,] at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.	64

D.	Ground 4: Basye in Combination With Shenhav and Conroy Renders Obvious Claims 12 and 17	67
1.	Motivation to Combine Basye, Shenhav, and Conroy	67
2.	Dependent Claims 12 and 17	68
a.	Claim 12: ...wherein at least one task specified by the voice command is performed with the device selecting to remain in the lower power mode.	68
	Claim 17: ...wherein responding to the at least one voice command comprises initiating performance of at least one task specified by the voice command with the device remaining in the low power mode.	68
X.	THE BOARD SHOULD NOT USE ITS DISCRETION TO DENY INSTITUTION UNDER <i>FINTIV</i>	70
A.	Stay	71
B.	Trial Date.....	72
C.	Diligence/Investment	73
D.	Overlap	74
E.	Parties	74
F.	Other considerations.....	74
XI.	MANDATORY NOTICES UNDER 37 C.F.R. § 42.8.....	75
A.	Real Parties-in-Interest	75
1.	Related Matters	75
B.	Lead and Backup Counsel.....	75
A.	Service Information	77
B.	Power of Attorney	78
XII.	FEES	78

PETITIONERS' EXHIBIT LIST

Exhibit No.	DESCRIPTION
1001	U.S. Patent 11,087,750 (“’750 Patent”)
1002	Declaration of Christopher M. Schmandt (“Schmandt”)
1003	Curriculum Vitae of Christopher M. Schmandt
1004	File History of U.S. Patent No. 11,087,750
1005	U.S. Patent Application Publication No. 2014/0163978 (“Basye”)
1006	U.S. Patent Application Publication No. 2014/0006825 (“Shenhav”)
1007	U.S. Patent Application Publication No. 2015/0106089 (“Parker”)
1008	U.S. Patent Application Publication No. 2012/0016678 (“Gruber”)
1009	U.S. Patent No. 9,275,637 (“Salvador”)
1010	U.S. Patent Application Publication No. 2013/0110521 (“Hwang”)
1011	U.S. Patent No. 9,634,855 (“Poltorak”)
1012	U.S. Patent No. 7,664,642 (“Espy-Wilson”)
1013	U.S. Patent No. 10,573,304 (“Gemmeke”)
1014	U.S. Patent Application Publication No. 2012/0034904 (“LeBeau”)
1015	U.S. Patent No. 9,998,977 (“Joshi”)
1016	U.S. Patent No. 6,594,630 (“Zlokarnik”)
1017	U.S. Patent Application Publication No. 2011/0314185 (“Conroy”)
1018	U.S. Patent Application Publication No. 2012/0296655 (“Kristjansson”)
1019	U.S. Patent No. 9,031,847 (“Sarin”)
1020	Claim Mapping Table

Exhibit No.	DESCRIPTION
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	U.S. Patent No. 8,694,316 ("Pickering")
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. 2003/0179888A1 ("Burnett")
1027	U.S. Patent No. 9,020,459 ("Waters")
1028	U.S. Patent Application Publication No. 2020/0279576A1 ("Binder")
1029	U.S. Patent Application Publication No. 2013/0006638 ("Lindahl")

LISTING OF CHALLENGED CLAIMS

Claim 1	
1[pre]	A device configured to operate in a lower power mode and a higher power mode, the device comprising:
1[a]	at least one processor configured to perform at least one processing stage on an acoustic input received from an environment of the device while the device is operating in the lower power mode, performing the at least one processing stage comprising:
1[a(i)]	determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input, wherein the determining comprises:
1[a(i(A))]	performing automatic speech recognition (ASR), while in the lower power mode, to recognize at least one word or phrase in the acoustic input;
1[a(i(B))]	determining, while in the lower power mode, whether the recognized at least one word or phrase matches the specific word or phrase;
1[a(ii)]	in response to determining that the acoustic input includes the specific word or phrase, requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command;
1[a(iii)]	in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode; and
1[a(iv)]	responding to the at least one voice command, in either the lower power mode or the higher power mode, depending on the selecting.

Claim 2	
2	The device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage.
Claim 3	
3	The device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.
Claim 4	
4	The device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.
Claim 5	
5	The device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.
Claim 6	
6	The device of claim 4, wherein the first processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.

Claim 7	
7	The device of claim 4, wherein the second processor is configured to perform the ASR on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.
Claim 8	
8	8. The device of claim 7, wherein the second processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.
Claim 9	
9	The device of claim 7, wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed.
Claim 10	
10	The device of claim 3, wherein performing the at least one first processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.
Claim 11	
11	The device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the

	device, recent activity on the device, and state of a battery powering the device.
Claim 12	
12	The device of claim 1, wherein at least one task specified by the voice command is performed with the device selecting to remain in the lower power mode.
Claim 13	
13	The device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.
Claim 14	
14[pre]	At least one non-transitory computer readable medium storing instructions that, when executed by at least one processor, cause the at least one processor to perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, the method comprising:
14[a]	performing at least one processing stage on the acoustic input, the at least one processing stage comprising:
14[a(i)]	performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input;
14[a(ii)]	determining, while in the low power mode, whether the recognized at least one word or phrase matches a trigger word or phrase;
14[a(iii)]	in response to determining that the recognized at least one word or phrase matches the trigger word or phrase, requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands;

14[a(iv)]	in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode, and
14[a(v)]	responding to the at least one voice command recognized by the at least one server, in either the lower power mode or the higher power mode, depending on the selecting.
Claim 15	
15	The at least one non-transitory computer readable medium of claim 14, wherein the method comprises performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the at least one first word or phrase matches the trigger word or phrase.
Claim 16	
16	The at least one non-transitory computer readable medium of claim 14, wherein the method comprises providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.
Claim 17	
17	The at least one non-transitory computer readable medium of claim 14, wherein responding to the at least one voice command comprises initiating performance of at least one task specified by the voice command with the device remaining in the low power mode.
Claim 18	
18	The at least one non-transitory computer readable medium of claim 14, wherein the method further comprises evaluating, while in the low power mode, at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

Claim 19	
19[pre]	A device comprising:
19[a]	at least one processor configured to process acoustic input from an environment of the device while the device is operating in a low power mode, the processing comprising:
19[a(i)]	performing automatic speech recognition (ASR), while in the low power mode, to recognize at least one word or phrase in the acoustic input;
19[a(ii)]	determining, while in the low power mode, whether the recognized at least one word or phrase matches a trigger word or phrase;
19[a(iii)]	in response to determining that the recognized at least one first word or phrase matches the trigger word or phrase, requesting that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands;
19[a(iv)]	in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode; and
19[a(v)]	performing at least one task to carry out the at least one voice command in the acoustic input recognized by the at least one server, in either the lower power mode or the higher power mode, depending on the selecting.

I. INTRODUCTION

Google LLC (“Google”), Samsung Electronics Co., Ltd. and Samsung Electronics America, Inc. (“Samsung”) (collectively “Petitioners”) request *inter partes* review of Claims 1-19 of U.S. Patent No. 11,087,750 (“the ’750 Patent”).

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

Petitioners certify that the ’750 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 ’750 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 1-19 (“Challenged Claims”) in view of:¹

Reference	Date(s)	Prior Art At Least Under
US20140163978A1 (“Basye”) (Ex-1005)	12/11/2012 (filed) 6/12/2014 (published)	§102(e)
US20140006825A1 (“Shenhav”) (Ex-1006)	6/30/2012 (filed) 1/2/2014 (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
US20150106089A1 ("Parker") (Ex-1007)	12/30/2010 (filed) 4/16/2015 (published)	§102(e)
US20120034904 ("LeBeau") (Ex-1014)	08/6/2010 (filed) 02/09/2012 (published)	§§102(a), (b), (e)
US20110314185 ("Conroy") (Ex-1017)	08/31/2011 (filed) 12/22/2011 (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, 49-74, 300-301; Ex-1003; Ex-1020.

Ground	Claims	Basis
1	1-8, 11, 13-15, and 19	§103: Basye, Shenhav
2	9, 16	§103: Basye, Shenhav, Parker
3	10, 18	§103: Basye, Shenhav, Lebeau
4	12, 17	§103: Basye, Shenhav, Conroy

IV. OVERVIEW OF THE '750 Patent

A. Specification

The '750 Patent is directed to techniques used by mobile devices to detect voice commands in received audio to minimize the device's power consumption.

Ex-1001, Abstract, 6:25-28. 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.*, 6:29-34.

The patent describes processing received acoustic input in both a low power mode and an active mode. The low power mode is “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.” *Id.*, 5:14-18. By contrast, when in an active mode “the display may be on, the primary processor enabled and/or other components active such that these resources may be more readily utilized without first having to activate them.” *Id.*, 12:30-33.

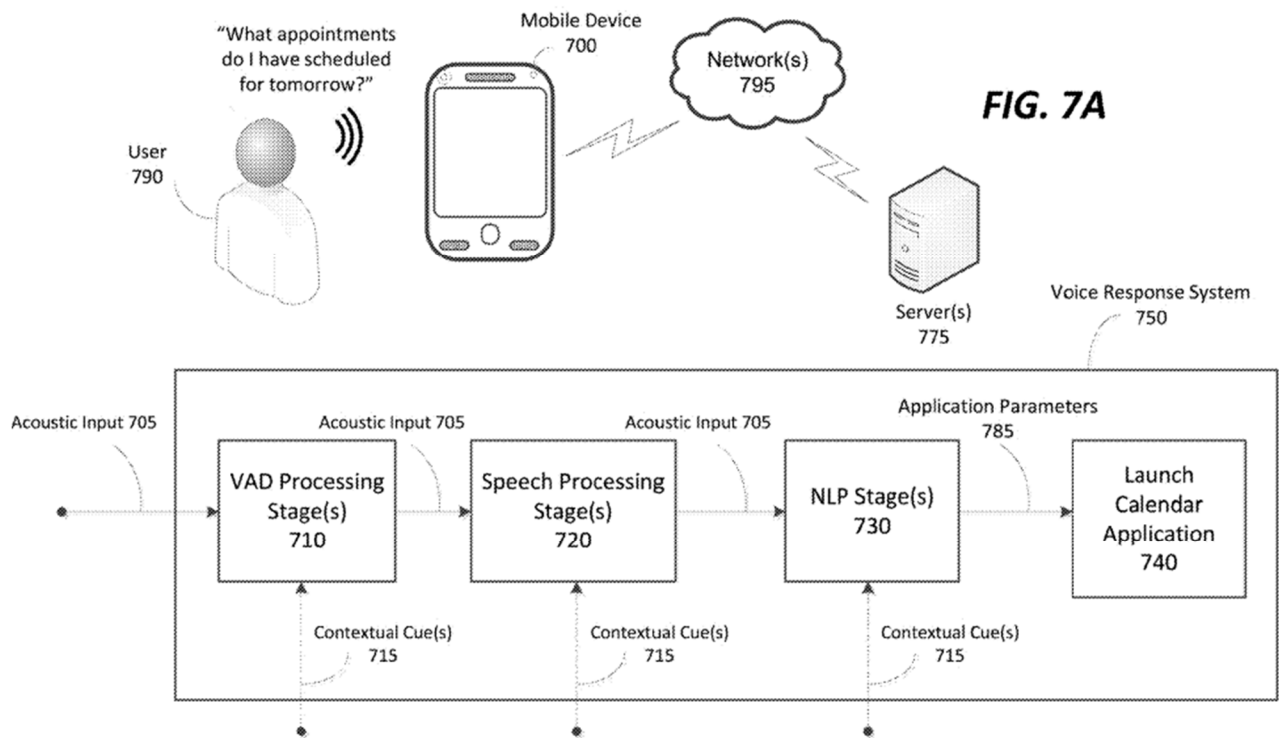


FIG. 7B

As shown in Figure 7B, acoustic input 705 is received by mobile device 700 and proceeds through a number of processing stages. For example, voice activity detection (VAD) “determine[s] whether the acoustic input includes speech content or corresponds to non-speech acoustic activity.” Ex-1001, 30:15-17. Speech processing stages 720 “determine whether the acoustic input includes a voice command.” *Id.*, 30:60-62. After a voice command is detected, natural language processing (“NLP”) stages 730 is used to “understand what the user intended the voice response system to do when speaking the voice command.” *Id.*, 31:42-46. After a voice command is fully recognized, an appropriate application may be

launched. *Id.*, 31:55-61. Moreover, “one or more ASR [automatic speech recognition] stages and/or one or more NLP stages may be performed by one or more servers 775 connected to network(s) 795.” *Id.*, 32:11-13; Ex-1002 ¶¶39-45.

B. Prosecution History

The '750 Patent was filed on August 16, 2016, as a continuation of U.S. Application 13/795,933, which subsequently issued as U.S. Patent No. 11,393,461. Ex-1002 ¶¶46-48.

Near the end of prosecution, on July 13, 2020, the applicant amended the independent claims to add new elements 1[a(iii)], 1[a(iv)], 14[a(iv)], 14[a(v)], 19[a(iv)], and 19[a(v)]. Ex-1004, 932-937. The Office subsequently issued a Notice of Allowance on September 2, 2020. *Id.*, 939-946. This Petition presents arguments not considered during prosecution. *See* Ex-1004.

V. 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 '750 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.

VI. CLAIM CONSTRUCTION

Petitioners are unaware of any “prior claim construction determination” related to the ’750 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.² *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc); Ex-1002 ¶118.

VII. OVERVIEW OF THE PRIMARY PRIOR ART REFERENCES

The prior art references analyzed in the grounds below are analogous to the ’750 Patent because they are within the same field of endeavor as the ’750 Patent and are reasonably pertinent to one or more problems addressed by the ’750 Patent. Ex-1002 ¶¶75-80.

² Petitioners reserve all rights to raise claim construction arguments and other arguments in any parallel or future litigation concerning the ’750 patent. For example, comparing the claims to the accused products in the litigation may raise controversies that require construction of certain claim terms.

A. Basye (Ex-1005)

Basye discloses a mobile device with a “power management subsystem,” which reduces power consumption when performing speech recognition. Ex-1002 ¶¶81-97. Basye recognizes that conventional speech recognition techniques, on mobile computing devices that rely on battery power, consume unacceptable amounts of energy by providing power to the speech recognition components. Ex-1005 ¶9.

As shown in Figure 3 below, in response to receiving audio input, Basye’s power management subsystem selectively activates one or more modules as the audio input progresses. Ex-1005 ¶10. For example, “audio detection module 106 may process audio input” and compares the audio input’s energy level to a threshold. *Id.* ¶19. If the threshold is met, speech detection module 108 is activated and applies “voice activity detection (VAD) techniques” “to determine whether the audio input includes speech.” *Id.* ¶¶20, 21. Once speech is detected in the input audio, speech processing module 110 is activated, which processes the audio input to determine whether the speech includes a keyword “that prompts the activation of one or more components,” such as a “wakeword.” *Id.* ¶¶10, 23, 24.

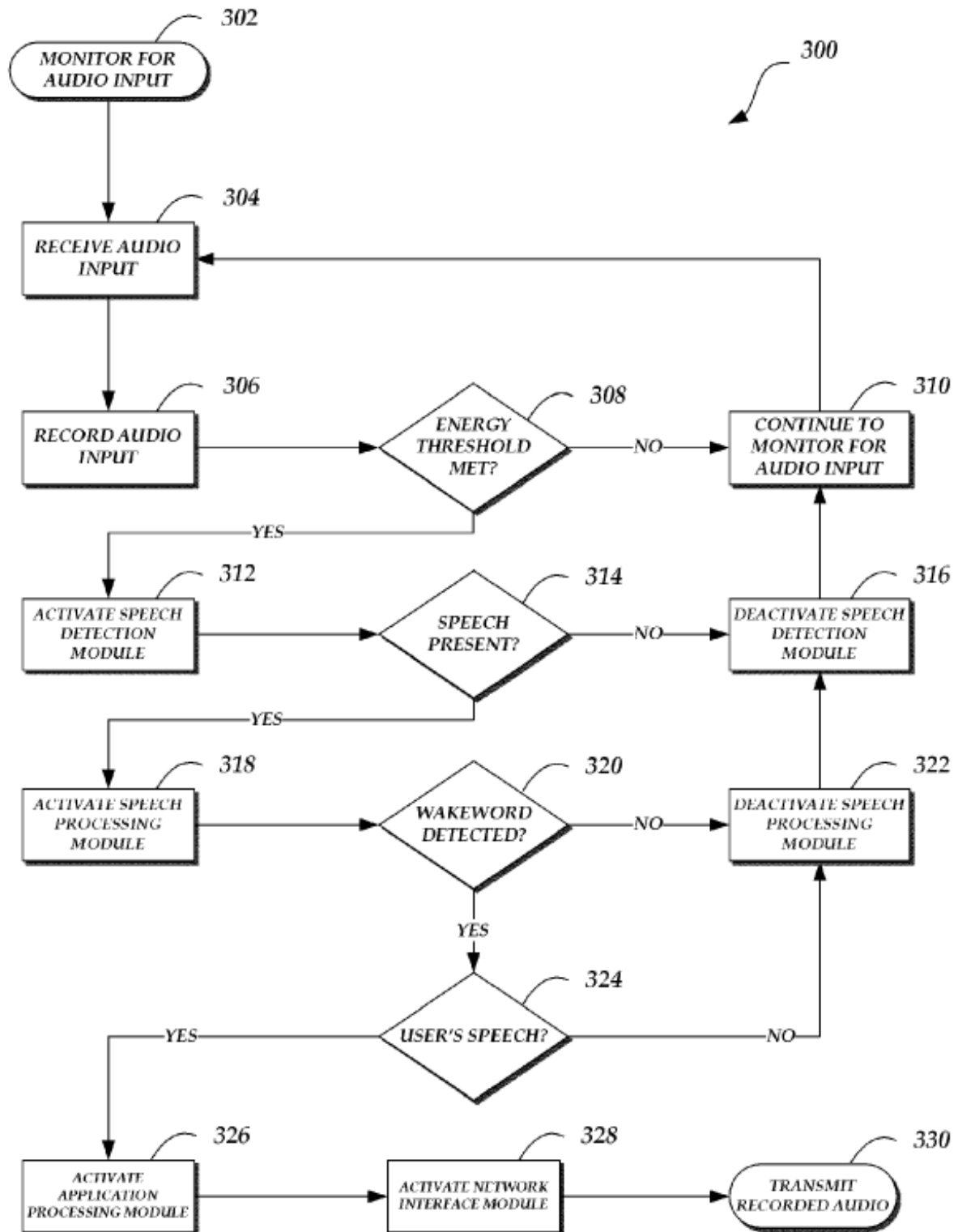


Fig.3.

If speech processing module 110 detects a keyword or wakeword, application processing module 112 and network interface module 206 may be activated. Ex-1005 ¶28. Application processing module 112 may implement “a speech recognition application,” such as “an intelligent agent frontend application.” *Id.* ¶29. Further, audio input (such as a query) can be transmitted to a remote speech recognition server, whereupon recognition results (such as a query response) are returned to application processing module 112. *Id.* ¶31. Basye’s computing device also can perform on-device speech recognition as opposed to relying on a server. Ex-1005 ¶79.

Basye, also incorporates U.S. Patent Application Publication No. 2012/0016678 (“Gruber,” Ex-1008) by reference in its entirety. Ex-1005 ¶13. Gruber discloses an “intelligent automated assistant system [that] engages with the user in an integrated, conversational manner using natural language dialog, and invokes external services when appropriate to obtain information or perform various actions.” Gruber, Abstract. Basye refers to Gruber’s system as an “intelligent agent frontend,” which may be implemented on Basye’s application processing module 112 for performing on-device speech recognition. Ex-1005 ¶¶13, 29; Ex-1002 ¶¶92-97.

B. Shenhav (Ex-1006)

Shenhav is generally directed to “waking up devices from power conservation states.” Ex-1006 ¶1; Ex-1002 ¶¶98-107. Shenhav’s mobile device detects and performs initial processing on a sound signal while operating in a standby, or low power, state, to determine if the sound signal indicates a wake-up phrase. Ex-1006 ¶14. A wake-up phrase may be a short phrase, such as “wake up,” “awake,” or “phone,” that brings the mobile device into an “active mode” from a “stand by mode” (*id.* ¶45), where the mobile device’s main processor(s) are in a “partially [powered] or low power state.” *Id.* ¶33.

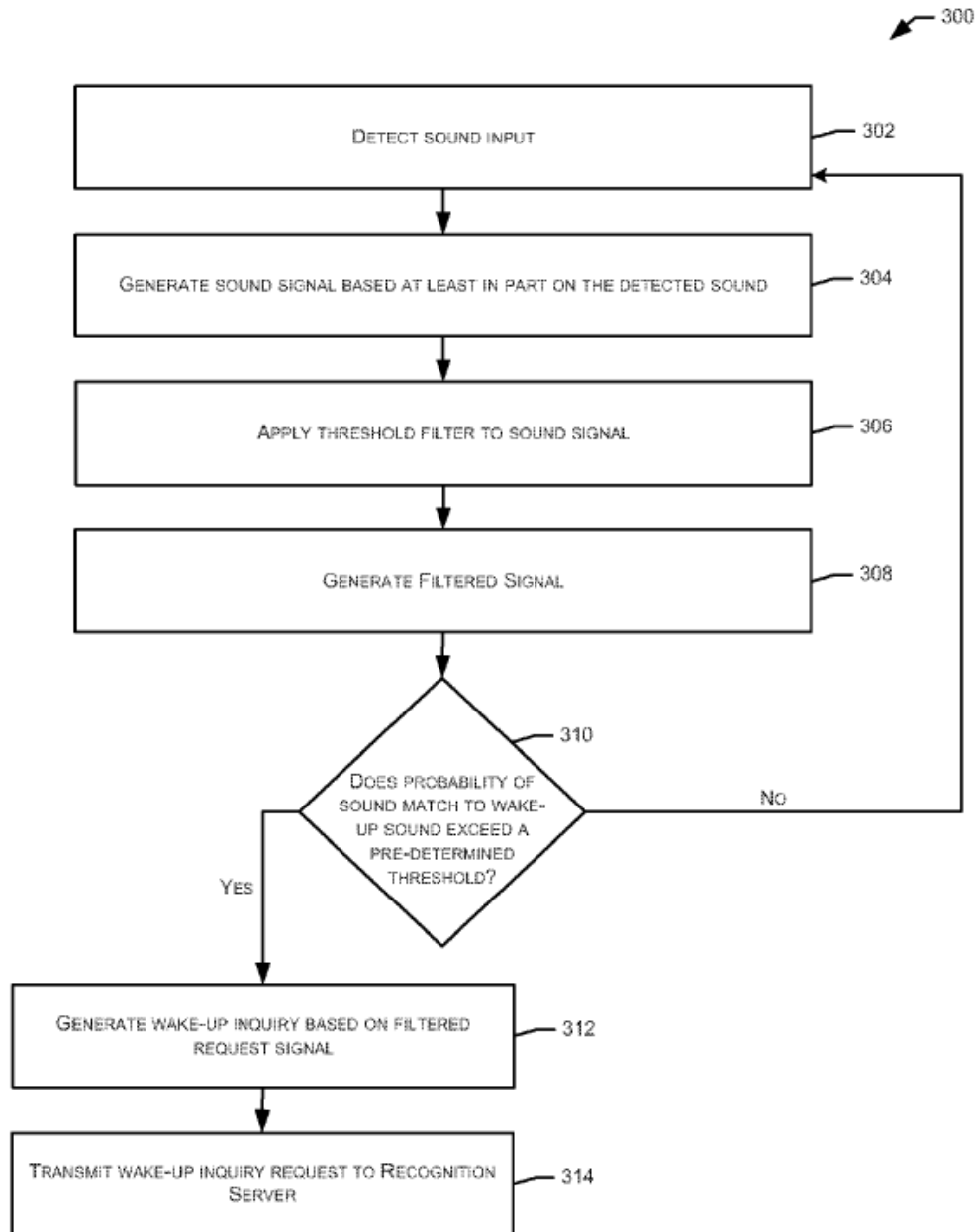


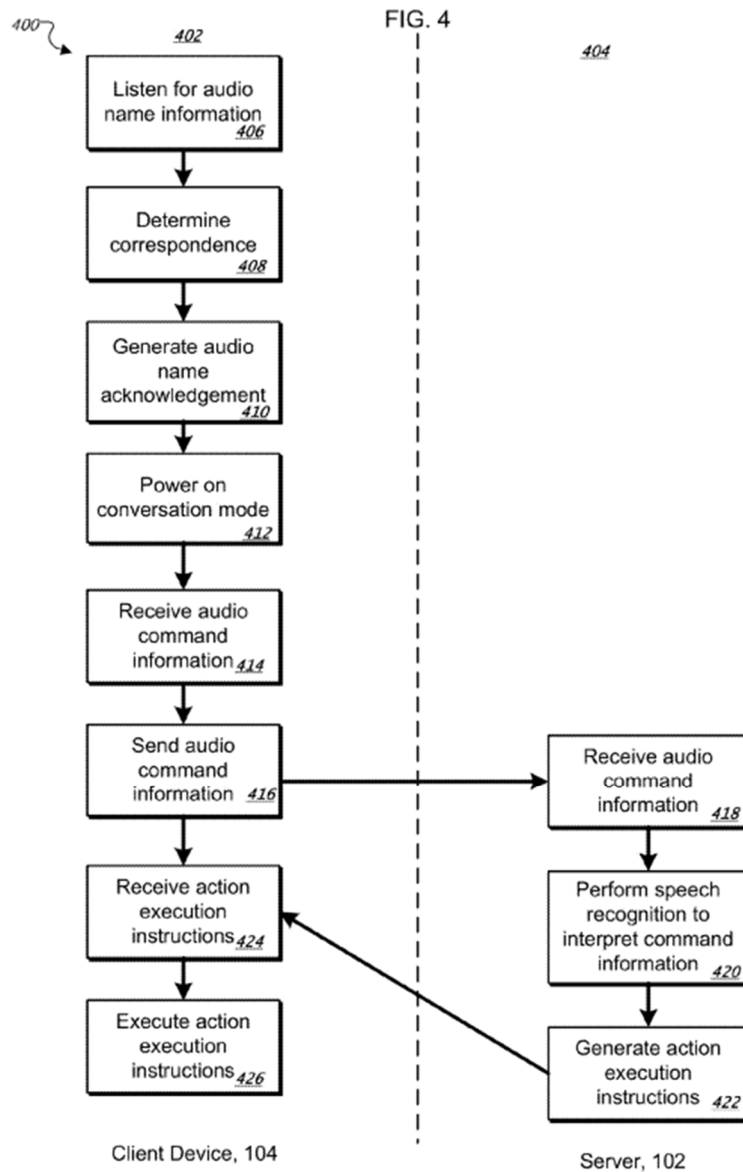
FIG. 3

As shown in Figure 3, if the mobile device determines that it is highly likely that the detected sound signal represents a wake-up phrase, then the device transmits the sound signal as part of a request to a remote recognition server for further analysis.

Id. The recognition server, upon receiving the request, extracts the sound signal and analyzes it “using speech and/or voice recognition methods to determine if the sound signal is indicative of one or more wake-up phrases.” Ex-1006 ¶15. If the server determines that the sound signal indicates a wake-up phrase, then the server generates and transmits a wake-up signal back to the mobile device which prompts the device “to wake up from a sleep or stand by state to a powered state.” *Id.*

C. Parker (Ex-1007)

Parker discloses a client/server system that emulates a “conversation” with an end user. Ex-1007 ¶22; Ex-1002 ¶¶108-113. As shown in Figure 4 below, the client device receives audio commands and sends the commands to a server, which performs speech recognition on the audio commands. Ex-1007 ¶¶36, 38. Based on its speech recognition, the server sends instructions to be performed back to the client device. *Id.* In view of the “time required for server 102 to process” audio commands, Parker generates “an audio acknowledgement to notify” the user that an audio command was received to provide the user with “an assurance that the audio command ... is being processed.” *Id.* ¶¶69-70.



D. Lebeau (Ex-1014)

Lebeau teaches a mobile device that operates in two modes of operation. Ex-1014, Abstract. Lebeau determines to “switch” the device to allow for voice input that indicates a request to perform an operation. Ex-1014 ¶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-1014 ¶23; Ex-1002 ¶¶114-115.

E. Conroy (Ex-1017)

Conroy teaches an electronic device, such as a media player or cell phone (Ex-1017 ¶23), with an “audio subsystem” that processes audio signals with little or no involvement of the processor so that the processor is free to “enter a low power state which extends the useful battery life of the electronic device.” *Id.* ¶¶31-34. Conroy’s audio subsystem handles all of the processing of audio data, so that the processor enters a low power state while the audio data is processed, by for example, a digital signal processor. *Id.* ¶14; Ex-1002 ¶¶116-117.

VIII. THE BOARD SHOULD NOT EXERCISE ITS DISCRETION UNDER §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 (Basye, Shenhav, Parker, and Conroy) were not cited or discussed by the Examiner. While the Basye reference (Ex-1005) was listed on an IDS, this would not be sufficient to satisfy the first prong of *Advanced Bionics* because the record is silent as to the Examiner's consideration or understanding of Basye's teachings. Similarly, while the Lebeau reference (Ex-1014) 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 Basye and Shenhav (Ground 3), a combination not of record. *See Godbersen-Smith Constr. Co. v. Guntert & Zimmerman Const. Div., Inc.*, IPR2021-00050, Paper 24 at 19-21 (PTAB May 7, 2021) ("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 March 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").

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 Basye and Lebeau, as applied here. That is because Basye clearly teaches the features of the

majority of the claims. *See* §IX.A. More particularly, the Office erred in not recognizing that Basye discloses a device that operates in a lower power mode and higher power mode, and which recognizes wakewords in audio input in the lower power mode in the same manner as described by the '750 Patent. The Office further erred in not recognizing that Basye discloses that its device transmits audio input to a server to determine if the audio input includes a voice command to determine whether to transition from the lower power mode to the higher power mode.

The Office also erred in not recognizing that Lebeau teaches to evaluate contextual cues, such as the device's battery power when monitoring for voice input. Ex-1014 ¶¶9, 23-24, 30. Patent Owner ("PO") did not specifically address Lebeau's teachings to overcome the rejections. But using contextual cues to process an acoustic input is obvious in view of other prior art that the Examiner failed to meaningfully consider. *See* §§IX.A-C.

Because the record is silent as to the Examiner's understanding of Basye and Lebeau, the Examiner must have overlooked it because Basye, in combination with Shenhav, Parker, Lebeau, and Conroy teach the claim limitations. *See K/S HIMPP v. Bragi GmbH*, IPR2023-01205, Paper 8 at 20-21 (PTAB March 19, 2024). In view of the foregoing, the Board should not exercise its discretion under §325(d) to deny institution because Shenhav, Parker, and Conroy were not before the Office and render a majority of the claims unpatentable.

IX. DETAILED EXPLANATION OF GROUNDS

A. Ground 1: Basye in Combination With Shenhav Renders Obvious Claims 1-8, 11, 13-15, and 19

1. Motivation to Combine Basye and Shenhav

A POSITA would have been motivated to combine the teachings of Basye and Shenhav (“Basye-Shenhav”) and would have had a reasonable expectation of success in doing so because each relates to the same well-known technologies. Ex-1002 ¶¶119-128.

A POSITA also would have been motivated to combine Shenhav’s teachings with Basye because they are analogous art and teach improvements to the same type of devices. Ex-1002 ¶¶120-121. Each teaches an electronic device in a low power state that is only activated upon a voice command. Ex-1005 ¶¶45, 74, Fig. 3; Ex-1006 ¶¶14 (standby mode), 41-45 (transition into “active mode”). Each teaches using speech processing to determine a voice command. Ex-1005 ¶¶45, 74, Fig. 3; Ex-1006 ¶¶36-37.

Basye discloses a remote speech recognition server that confirms a wakeword was spoken by an end user. Ex-1005 ¶65. Shenhav teaches additional benefits applicable to Basye’s mobile device. Ex-1002 ¶¶122-126. For example, Shenhav recognizes that while the main processor of a mobile device “may provide a relatively high level of processing bandwidth,” it can also “consume a relatively high

level of power and/or energy during operation.” Ex-1006 ¶32. Indeed, the main processor “may consume the most power of all of the elements of the mobile device.” *Id.* Accordingly, Shenhav maintains the main processor in a “stand by” or low power state when processing bandwidth is not required by the mobile device, and to “wake up” the processor when processing bandwidth is needed. *Id.* ¶¶33-34.

Based on Shenhav’s teachings, a POSITA would have been motivated to achieve the power-saving benefit of delaying activation of Basye’s application processing module until after the server confirms the wakeword, processes the query, and provides a result for display. Ex-1002 ¶127; Ex-1005 ¶79 (speech recognition capabilities by “application processing module 112 may require a relatively large power draw”). As such, a POSITA would have understood that applying Shenhav’s teachings (specifically, keeping the main processor in a low-power state until woken up by a command communicated from the remote server) to Basye would further Basye’s goal of reducing power consumption during speech recognition operations. Ex-1002 ¶127.

A POSITA would therefore have been motivated to modify Basye in view of Shenhav’s disclosure such that the audio input that is first analyzed by Basye’s device is transmitted to the speech recognition server to confirm the audio input includes the wakeword without activating Basye’s main processor (application

processing module 112), and would have had a reasonable expectation of success in doing so. Ex-1002 ¶128.

2. Independent Claim 1

a. 1[pre]: A device configured to operate in a lower power mode and a higher power mode, the device comprising:

To the extent the preamble is limiting, Basye teaches³ this element. Ex-1002 ¶¶129-133.

Basye teaches a device configured to operate in lower power and higher power modes. Specifically, Basye’s “computing device” includes a “power management subsystem” (Ex-1005 ¶15) that “selectively activates or deactivates one or more modules of the computing device” (*id.* ¶10). Basye’s modules include “an audio detection module 106; a speech detection module 108; a speech processing module 110; [and] an application processing module 112.” *Id.* ¶15, Fig. 1.

Basye selectively activates its modules based on “an *escalation of processing and/or power consumption*, as modules that are activated later ... may have relatively greater processing requirements and/or power consumption.”⁴ *Id.* ¶44. For

³ 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)).

⁴ All emphases/annotations/color are added unless noted otherwise.

example, while “the audio detection module may persistently monitor for audio input, the remaining components of the power management subsystem may remain in a low-power, inactive state.” *Id.* ¶13. Audio detection module 106 and speech detection module 108 include a “low-power digital signal processor,” while speech processing module 110 includes “a microprocessor configured to detect a keyword in the speech, such as a wakeword or sleepword.” Ex-1005 ¶¶19, 21, 24. Furthermore, application processing module 112 may include “a dedicated microprocessor for implementing [a] speech recognition application.” *Id.* ¶30. Basye observes that speech recognition implemented in application processing module 112 “require[s] a relatively large power draw.” *Id.* ¶79; *see also id.* ¶54 (“on-device speech recognition may have especially high power consumption”).

Thus, a POSITA would have understood that Basye’s device is in a “lower power mode” when the audio detection module 106, speech detection module 108, and/or speech processing module 110 are activated, and application processing module 112 is *not* activated. Ex-1002 ¶132. When the application processing module 112 is also activated, however, Basye’s device is in a “higher power mode” due to the application processing module’s relatively large power requirements. *Id.*

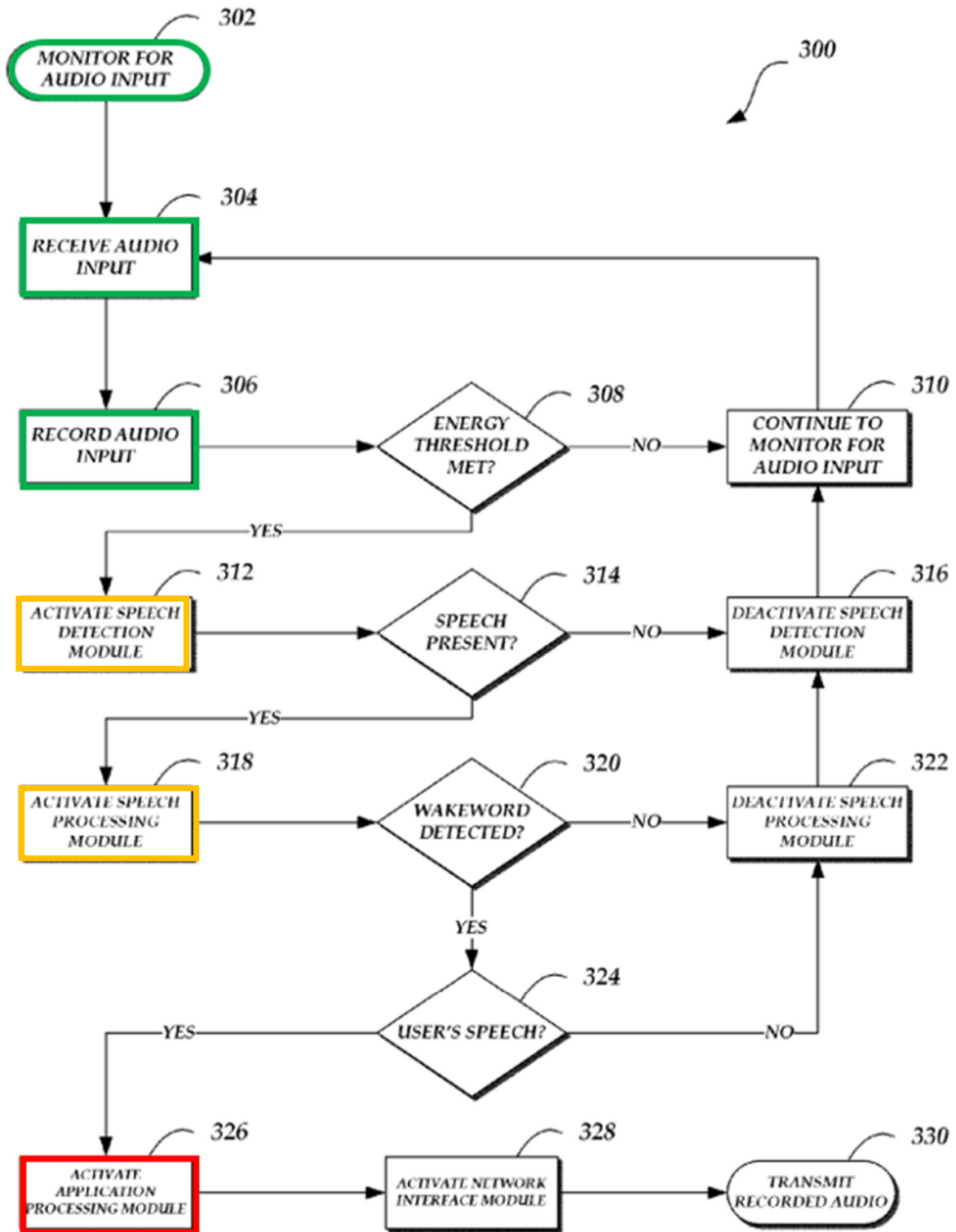
Accordingly, Basye’s computing device progressively activates an audio detection module, a speech detection module, and a speech processing module (“a device configured to operate in a lower power mode”), and which also activates an

application processing module having a large power draw (“and configured to operate in a higher power mode”). *Id.* ¶133.

- b. 1[a]: at least one processor configured to perform at least one processing stage on an acoustic input received from an environment of the device while the device is operating in the lower power mode, performing the at least one processing stage comprising:**

Basye teaches this element. Ex-1002 ¶¶134-143.

Basye teaches a processor that performs at least one processing stage on an acoustic input from the device’s environment, while the device is operating in a lower power mode. As shown below in Figure 3, Basye’s computing device audio input module 208 monitors for audio input (using a microphone, Ex-1005 ¶¶33, 40) and records the received audio input to a memory buffer module (blocks 302, 304, and 306, annotated green). Ex-1005 ¶45. Basye uses the audio input module to determine if the device “is in a relatively *noisy environment*,” so that the user may wish to increase the energy level threshold to activate speech detection module 108 (yellow), in order to detect speech in the audio input. Ex-1005 ¶74. Basye’s audio input module 208 thus receives audio (acoustic) input from the environment of Basye’s computing device when detecting that the environment is noisy or quiet. Ex-1005 ¶74; Ex-1002 ¶136.



The acoustic input is then processed by Basye's device. First, "speech detection module 108" includes a *low-power digital signal processor* ("DSP") that uses "voice activity detection (VAD) techniques" to determine whether the audio input includes speech. Ex-1005 ¶21. Basye's "speech processing module 110 includes a *microprocessor* configured to detect a keyword in the speech." *Id.* ¶24.⁵ Each of the speech detection module and the speech processing module, alone or together, thus perform at least one processing stage performed by low-power digital signal processor and the microprocessor, respectively. Ex-1002 ¶137. Indeed, these processing stages are consistent with '750 Patent's disclosure of what constitutes a processing stage. *See* Ex-1001, 10:58-62 ("employ one or more *VAD processing stages*, a *speech processing stage* ... may be performed on acoustic input"); Ex-1002 ¶138. Therefore Basye's computing device teaches a processor that performs

⁵ Basye teaches a flexible architecture, whereby "[t]he modules of the power management subsystem 100," including speech detection module 108 and speech processing module 110, "may be combined or rearranged." Ex-1005 ¶32. In particular, "[t]he functionality of any module ... may be allocated among multiple modules, or combined with a different module." *Id.* Further, "any or all of the modules may be embodied in one or more integrated circuits, one or more general-purpose microprocessors, or in one or more special-purpose digital signal processors." *Id.* Accordingly, a POSITA would have understood, for example, that speech processing module 108 and speech processing module 110 could be implemented in separate processors (as disclosed in the exemplary embodiment), or, alternatively, within the same processor, such as a microprocessor or a special-purpose digital signal processor. Ex-1002 ¶137 n.4.

at least one processing stage on an acoustic input received from an environment of the device, as claimed.

Basye teaches that its processor processes an acoustic input while the device is operating in a lower power mode. As shown in Figure 3, “power management module 120 *may activate the application processing module 112*” at block 326 (red), after audio input module 208, speech detection module 108 and speech processing module 110 perform their processing steps. Ex-1005 ¶49. Specifically, the processing is performed in lower power mode because speech detection module 108’s low power **DSP** uses VAD techniques to determine whether the audio input includes speech. *Id.* ¶21. If speech is present in the audio input, speech processing module 110 (and its microprocessor) is then activated. *See* Ex-1005 ¶23. Therefore, the processing steps (annotated yellow) are executed while application processing module 112 has not yet been activated (326, red). Ex-1005 ¶45. As discussed in §IX.A.1.a, a POSITA would have understood that Basye’s computing device is operating in a “lower power mode” when any or all of Basye’s audio detection module 106, speech detection module 108, and speech processing module 110 are activated, and application processing module 112 is **not** activated. Ex-1002 ¶¶140-142.

Accordingly, Basye’s speech detection module that uses a digital signal processor and a speech processing module that uses a microprocessor (“at least one

processor”) detect speech and a wakeword (“configured to perform at least one processing stage”) in audio input received by the audio input module’s microphone(s) (“on an acoustic input received from an environment of the device”), where the audio input is received and processed to detect a wakeword while the application processing module is not activated (“while the device is operating in the lower power mode”). Ex-1002 ¶143.

- (i) **1[a(i)]: determining in the lower power mode whether the acoustic input includes a specific word or phrase that, when spoken, indicates that a voice interface of the device is to be engaged for provision of subsequent input, wherein the determining comprises:**

Basye teaches this element. Ex-1002 ¶¶144-151.

Basye determines whether the acoustic input includes a specific word or phrase while in lower power mode. As discussed in §IX.A.2.a and the previous section, speech processing module 110 processes “the audio input to determine whether a *keyword* is included in the speech” while the application processing module 112 is not activated. Ex-1005 ¶24. An example of a keyword that Basye provides is a “wakeword.” *Id.*; Ex-1002 ¶¶145-146.

Basye teaches that the specific word indicates to activate a voice interface for subsequent voice input. Ex-1002 ¶¶147-150. Specifically, Basye teaches that, “[i]f the speech processing module determines that the speech includes the wakeword, an

application processing module may be activated.” Ex-1005 ¶13. After application processing module 112 is activated in response to speech processing module 110’s detecting a wakeword, “subsequently obtained audio inputs are provided from the audio input module 208 directly to the application processing module 112.” Ex-1005 ¶49. Application processing module 112 implements a “speech recognition application module,” such as Gruber’s intelligent agent. Ex-1005 ¶¶13, 29. For example, Gruber’s intelligent agent “provides a *conversational interface* that the user may find more intuitive and less burdensome than conventional graphical user interfaces.” Ex-1008 ¶11. A POSITA would have understood that Gruber’s “conversational interface” is a voice interface. Ex-1002 ¶¶148-150 (discussing conversational interfaces). Therefore, once activated, a POSITA would have understood that Basye’s application processing module is configured to receive audio input that is subsequent to (i.e., received after) the wakeword. *Id.*

Accordingly, Basye’s speech processing module detects wakewords when the computing device’s application processing module is not activated (“determines in the lower power mode whether the acoustic input includes a specific word or phrase”) where the wakewords are an indication to activate an application processing module running an intelligent agent that provides a conversational user interface (“that, when spoken, indicates a voice interface of the device is to be engaged”),

where, subsequent audio input is provided directly from the audio input module to the application processing module (“for provision of subsequent input”). *Id.* ¶151.

(a) 1[a(i(A))] performing automatic speech recognition (ASR), while in the lower power mode, to recognize at least one word or phrase in the acoustic input;

Basye teaches this element. Ex-1002 ¶¶152-160. The ’750 Patent explains that “automatic speech recognition (ASR) processes such as limited vocabulary ASR” may be used to “determine whether a user spoke an explicit voice trigger.” Ex-1001, 10:66-11:6; *see also id.*, 18:20-24 (“performing ASR using a vocabulary restricted to the words in the explicit voice trigger phrase (which may include as few as a single word.)”).

Basye uses models to determine whether the acoustic input includes a word from a limited set of specific words in the same manner described by the ’750 Patent. As discussed in §IX.A.2.a, Basye’s computing device is operating in the “lower power mode” when any or all of Basye’s audio detection module 106, speech detection module 108, and speech processing module 110 are activated, and application processing module 112 is not activated. Ex-1002 ¶¶153-155. Specifically, speech processing module 110 “us[es] HMM techniques, GMM techniques, or other speech recognition techniques” “to detect a keyword in the speech, such as a wakeword.” Ex-1005 ¶24. A POSITA would have recognized that

HMM and GMM are known automatic speech recognition techniques to recognize at least one word in the audio by detecting speech as compared to noise or silence. Ex-1002 ¶¶156-159 (discussing HMM and GMM).

Accordingly, Basye's speech processing module employs HMM, GMM or other speech recognition techniques ("performing ... (ASR)") while the application processor is not activated ("while in the lower power mode") to determine whether the audio input includes a wakeword ("to recognize at least one word or phrase"). *Id.* ¶160.

(b) 1[a(i(B))]: determining, while in the lower power mode, whether the recognized at least one word or phrase matches the specific word or phrase;

Basye teaches this element. Ex-1002 ¶¶161-168.

Basye determines whether the recognized word matches the specific word or phrase, while in lower power mode. As discussed in §IX.A.2.a, Basye's computing device is operating in the "lower power mode" when any or all of Basye's audio detection module 106, speech detection module 108, and speech processing module 110 are activated, and application processing module 112 is not activated. Ex-1002 ¶162.

Basye teaches that speech processing module 110 processes the audio input "to detect the keyword using HMM techniques, GMM techniques, or other speech

recognition techniques.” Ex-1005 ¶24; *see also id.*, ¶21 (discussing speech detection module applies HMM or GMM techniques). A POSITA would have recognized that Basye’s use of HMM and GMM techniques involves the comparison of audio input to keyword models to determine whether a keyword (a specific word or phrase) was detected. Ex-1002 ¶¶163-164 (discussing Ex-1016).

Moreover, Basye’s computing device locally stores “keyword models (e.g., wakeword models or sleepword models), to which an audio input may be compared by the power management subsystem 100.” Ex-1005 ¶36. Similarly, Basye’s speech recognition server utilizes a data store (*id.* ¶59) that “include[s] one or more models of wakewords” so that “[u]pon receiving an audio input determined ... to include a wakeword, the speech recognition server may compare the audio input to a known model of the wakeword stored in the data store 430.” *Id.* ¶60; Ex-1002 ¶165. Because both speech processing module 110 and speech recognition server 420 determine that a wakeword is present in an audio input, and speech recognition server 420 accesses a data store that stores known wakeword models to detect a wakeword in the audio input, a POSITA would have been motivated to implement speech processing module 110 to perform its wakeword detection in a similar way. Ex-1002 ¶¶166-167.

Accordingly, Basye’s speech recognition module compares, while the computing device’s application processing module is not activated (“determining,

while in the lower power mode”), a detected wakeword in audio input to a stored wakeword model (“whether the recognized at least one word or phrase matches the specific word or phrase”). *Id.* ¶168.

- (ii) **1[a(ii)]: in response to determining that the acoustic input includes the specific word or phrase, requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command;**

Basye-Shenhav teaches this element. Ex-1002 ¶¶169-182.

Basye requests, via a network, that a server performs recognition to determine if the input includes a voice command, in response to determining that the acoustic input includes a specific word/phrase. Specifically, after Basye’s computing device determines that the acoustic input includes a wakeword, Basye’s device transmits the “audio input” that includes the wakeword and “may also include a voice command or query” from the speech processing module to a speech recognition server over “network 410.” Ex-1005 ¶65.

Basye further teaches that speech recognition server 420 is “in communication ... with a data store 430” (Ex-1005 ¶59), which includes “one or more models of wakewords” and “one or more acoustic and/or language models for use in speech recognition,” including “general-purpose models as well as specific models.” *Id.*

¶¶60, 61. Using the models, the speech recognition server performs voice command recognition in response to Basye's computing device detecting a wakeword because the server receives audio input containing the wakeword and a separate voice command, and accesses a data store that includes models to identify a wakeword and the voice command. Ex-1002 ¶¶171-177 (explaining Ex-1005 ¶¶55, 65-66). For example, in Figure 4A, Basye's speech recognition server recognizes a wakeword, and in Figure 4B, Basye's speech recognition server recognizes a voice command/agent query. Ex-1005 ¶55; Ex-1002 ¶¶173-177. While these scenarios are provided as distinct examples, this would not have changed a POSITA's understanding that, when the server is provided an audio input containing both a wakeword and a voice command, the server can, and in fact does, process the audio to recognize both. *Id.* Indeed, Basye suggests this by describing speech recognition server 420 as being capable of processing audio input "in a number of ways, depending on [its] implementation." Ex-1005 ¶57; *see also id.* ("speech recognition server 420 processes the audio input ... to confirm that a wakeword is present."), ¶58 ("speech recognition server 420 may process the audio input ... to determine speech recognition results from the audio input.").

To the extent that PO argues or the Board finds that Basye does not expressly disclose that its speech recognition server recognizes a wakeword and a separate voice command within the same acoustic input, a POSITA would have been

motivated to implement the server to do so with a reasonable expectation of success. Ex-1002 ¶178. In particular, a POSITA would have found it advantageous to implement the server to confirm a wakeword and recognize a voice command in a single audio input transmission because doing so would increase reliability, as a single audio input transmission would be less prone to network instability and latency. *Id.* (noting latency between a local device and a remote server as a factor in deciding where to perform speech recognition). Furthermore, allowing for the transmission of a wakeword and separate voice command in the same audio input would be easier and less time consuming for an end user, who would not need to speak a wakeword, wait for confirmation, and then speak a different voice command. *Id.*

A POSITA would also have found it obvious to have Basye's computing device request the server perform recognition while Basye's computing device is in lower power mode. Ex-1002 ¶179. While Basye teaches that the application processing module can be activated to transmit the audio input, Basye also teaches application processing module 112 is not required to process audio input that will be sent to a speech recognition server because Basye's speech processor module 110 may communicate directly with network interface module 206. Ex-1005, Abstract, ¶28; *see also id.* ¶¶39, 49. A POSITA would have found this implementation detail obvious because Basye teaches methods to reduce power consumption, and keeping

the on-device application processor deactivated when not needed would have furthered this goal. Ex-1002 ¶179.

While Basye teaches that the application processor module can be deactivated when a server performs recognition, Shenhav explicitly teaches requesting that a server perform recognition while the device is in a lower power mode. Ex-1002 ¶180. Shenhav teaches that its low power communications processor (1) determines if a received sound signal “is indicative of a probability greater than a predetermined probability level that the sound matches a wake-up phrase” and (2) if so, transmits the signal associated with the sound to a recognition server to “further analyze and determine with a relatively higher level of probability whether the received sound matches the wake-up phrase.” Ex-1006 ¶¶36-37. Shenhav teaches that its communications module “analyze[s] the signal representing the sound,” the mobile device’s main processor is “in a sleep mode or a stand by mode.” *Id.* ¶37.

In addition to the reasons explained in §IX.A.1, a POSITA would have been motivated to implement Shenhav’s requesting a server to perform recognition while in lower power mode into Basye because Shenhav addresses known problems in speech recognition technology, such as performing speech recognition accurately in noisy environments and the high level of processing bandwidth needed for speech recognition that may not be available on a mobile device in a power conservation state. *Id.* ¶4; *see also id.* ¶32 (explaining that the mobile device’s main processor

“may provide a relatively high level of processing bandwidth” and “consume a relatively high level of power and/or energy during operation.”). Implementing Shenhav’s teachings (transmitting the acoustic input to a server in a low power mode before waking up the main processor in the device) with Basye would have furthered Basye’s objectives of reducing on-device power consumption when possible. Ex-1002 ¶181.

Accordingly, Basye-Shenhav’s computing device, in response to detecting a wakeword in audio input (“in response to determining that the acoustic input includes the specific word or phrase”), transmits the audio input to a speech recognition server via a network while the computing device’s application processing module is not activated (“requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input”), where the audio input includes the wakeword and a separate voice command/query (“to determine whether the acoustic input includes at least one voice command”). Ex-1002 ¶182.

- (iii) **1[a(iii)]: in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command, selecting whether to remain in the lower power mode or to transition to a higher power mode; and**

Basye-Shenhav teaches this element. Ex-1002 ¶¶183-188. Basye teaches selecting to transition to a higher power mode in response to an indication of a voice command. As discussed in the previous section, Basye transmits the acoustic input of the wakeword and the voice command to the speech recognition server while in lower power mode (i.e., the application processing module is not activated). Speech recognition server 420 recognizes the speech and generates speech recognition results. Ex-1005 ¶66, Fig. 4B. The speech recognition results are transmitted to application processing module 112 to be presented “in audible format (e.g., via text-to-speech) or in visual format (e.g., via a display of the user computing device 200).” *Id.* A POSITA would have also understood that, in order for application processing module 112 to “cause presentation of the results ... in audible format ... or in visual format,” the computing device would select to activate application processing module 112, thereby transitioning the computing device into a higher power mode. Ex-1002 ¶¶184-186; Ex-1005 ¶66.

Moreover, this implementation detail would have further been obvious in view of Shenhav. Ex-1002 ¶187. Shenhav first determines “if the received signal

associated with the sound is indicative of a probability greater than a predetermined probability level that the sound matches a wake-up phrase.” Ex-1006 ¶36. If there is some likelihood that the received sound is the wake-up phrase, (but there is not a definitive match), then the sound is transmitted to the recognition server 204 to further analyze and determine with a relatively higher level of probability whether the received sound matches the wake-up phrase. Ex-1006 ¶37. If the server determines that the sound signal indicates a wake-up phrase, then the server generates and transmits a wake-up signal back to the mobile device, which prompts the mobile device “to wake up from a sleep or stand by state to a powered state.” *Id.* ¶15. Therefore, Shenhav teaches a two-stage wake-up word confirmation technique, where the server confirms that there is a voice command to determine whether to remain in idle mode or activate the device. Ex-1002 ¶187.

Accordingly, Basye-Shenhav’s computing device receives speech recognition results from a speech recognition server (“in response to receiving an indication from the at least one server that the acoustic input includes at least one voice command”) and activates an application processing module 112 (“selects whether...to transition to a higher power mode”). Ex-1002 ¶188.

- (iv) **1[a(iv)]: responding to the at least one voice command, in either the lower power mode or the higher power mode, depending on the selecting.**

Basye-Shenhav teaches this element. Ex-1002 ¶¶189-192.

Basye-Shenhav responds to the voice command in either power mode based on the selection. As discussed in §IX.A.2.b(iii), Basye-Shenhav activates application processing module 112 when it receives speech recognition results from the speech recognition server. In doing so, the device selects to make a transition from a lower power mode to a higher power mode to activate application processing module 112. Furthermore, once activated, application processing module 112 presents the results audibly or visually. Ex-1005 ¶66. Displaying speech recognition results on the display of the computing device, for example, is a response to a voice command, such as an intelligent agent query. Ex-1002 ¶¶190-191; Ex-1005, Fig. 4B (voice command is “Tell me about flights to Seattle”).

Accordingly, Basye-Shenhav’s computing device displays speech recognition results on a device display (“responds to the at least one voice command”) after selecting to activate an application processing module (“in ... the higher power mode, depending on the selecting.”). Ex-1002 ¶192.

3. Dependent Claims 2-8 and 11-13

- a. Claim 2: The device of claim 1, wherein the at least one processor comprises a first processor configured to perform at least one first processing stage and a second processor configured to perform at least one second processing stage.**

Basye-Shenhav teaches this claim. Ex-1002 ¶¶193-200. Basye teaches a first processor that performs a first processing stage. As discussed in §IX.A.2.b, Basye’s low-power digital signal processor applies voice activity detection (VAD) techniques, a “processing stage” consistent with the ’750 Patent. Ex-1002 ¶193.

Basye further teaches a second processor configured to perform a second processing stage. Ex-1002 ¶¶194-197. Although, as discussed in §IX.A.2.b(ii), Basye-Shenhav’s computing device may rely on a remote server to perform speech recognition, Basye’s computing device is also configured to perform on-device speech recognition using “application processing module 112,” which “includes a dedicated microprocessor for implementing the speech recognition application.” Ex-1005 ¶30. An exemplary speech recognition application implemented by application processing module 112 is the intelligent agent disclosed by Gruber. *Id.* ¶13. Gruber (which Basye incorporates by reference) teaches that its intelligent agent can perform NLP on voice input. Ex-1008 ¶410.

Thus, Basye’s user computing device is configured to perform at least one second processing stage consistent with the ’750 Patent, which also performs NLP

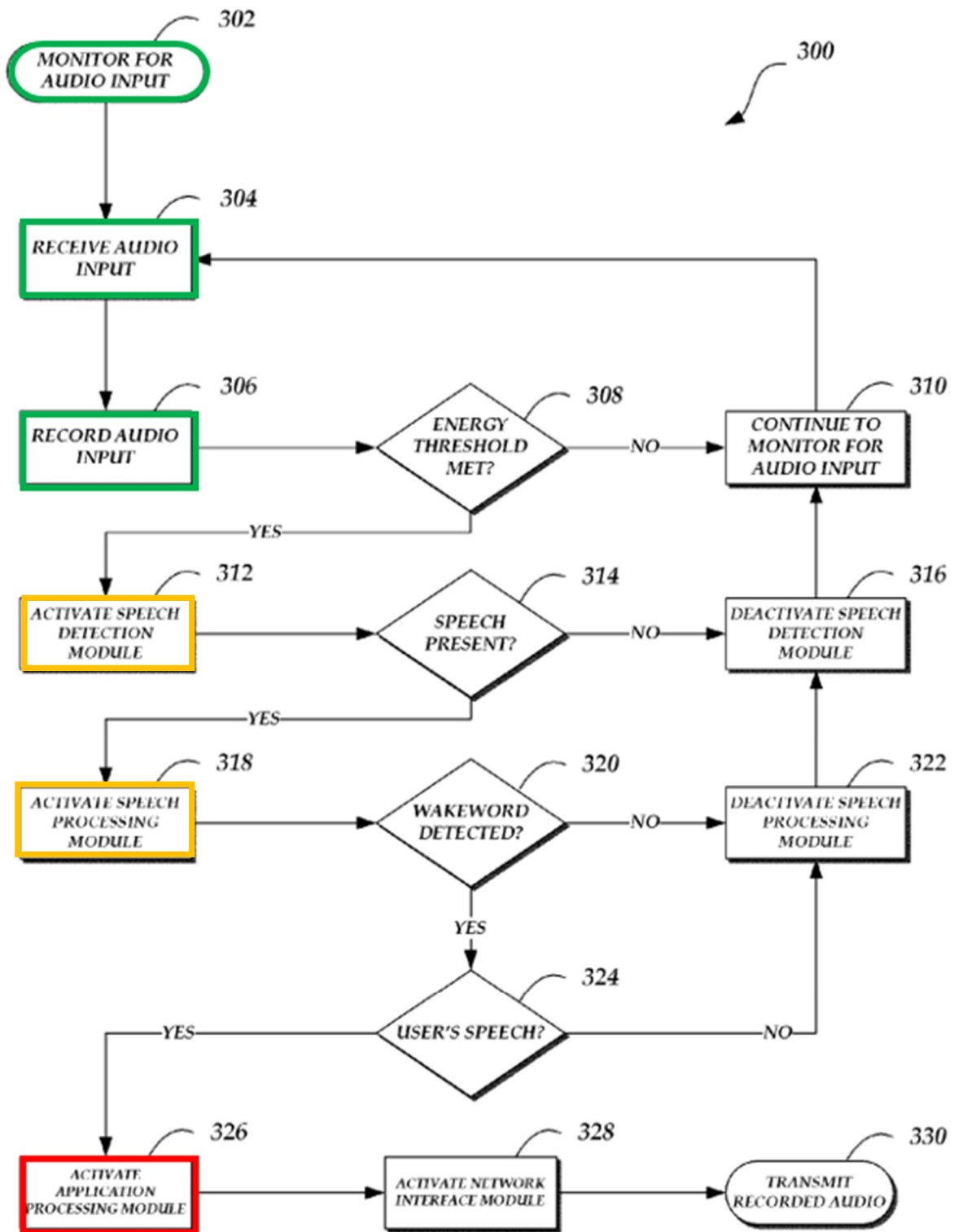
on voice input. *See* Ex-1001, 20:3-5. Moreover, a POSITA would have recognized that, even if the application processing module 112's speech recognition application is bypassed because the mobile device relies on the server to perform speech recognition in some instances, Basye's computing device is nonetheless "configured" to perform speech recognition in other instances. Ex-1002 ¶¶198-199.

Accordingly, Basye-Shenhav's computing device includes a low-power digital signal processor that performs VAD ("a first processor configured to perform at least one first processing stage") and a microprocessor that implements an intelligent agent that is configured to perform NLP on voice input ("a second processor configured to perform at least one second processing stage"). Ex-1002 ¶200.

b. Claim 3: The device of claim 2, wherein the first processor is configured to perform at least one first processing stage prior to the second processor being engaged to perform at least one second processing stage.

Basye-Shenhav teaches Claim 3. Ex-1002 ¶¶201-204. Basye's first processor performs a first processing stage prior to the second processor being engaged for a second processing stage. As discussed in §IX.A.2.b and the previous section, Basye-Shenhav's device includes a low-power digital signal processor that applies VAD techniques and an application processing module that includes a microprocessor configured to perform a second processing stage of NLP.

As shown in annotated Figure 3 below, in response to detecting a wakeword, the computing device, in step 326, may “activate a processing unit that implements any on-device speech recognition capabilities of the computing device.” Ex-1005 ¶13. Thus, when Basye’s computing device is configured for on-device speech recognition, the processing unit that implements on-device speech capabilities (i.e., application processing module 112’s microprocessor) is activated *after* speech detection module 108 determines that input audio includes speech. Ex-1005, Fig. 3 (blocks 312 and 326); Ex-1002 ¶203.



Accordingly, Basye-Shenhav's low-power digital signal processor determines whether audio input includes speech ("first processor is configured to perform at least one first processing stage") before the application processing module's microprocessor is activated to perform NLP ("prior to the second processor being engaged to perform at least one second processing stage."). Ex-1002 ¶204.

c. Claim 4: The device of claim 3, wherein the first processor is a digital signal processor configured to operate at a lower power than the second processor.

Basye-Shenhav teaches this claim. Ex-1002 ¶¶205-206. As discussed in §IX.A.3.a, Basye's first processor is speech detection module 108's *low-power digital signal processor*. Ex-1005 ¶21. Basye's application processing module 112 (the claimed "second processor") "require[s] *a relatively large power draw*" when implementing on-device speech recognition capabilities *Id.* ¶79. Therefore a POSITA would have understood that Basye's first processor operates at a lower power than the second processor. Ex-1002 ¶¶205-206.

d. Claim 5: The device of claim 4, wherein the first processor is configured to perform at least one voice activity detection processing stage including spectral analysis on the acoustic input to evaluate whether the spectrum of the acoustic input is indicative of voice activity, periodicity analysis to evaluate whether the signal periodicity is indicative of voice activity and/or phone loop analysis to evaluate whether the acoustic input includes speech.

Basye-Shenhav teaches Claim 5. Ex-1002 ¶¶207-210.

As discussed in §IX.A.3.b, Basye's speech detection module 108's low-power digital signal processor performs a processing stage in which the module "applies voice activity detection (VAD) techniques." Ex-1005 ¶21. Moreover, Basye teaches using "quantitative aspects of the audio input, such as the spectral slope between one or more frames of the audio input" and the energy level or signal-to-noise ratios of the audio input in one or more spectral bands. *Id.*; Ex-1002 ¶¶208-209.

Accordingly, Basye-Shenhav's speech detection module that includes a low-power digital signal processor applies VAD techniques to audio input ("first processor is configured to perform at least one voice activity detection processing stage"), including analysis of the spectral slope between one or more frames of the audio input ("including spectral analysis on the acoustic input") to determine whether speech is present in the audio input ("to evaluate whether the spectrum of the acoustic input is indicative of voice activity"). Ex-1002 ¶210.

- e. Claim 6: The device of claim 4, wherein the first processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.**

Basye-Shenhav teaches Claim 6. Ex-1002 ¶¶211-222. The '750 Patent describes its implementation of limited vocabulary ASR as performing explicit voice

trigger detection by “using a vocabulary restricted to the words in the explicit voice trigger phrase (which may include as few as a single word).” Ex-1001, 18:20-24.

Basye’s first processor uses models to determine whether the acoustic input includes a word from a limited set of specific words in the same manner as described by the ’750 Patent. As discussed in §IX.A.3.b, Basye’s speech detection module 108’s low-power digital signal processor performs a processing stage in which the module “applies voice activity detection (VAD) techniques.” Ex-1005 ¶21. Basye’s speech processing module 110⁶ also detects keywords using HMM techniques, GMM techniques, or other speech recognition techniques, which a POSITA would have understood are examples of automatic speech recognition techniques. Ex-1005 ¶24. Basye describes keywords as “one or more pre-designated spoken words.” Ex-1005 ¶10. As such, a POSITA would have also understood that Basye’s keyword detection is based upon a limited vocabulary comprising a limited set of words or phrases because they are a “pre-designated” subset of all possible words in the same manner as described by the ’750 Patent. Ex-1002 ¶214.

Moreover, Basye incorporates by reference (Ex-1005 ¶80), U.S. Patent No. 9,275,637 (“Salvador,” Ex-1009), which confirms that Basye’s keyword detection

⁶ Basye teaches that “any or all of the modules may be embodied ... in one or more special-purpose digital signal processors,” so that the functionality of speech processing module 110 could be combined with the functionality of speech detection module 108 into a single module on a digital signal processor. Ex-1005 ¶32.

is based on a smaller, limited vocabulary by contrasting wakeword recognition with more general speech recognition, just like the '750 Patent. *See* Ex-1009, 4:40-44 (“speech recognition module 110 may employ large vocabulary speech recognition techniques and the wake word recognition module may include *speech or phrase recognition particular to the wake word*”); Ex-1001, 18:20-24 (“performing ASR using a vocabulary restricted to the words in the explicit voice trigger phrase (*which may include as few as a single word.*)”); Ex-1002 ¶¶215-216.

Basye’s keyword detection (and Salvador’s wakeword detection) is performed by Basye’s first processor, i.e., speech detection module 108’s *low-power digital signal processor*. Ex-1005 ¶21. A POSITA would have been motivated to implement speech detection module 108 and speech recognition module 110 on a single low power DSP, as suggested by Basye. Ex-1002 ¶¶217-219. A POSITA would have recognized benefits to doing so, such as the fact that implementing Basye’s power management functionality on fewer processors would result in lower overall power consumption on Basye’s computing device. *Id.* ¶220 (discussing Ex-1015). Thus, a POSITA would have recognized that implementing both speech detection and speech processing functionality on a single low-power digital signal processor would further Basye’s goal of conserving device power. *Id.* ¶221.

Accordingly, Basye-Shenhav implements speech detection and speech processing modules on one low-power digital signal processor (“wherein the first

processor”) that performs keyword detection using HMM techniques, GMM techniques (which a POSITA would have known to be a form of ASR in the same manner as described by the ’750 Patent, *see supra* §IX.A.2.b(i)) using a vocabulary that is particular to a wakeword (“performing limited vocabulary automatic speech recognition on the acoustic input”) to determine that the audio input includes a wakeword (“determine whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase”). Ex-1002 ¶222.

f. Claim 7: The device of claim 4, wherein the second processor is configured to perform the ASR on the acoustic input, at least one classification processing stage and/or at least one natural language processing stage to understand the meaning of the voice command.

Basye-Shenhav teaches Claim 7. Ex-1002 ¶¶223-226. Basye’s second processor performs automatic speech recognition that performs a NLP stage to understand the meaning of the voice command. Basye’s application processing module 112’s microprocessor is the claimed second processor. *See* §IX.A.3.c. As discussed in §IX.A.3.a, Gruber’s intelligent agent implemented on Basye’s application processing module 112’s microprocessor (incorporated by reference, Ex-1005 ¶13), performs NLP on the voice input. Ex-1008 ¶410. In one example, Gruber’s NLP ascertains the meaning of the query “who is playing this weekend at the filmore” by using, for example, language pattern recognizers, a vocabulary database, and an active ontology. Ex-1008 ¶412, Fig. 28; Ex-1002 ¶224.

Accordingly, Basye-Shenhav's application processing module 112's microprocessor performs NLP to understand the meaning of voice input commands, such as queries ("the second processor is configured to perform ... at least one natural language processing stage to understand the meaning of the voice command"). Ex-1002 ¶¶225-226. Moreover, as the claim is drafted, only one of ASR, classification processing stage or NLP stage is required to understand the meaning of the voice command as explained by the '750 Patent itself, and Basye-Shenhav teaches to use NLP. *See* Ex-1001, 10:66-11:15; *Cochlear Bone Anchored Sols. AB v. Oticon Med. AB*, 958 F.3d 1348, 1359-60 (Fed. Cir. 2020) ("use of the disjunctive creates three alternative subsets of claim coverage"); *SuperGuide Corp. v. DirecTV Enters., Inc.*, 358 F.3d 870 (Fed. Cir. 2004).

g. Claim 8: The device of claim 7, wherein the second processor is configured to perform the ASR at least in part by performing limited vocabulary automatic speech recognition on the acoustic input to determine whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase.

Basye-Shenhav teaches Claim 8. Ex-1002 ¶¶227-232.

Basye's second processor is configured to perform limited vocabulary automatic speech recognition on the acoustic input to determine whether the acoustic input includes a word or phrase from a limited set that includes a specific word. As discussed above in §IX.A.3.a, Basye's application processing module 112's

microprocessor is the second processor. As recited in Claim 3 of Basye, a “wakeword recognition module is further configured to determine the second score [i.e., a likelihood that audio input comprises a wakeword] using at least one of an *application processing module*, a hidden Markov model, and a Gaussian mixture model.” Ex-1005, Cl. 3.

A POSITA would have understood that Basye’s use of HMM and GMM techniques involves the comparison of audio input to keyword models to determine whether a keyword, such as a wakeword, was detected. Ex-1002 ¶229 (discussing Ex-1016). As previously explained, Basye’s wakeword detection is based upon a limited vocabulary comprising a limited set of words or phrases. Ex-1002 ¶230. Thus, by virtue of what is recited in claim 3, Basye teaches that wakeword recognition can be performed using the application processing module which is implemented on Basye’s second processor.

While Basye explains that application processing module 112’s microprocessor (the claimed “second processor”) “require[s] *a relatively large power draw*” when implementing on-device speech recognition capabilities (Ex-1005 ¶79), Basye also teaches a flexible architecture where “operations or events can be performed” on “multiple processors or processor cores.” *Id.* ¶82.

Accordingly, a POSITA would have found it obvious to perform the limited vocabulary automatic speech recognition (as performed by speech processing

module 110) on certain cores of application processing module 112, such that Basye's mobile device remains in a low power mode, in view of Basye's suggested power gating and because speech recognition is merely software that can run on a processor. Ex-1005 ¶82 ("processor cores"); Ex-1002 ¶231 (discussing Ex-1027).

Further, a POSITA would have been motivated to implement the ASR using Basye's application processor module 112's microprocessor (the claimed second processor) due to the limited processing power of Basye's low-power digital signal processor. Specifically, using specific core(s) in the application processor instead would have allowed for additional processing resources to be applied, such as when Basye implements its on-device speech recognition capabilities. Ex-1005 ¶79; Ex-1002 ¶232. Doing so would have further been advantageous when network connectivity is down, preventing the server from performing the ASR. *Id.*

h. Claim 11: The device of claim 3, wherein performing the at least one second processing stage comprises evaluating at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.

Basye-Shenhav teaches Claim 11. Ex-1002 ¶¶233-241.

Basye's second processing stage comprises evaluating a contextual cue, such as time of day and location of the device. As discussed above in §IX.A.3.b (Claim 3), Basye's application processing module 112's microprocessor is the claimed

second processor that performs the second processing stage. As discussed in §IX.A.3.a, Gruber's intelligent agent implemented on Basye's application processing module 112's microprocessor (incorporated by reference, Ex-1005 ¶13), performs natural language processing on the voice input. Ex-1008 ¶¶410-411, Fig. 28. Gruber's NLP considers "context information" such as the time of day (*id.* ¶412 ("current user location and local time")) and location of the device (*id.* ¶142 (***location information*** "to detect when a user is at home.")); Ex-1002 ¶¶235-240.

Accordingly, Basye's NLP stage ("at least one second processing stage") uses the current user location and local time to interpret voice input ("evaluating at least one contextual cue selected from a group consisting of: time of day" and "location"). Ex-1002 ¶241.

i. Claim 13: The device of claim 1, wherein the device is further configured to perform one or more default actions when a specific task could not be ascertained from the acoustic input.

Basye-Shenhav teaches Claim 13. Ex-1002 ¶¶242-247. Basye teaches that its device performs a default action when a specific task could not be ascertained from the acoustic input. Basye incorporates Gruber (Ex-1008), in its entirety. Ex-1005 ¶13. 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, to obtain the needed information so that the

correct services ... are called and the intended action taken.” Ex-1008 ¶88. In one embodiment, Gruber “prompt[s] 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 ’750 Patent when a command or task cannot be ascertained. *See* Ex-1001, 37:37-40 (“default action may include formulating a question”); Ex-1002 ¶245.

A POSITA would have been motivated to apply Gruber’s teaching of prompting an end user for clarification to address a scenario in which the server, as opposed to the computing device, finds a recognized voice command to be ambiguous because Gruber teaches that its intelligent agent requests clarification, not only when speech recognition is performed locally, but also when a server is relied upon. Ex-1008 ¶¶84, 110-111; Ex-1002 ¶246.

Accordingly, Basye-Shenhav’s computing device that prompts a user for confirmation (“performs one or more default actions”) when it does not understand an ambiguous voice command (“when a specific task could not be ascertained from the acoustic input”). Ex-1002 ¶247.

4. Independent Claim 14

- a. 14[pre]: At least one non-transitory computer readable medium storing instructions that, when executed by at least one processor, cause the at least one processor to perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, the method comprising:**

To the extent the preamble is limiting, Basye teaches this element. Ex-1002 ¶¶248-253. Basye's computing device has "memory 210 to implement various software modules stored therein, such as a user interface module 212, operating system 214, and speech recognition application module 216." Ex-1005 ¶34. Basye's "memory 210 generally includes RAM, *ROM*, and/or *other persistent or non-transitory computer-readable storage media*." *Id.* ¶35. Basye teaches its power management subsystem's modules "are implemented as software that includes computer-executable code carried out by one or more general-purpose processors." *Id.* ¶11. In addition, "[a] software module may be activated (or deactivated) by activating (or deactivating) a general-purpose processor configured to or capable of carrying out the computer-executable code included in the software." *Id.*

Although Basye teaches that its power management subsystem and software modules may be carried out by "general-purpose processors," speech detection module 108's voice activity detection and speech processing module 110's keyword detection can be implemented on a single low power DSP. As such, a POSITA would

have understood that Basye's low power DSP was capable of accessing non-transitory computer readable media as part of executing its software. Ex-1002 ¶¶249-251 (citing Ex-1010). Indeed, a POSITA would have been motivated to configure Basye's DSP to do so because performing the claimed monitoring steps on a low power DSP, rather than application processing module 112's microprocessor, would have furthered Basye's goal of improving the energy efficiency of the computing device. Ex-1005 ¶14.

Basye teaches that the processor performs a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device while the device is operating in a low power mode, for the reasons discussed in §§IX.A.2.a, .b (Elements 1[pre] and 1[a]). A POSITA would also have understood that Basye's device is in a "low power mode," as recited in Claim 14, when the microprocessor of application processing module 112 is not activated, which is consistent with the '750 Patent's disclosure of "[e]ntering a low power mode may include turning off the display, shutting down one or more processors." Ex-1001, 5:14-15; Ex-1002 ¶252.

Accordingly, Basye's non-transitory memory ("at least one non-transitory computer readable medium") stores computer-executable code that can be carried out by one or more processors ("storing instructions that, when executed by at least one processor") to implement the device's audio detection, speech detection, speech

processing, and application processing modules (“cause the at least one processor to perform a method of monitoring an acoustic environment of a device to receive acoustic input from the environment of the device”) on audio input received by an audio input module prior to the application processing module’s activation (“while the device is operating in a low power mode”). Ex-1002 ¶253.

b. 14[a] and 14[a(i)]

Basye-Shenhav teaches Elements 14[a] and 14[a(i)] for the reasons discussed in §§IX.A.2.b, IX.A.2.b(i)(a). Ex-1002 ¶¶254-255.

c. 14[a(ii)]: determining, while in the low power mode, whether the recognized at least one word or phrase matches a trigger word or phrase;

Basye-Shenhav teaches this element for the reasons discussed in §IX.A.2.b(i)(b). Ex-1002 ¶256. Additionally, Basye’s wakeword is an example of a “a trigger word or phrase,” in a similar way as the ’750 Patent, which refers to “explicit voice triggers” as “a specific, designated word or phrase to engage the voice response system.” Ex-1001, 4:46-48. Similarly, Basye describes wakewords as “pre-designated spoken words” that “prompt[] the activation of one or more components.” Ex-1005 ¶10. The ’750 Patent’s “trigger word or phrase” and Basye’s “wakeword” thus are commensurate in scope. Ex-1002 ¶256.

d. 14[a(iii)]

Basye-Shenhav teaches this element for the reasons discussed in §IX.A.2.b(ii). Ex-1002 ¶257. While Elements 14[a(iii)]’s and 1[a(ii)]’s language differ slightly (as shown in the table below), the scope and meaning are not substantively different, as the ’750 Patent uses “explicit trigger word or phrase” and “specific and predetermined word or phrase” interchangeably, and refers to them as an “explicit voice trigger.” Ex-1001, 1:38-44. Accordingly, Element 14[a(iii)] is taught for the same reasons discussed above.

1[a(ii)]	14[a(iii)]
in response to determining that the acoustic input includes the specific word or phrase , requesting, via a network while in the lower power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes at least one voice command ;	in response to determining that the recognized at least one word or phrase matches the trigger word or phrase , requesting, via a network while in the low power mode, that at least one server perform recognition on at least a portion of the acoustic input to determine whether the acoustic input includes one or more voice commands ;

e. 14[a(iv)]

Basye-Shenhav teaches this element for the reasons discussed in §IX.A.2.b(iii). Ex-1002 ¶258.

f. 14[a(v)]

Basye-Shenhav teaches this element for the reasons discussed in §IX.A.2.b(iv). Ex-1002 ¶259. While Elements 14[a(v)] and 1[a(iv)] use slightly different language (shown in the table below), Basye-Shenhav’s voice command is “recognized by the at least one server” when the voice command can be an intelligent agent query, such as “[t]ell me about flights to Seattle.” Basye’s intelligent agent query is processed (i.e., “recognized”) by the speech recognition server, which provides results that Basye’s computing device displays in response. Ex-1005 ¶66. Accordingly, Element 14[a(v)] is taught for the same reasons discussed above, including the server-side recognition discussion.

1[a(iv)]	14[a(v)]
responding to the at least one voice command, in either the lower power mode or the higher power mode, depending on the selecting.	responding to the at least one voice command recognized by the at least one server , in either the lower power mode or the higher power mode, depending on the selecting.

5. Dependent Claim 15

a. Claim 15

Basye-Shenhav teaches Claim 15 for the same reasons as explained in regard to Claim 6. *See* §IX.A.3.e; Ex-1002 ¶260.

6. Independent Claim 19

a. 19[pre]-19[a(iv)]

Basye-Shenhav teaches Elements 19[pre]-19[a(iv)] for the same reasons as described in §§IX.A.2.a-IX.A.2.b(iii), according to the table below. Ex-1002 ¶¶261-265.

Element	Section
19[pre]	IX.A.2.a
19[a]	IX.A.2.b
19[a(i)]	IX.A.2.b(i)(a)
19[a(ii)]	IX.A.2.b(i)(b)
19[a(iii)]	IX.A.2.b(ii)
19[a(iv)]	IX.A.2.b(iii)

b. 19[a(v)]

Basye-Shenhav teaches Element 19[a(v)], for the same reasons as described in Elements 1[a(iv)] and 14[a(v)]. See §§IX.A.2.b(iv), IX.A.4.f. While Elements 19[a(v)] and 14[a(v)] differ (shown in the below table), the '750 Patent itself makes clear that responding to a voice command is “performing at least one task to carry out.” Ex-1001, 1:15-18 (“*respond to voice commands to perform desired tasks.*”);

Ex-1002 ¶266. Accordingly, Element 19[a(v)] is taught for the same reasons discussed above.

14[a(v)]	19[a(v)]
responding to the at least one voice command recognized by the at least one server, in either the lower power mode or the higher power mode, depending on the selecting.	performing at least one task to carry out the at least one voice command in the acoustic input recognized by the at least one server, in either the lower power mode or the higher power mode, depending on the selecting.

B. Ground 2: Basye in Combination With Shenhav and Parker Renders Obvious Claims 9 and 16

1. Motivation to Combine Basye, Shenhav, and Parker

A POSITA would have been motivated to combine the teachings of Basye, Shenhav, and Parker (“Basye-Shenhav-Parker”) 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 ¶¶267-273. Each teaches using speech processing to determine a voice command. Ex-1005 ¶¶21, 24; Ex-1014 ¶¶36-37; Ex-1007 ¶¶36-37. Specifically, Parker receives “audio command information,” and sends the audio command information to a server that “is configured to perform speech recognition on [the] audio command information.” Ex-1007 ¶¶36-38. Based on its speech

recognition, a “speech recognition manager” on the server “generates action execution instructions” specifying one or more actions to be performed by the client device. *Id.* ¶38. The client device then performs the specified actions. *Id.*

Parker teaches additional benefits to Basye-Shenhav’s computing device. Parker recognizes that, when audio command information is processed by a speech recognition server, “a time lag may exist” from the time the user speaks to when the user receives action execution instructions. Ex-1007 ¶70. To address the problem of user uncertainty due to time lags, Parker’s client device acknowledges receipt of the user’s command “immediately after receiving the command and before [Parker’s] client device ... has processed the command either locally or by sending the command over a network to a server for processing.” *Id.* ¶93. Parker notes that the client device is thus “able to process the command, which may take a few seconds, without the user being unsure as to whether [the] client device ... received the command.” *Id.* Parker’s acknowledgment may be either visual or audio notifications. *Id.* ¶¶54, 64; Ex-1002 ¶¶269-271.

A POSITA would have been motivated to implement Parker’s technique of immediately providing an acknowledgement of the received audio input into Basye-Shenhav’s computing device to reduce user uncertainty that their audio input was received and potentially conserve system resources. Ex-1002 ¶272.

A POSITA would have had a reasonable expectation of success in applying Parker's teachings to Basye-Shenhav's computing device because Parker's acknowledgement technique is compatible with the Basye-Shenhav computing device. *Id.* ¶273. For example, Basye-Shenhav's computing device includes speakers (Ex-1005 ¶42), which a POSITA would have understood as enabling the device to provide Parker's audio acknowledgements. Ex-1007 ¶69 (generating "an audio acknowledgement to notify user 110 of receipt of audio command").

2. Dependent Claims 9 and 16

- a. Claim 9: ...wherein the device is further configured to provide, prior to activating the second processor or turning on a display of the device, at least one indication to the user to confirm that the acoustic input was received and is being processed.**

Claim 16: ...wherein the method comprises providing, while in the low power mode, at least one indication to the user to confirm that the acoustic input was received and is being processed.

Basye-Shenhav-Parker teaches Claims 9 and 16. Ex-1002 ¶¶274-278. Basye-Shenhav-Parker teaches providing an indication to the user to confirm that the acoustic input was received and is being processed while in the low power mode. Basye-Shenhav teaches that the speech processing steps are performed prior to activating the application module 112's microprocessor (the claimed "second processor"). *See* §§IX.A.3.b, IX.A.4.a (low power mode).

Parker further notes “a time lag” from when the user speaks the command to when the user receives action execution instructions due to the server’s recognizing the user’s speech, and teaches providing an “audio command acknowledgement provides [the] user ... with an assurance that the audio command information ... is being processed.” Ex-1007 ¶70. Specifically, Parker acknowledges receipt of the user’s command “*immediately after receiving the command and before [Parker’s] client device ... has processed the command either locally or by sending the command over a network to a server for processing.*” *Id.* ¶93. The notification can be either visual or audio. *Id.* ¶¶54, 64; Ex-1002 ¶¶275-276.

In addition to the reasons explained in §IX.B.1, a POSITA would have been motivated to implement Parker’s user indication into Basye-Shenhav’s computing device because a user may choose to repeat the command unnecessarily, perhaps multiple times, further overburdening the device, as well as the network and server. Ex-1002 ¶277.

Accordingly, Basye-Shenhav-Parker’s computing device provides an acknowledgement to the user immediately after receiving audio input (“providing ... at least one indication to the user to confirm that the acoustic input was received and is being processed”) before activating the application processing module’s microprocessor (“prior to activating the second processor”). Ex-1002 ¶278

**C. Ground 3: Basye in Combination With Shenhav and LeBeau
Renders Obvious Claims 10 and 18**

1. Motivation to Combine Basye, Shenhav, and Lebeau

A POSITA would have been motivated to combine the teachings of Basye, Shenhav, and Lebeau (“Basye-Shenhav-Lebeau”) 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 ¶¶279-283. Each teaches using speech processing to determine a voice command. Ex-1005 ¶¶21, 24; Ex-1006 ¶¶36-37; Ex-1014 ¶¶7-9 (“monitors ambient sounds for voice input that indicates a request to perform an operation”).

Lebeau offers additional benefits to Basye-Shenhav’s computing device. Lebeau detects the mobile device’s condition when determining to switch the device to monitor for voice input. Ex-1014 ¶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-1014 ¶23. Lebeau’s “context” includes whether the mobile device’s battery is “running low on stored power (e.g., below 25% charge remaining in the battery)” (Ex-1014 ¶24) or whether the mobile device is receiving power from a power cord. Ex-1014 ¶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 contextual cues, such as 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 Basye-Shenhav. Ex-1005 ¶9; Ex-1002 ¶¶281-282.

A POSITA would have had a reasonable expectation of success in applying Lebeau’s teachings to Basye-Shenhav’s computing device because Lebeau’s detection of contextual cues is compatible with the Basye-Shenhav computing device, which already considers contextual cues when processing the acoustic input. Ex-1005 ¶43 (considering power level to selectively activate or deactivate modules); Ex-1002 ¶¶283. The electronic components of Lebeau with Basye-Shenhav’s devices are similar and interchangeable. 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. Dependent Claims 10 and 18

- a. **Claims 10/18: The [device of claim 3]/[at least one non-transitory computer readable medium of claim 14], wherein [performing the at least one first processing stage comprises evaluating]/[the method further comprises evaluating, while in the low power mode,] at least one contextual cue selected from a group consisting of: time of day, motion of the device, location of the device, an occurrence of an event on the device, recent activity on the device, and state of a battery powering the device.**

Basye-Shenhav-Lebeau teaches Claims 10 and 18. Ex-1002 ¶¶284-288. Basye-Shenhav teaches performing a first processing stage while in the lower power mode. *See* §§IX.A.3.b (Claim 3), IX.A.4.a (Element 14[pre]). As described in §IX.A.3.h (Claim 11), Basye-Shenhav also teaches processing the acoustic input by evaluating a contextual cue, such as time of day and location of the device. Basye also considers whether the mobile device’s power supply “drops below a threshold power supply level (e.g., a battery charge percentage).” Ex-1005 ¶79; *see also id.* ¶43.

A POSITA would have been motivated to evaluate a contextual cue while in the low power mode, in view of Lebeau. As shown in Figure 1B below, at step A, Lebeau detects contextual cues (such as the user’s location, the time of day, calendar, and battery status) and, at step B, uses the contextual cues to determine whether to monitor for voice input autonomously. Ex-1014 ¶42.

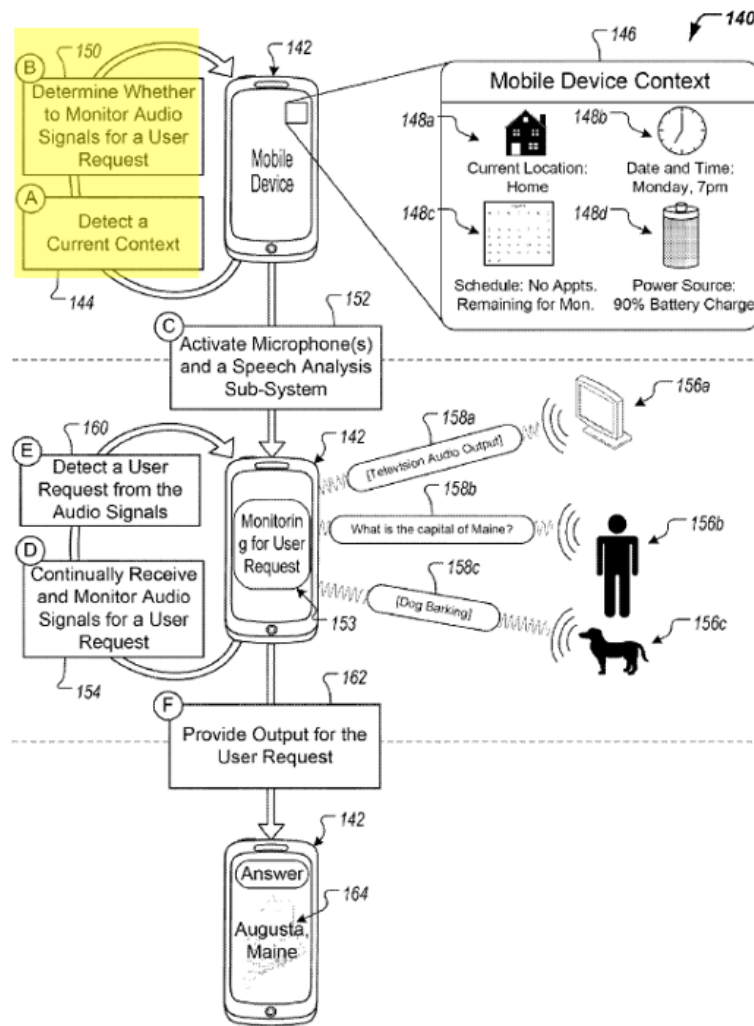


FIG. 1B

For example, Lebeau uses the time of day as a contextual cue when determining “a threshold likelihood of [the user] providing a voice-based request to the mobile computing device.” *Id.* ¶43. Moreover, Lebeau’s “context” includes whether the mobile device’s battery is “running low on stored power (e.g., below 25% charge remaining in the battery)” (Ex-1014 ¶24) or whether the mobile device is receiving power from a power cord. Ex-1014 ¶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).” Ex-1014 ¶31.

In view of Lebeau’s teachings, a POSITA would have understood that Lebeau uses a contextual cue, while in low power mode, because 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. Ex-1002 ¶¶286-287.

Further, although Lebeau employs contextual cues to determine whether to begin monitoring audio (Ex-1014 ¶43), a POSITA would have nonetheless been motivated to evaluate contextual cues, such as battery level as taught in Lebeau, in Basye-Shenhav’s first processing stage (i.e., determining whether already received acoustic input includes a specific word or phrase) because the contextual cue could be used to determine the likelihood that the user provided an acoustic input with a specific word or phrase. Ex-1014 ¶¶23, 43, 87. A POSITA would have been motivated to do so because constantly monitoring for voice input can drain more power than normal standby operation and can go on for an indeterminate amount of time. Ex-1014 ¶31; Ex-1005 ¶79; Ex-1002 ¶288.

**D. Ground 4: Basye in Combination With Shenhav and Conroy
Renders Obvious Claims 12 and 17**

1. Motivation to Combine Basye, Shenhav, and Conroy

A POSITA would have been motivated to combine the teachings of Basye, Shenhav, and Conroy (“Basye-Shenhav-Conroy”) and would have had a reasonable expectation of success in doing so because each relates to the same well-known technologies, such as computing devices that operate in low power state in order to conserve battery. Ex-1005 ¶15; Ex-1006 ¶27; Ex-1017 ¶¶14, 23, 33; Ex-1002 ¶¶289-293.

A POSITA also would have been motivated to combine the teachings of Conroy with Basye-Shenhav because they are analogous art and teach improvements to the same type of devices. Ex-1002 ¶¶290-291. Conroy offers additional benefits to Basye-Shenhav’s computing device. Conroy teaches to process and play audio signals with little or no involvement of the CPU so that the CPU is free to “enter a low power state which extends the useful battery life of the electronic device.” Ex-1017 ¶¶31-34.

A POSITA would have recognized that by performing tasks, such as processing and playing audio, while the device is in a low power mode, would tend to conserve power usage and extend the battery’s life, which is a goal of Basye-Shenhav. Ex-1005 ¶9; Ex-1002 ¶292.

A POSITA would have had a reasonable expectation of success in applying Conroy's teachings to Basye-Shenhav's computing device because the electronic components of Conroy with Basye-Shenhav's devices are similar and interchangeable. Ex-1002 ¶293. 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 perform the requested tasks while saving power. *Id.*

2. Dependent Claims 12 and 17

- a. Claim 12: ...wherein at least one task specified by the voice command is performed with the device selecting to remain in the lower power mode.**

Claim 17: ...wherein responding to the at least one voice command comprises initiating performance of at least one task specified by the voice command with the device remaining in the low power mode.

Basye-Shenhav-Conroy teaches Claims 12 and 17. Ex-1002 ¶¶294-299. As discussed in §§IX.A.2.b(ii)-IX.A.2.b(iii), Basye teaches selecting to transition to a higher power mode in response to receiving an indication of a voice command. Specifically, Basye's speech recognition server 420 recognizes the speech in the audio input, generates and transmits speech recognition results to application processing module 112 to be presented audibly or visually. Ex-1005 ¶66, FIG. 4B. While the presentation of the results would thereby wake up the application

processor, it would have been obvious for the device to respond to the at least one voice command by performing at least one task specified by the voice command with the device remaining in the low power mode, as taught by Conroy. As the '750 Patent acknowledges, it was known in the art that certain voice commands may be implemented with a mobile device remaining in a low power state. The example of such a known application described by the '750 Patent is that “some mobile devices are capable of rendering music while in a low power mode.” Ex-1001, 32:22-24.

This is true, and a POSITA would have known of references, such as Conroy (Ex-1018), which do just this. For example, Conroy teaches an electronic device, such as a media player or cell phone (Ex-1017 ¶23), with an “audio subsystem” that processes the audio signals with little or no involvement of the CPU so that the CPU is free to “enter a low power state which extends the useful battery life of the electronic device.” *Id.* ¶¶31-34. When an audio track is to be played, Conroy’s CPU remains in a low power state and the controller sends a signal to the Audio DSP indicating that data is available in the subsystem to be processed and played. Ex-1017 ¶¶62-64. The audio subsystem will “draw audio data” until all of the audio data stored in host memory has been transferred, which triggers an interrupt for the CPU to power up. *Id.* ¶¶67-68. Thus, in Conroy, the audio subsystem handles all of the processing of audio data, so that the CPU and the rest of the host subsystem enter a

low power state while the audio data is processed, by for example, a digital signal processor. *Id.* ¶¶14, 31.

As taught by Conroy, when an application, such as a music playing application, can be run in a lower power state, it is beneficial to do so to conserve system resources. Ex-1002 ¶¶297-298. This allows a voice command (such as a command to play music) to be performed without requiring extra power or resources of the main application processor, such that the device will select to remain in a low power mode when performing the user's request. Ex-1017 ¶¶14, 31, 33.

Accordingly, a POSITA would have been motivated to use such known low-power applications in response to a voice command to, e.g., play an audio track, since this would conserve resources. Ex-1002 ¶299. Therefore it would have been obvious to implement performing the task specified by the voice command after the device selects to remain in a lower power mode because there exists a number of well-known applications that do just that, and doing so would achieve the predictable result of low-power processing. *Id.*

X. THE BOARD SHOULD NOT USE ITS DISCRETION TO DENY INSTITUTION UNDER *FINTIV*

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).

A. 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).

B. 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 17, 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).

C. 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.*

D. 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.

E. 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.

F. 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.

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

A. Real Parties-in-Interest

Petitioners identify themselves as the real parties-in-interest.⁷

1. Related Matters

To the best of Petitioners' knowledge, the '750 Patent has only been involved in the following district court litigation: *Cerence Operating Co. v. Samsung Elecs. Co., Ltd. et al.*, 2:23-cv-00482-JRG-RSP (E.D. Tex.), filed October 13, 2023. Ex-1021.

To the best of Petitioners' knowledge, the '750 Patent has not been challenged in any other *inter partes* review or post-grant review prior to this proceeding.

Petitioners also filed IPR2024-01464 directed to U.S. Patent No. 11,393,461, which is the parent of the '750 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	Jeff Baxter (Reg. No. 45,560) jbaxter@omm.com O'Melveny & Myers LLP

⁷ 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.

Lead Counsel	Back-Up Counsel
400 South Hope Street, 18th Floor Los Angeles, CA 90071 Tel: 213-430-6000 Fax: 213-430-6407	2501 North Harwood Street, Suite 1700 Dallas, TX 75201-1663 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) 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

Lead Counsel	Back-Up Counsel
	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

A. 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

B. Power of Attorney

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

XII. 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 17, 2024

Respectfully submitted,

/s/ 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 13,992 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.

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

CERTIFICATE OF SERVICE

I certify that on October 17, 2024, I caused a true and correct copy of the foregoing Petition for *Inter Partes* Review of U.S. Patent No. 11,087,750 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

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