

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

**DECLARATION OF CHRISTOPHER M. SCHMANDT IN SUPPORT
OF PETITION FOR INTER PARTES REVIEW
OF U.S. PATENT NO. 11,087,750**

U.S. Patent No. 11,087,750
Declaration of Mr. Christopher M. Schmandt

TABLE OF CONTENTS

I.	INTRODUCTION	9
A.	BACKGROUND AND QUALIFICATIONS	10
B.	INFORMATION CONSIDERED	14
II.	LEGAL STANDARDS	16
A.	Anticipation	16
B.	Obviousness.....	17
C.	Level of Ordinary Skill in the Art	20
D.	Claim Construction	21
III.	SUMMARY OF OPINIONS.....	23
IV.	THE '750 PATENT.....	23
A.	Overview of the '750 Patent.....	23
B.	Prosecution History	27
V.	TECHNOLOGY BACKGROUND.....	27
A.	It was well known to use contextual cues in speech recognition systems for power saving benefits.	27
B.	Noise reduction and/or cancellation techniques were well known in order to better process speech input and speech recognition.	37
C.	Voice activity detection was well known to detect human speech.	43
VI.	BRIEF DESCRIPTION OF THE APPLIED PRIOR ART	44
A.	Basye (Ex-1005).....	46
B.	Shenhav (Ex-1006).....	56
C.	Parker (Ex-1007)	62
D.	Lebeau (Ex-1014).....	66

E.	Conroy (Ex-1017).....	66
VII.	CLAIM CONSTRUCTION	67
VIII.	DETAILED EXPLANATION OF GROUNDS.....	67
A.	Ground 1: Basye in Combination With Shenhav Renders Obvious Claims 1-8, 11, 13-15, and 19	67
1.	A POSITA would have been motivated to combine Basye’s and Shenhav’s teachings and would have had a reasonable expectation of success.....	67
2.	Independent Claim 1	72
a.	1[pre]: A device configured to operate in a lower power mode and a higher power mode, the device comprising:	72
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:.....	74
(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:	79
(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;	90
(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	

Declaration of Mr. Christopher M. Schmandt

the lower power mode or to transition to a higher power mode; and.....98

(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. 100

3. Dependent Claims 2-8 and 11-13 102

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. 102

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. 106

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. ...109

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. 110

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. 112

- 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. 118
- 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. 120
- 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. 123
- 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. 125
- 4. Independent Claim 14 128
 - 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: 128
 - b. 14[a]: performing at least one processing stage on the acoustic input, the at least one processing stage comprising: 132

- c. 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; 132
- d. 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; 132
- e. 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; 133
- f. 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.....134
- g. 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. 134
- 5. Dependent Claim 15 135
 - a. Claim 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.....135
- 6. Independent Claim 19.....135
 - a. Elements 19[pre], 19[a]: A device comprising: at least one processor configured to process acoustic input from an environment of the device while the device is

Declaration of Mr. Christopher M. Schmandt

operating in a low power mode, the processing
comprising: 135

- b. 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; 136
- c. 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; 136
- d. 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; 136
- e. 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.....136
- f. 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. 137

B. Ground 2: Basye in Combination With Shenhav and Parker Renders Obvious Claims 9 and 16 138

- 1. A POSITA would have been motivated to combine Basye's, Shenhav's, and Parker's teachings and would have had a reasonable expectation of success.....138
- 2. Dependent Claims 9 and 16 143
 - a. Claim 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

Declaration of Mr. Christopher M. Schmandt

least one indication to the user to confirm that the acoustic input was received and is being processed.....143

Claim 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. 143

C. Ground 3: Basye in Combination With Shenhav and LeBeau Renders Obvious Claims 10 and 18 145

1. A POSITA would have been motivated to combine Basye’s, Shenhav’s, and Lebeau’s teachings and would have had a reasonable expectation of success.....145

2. Dependent Claims 10 and 18 147

a. Claim 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 147

Claim 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. 147

D. Ground 4: Basye in Combination With Shenhav and Conroy Renders Obvious Claims 12 and 17 151

1. A POSITA would have been motivated to combine Basye’s, Shenhav’s, and Conroy’s teachings and would have had a reasonable expectation of success.....151

2.	Dependent Claims 12 and 17	153
a.	Claim 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.	153
	Claim 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.....	153
IX.	SECONDARY CONSIDERATIONS	157
X.	CONCLUSION.....	157

I. INTRODUCTION

1. I have been retained by the Samsung Electronics Co., Ltd. and Samsung Electronics America, Inc. (“Samsung” or “Petitioners”) as an independent expert consultant in this proceeding before the United States Patent and Trademark Office (“PTO”).

2. I am being compensated at a rate of \$550/hour for my services in this proceeding, which is my regular and customary rate.

3. My compensation is in no way contingent on the nature of my finding, the presentation of my findings in testimony or this declaration, or the outcome of this or any other proceeding. I have no other interest in this proceeding.

4. I have been asked to consider whether certain references disclose and/or suggest the features recited in Claims 1-19 of U.S. Patent No. 11,087,750 (“the ’750 Patent”). (Ex-1001).¹ My opinions are set forth below.

5. I have been asked to provide my independent analysis of the ’750 Patent in light of the prior art publications cited below. I have also been asked to consider the state of the art and prior art available as of March 12, 2013. Based on the prior art discussed in this declaration, it is my opinion that Claims 1-19 of the ’750 Patent are unpatentable for the reasons provided below.

¹ Where appropriate, I refer to exhibits which I understand will be attached to the petition for inter partes review of the ’750 Patent (the “Petition”).

A. BACKGROUND AND QUALIFICATIONS

6. I retired several years ago after a 40-year career at the Massachusetts Institute of Technology (“MIT”); for most of that time I was employed as a Principal Research Scientist at the Media Laboratory. In that role I also served as faculty for the MIT Media Arts and Sciences academic program. I was a founder of the Media Laboratory, a research lab which now spans two buildings.

7. I received my B.S. degree in Electrical Engineering and Computer Science from MIT in 1978, and my M.S. in Visual Studies (Computer Graphics) also from MIT. I was employed at MIT since 1980, initially at the Architecture Machine Group which was an early computer graphics and interactive systems research lab. In 1985, I helped found the Media Laboratory and continued to work there until retirement. I was director of a research group titled “Living Mobile.” My research spanned distributed communication and collaborative systems, with an emphasis on multi-media and user interfaces, with a strong focus on speech-based systems. I have over 70 published conference and journal papers and one book in the field of speech technology and user interaction.

8. For the first fifteen years of my career, my research emphasized speech recognition and speech user interfaces. I built the first conversational computer system utilizing speech recognition and synthesis (“Put That There”) starting in 1980. Put That There parsed spoken input into a semantic structure representing a

command or intent, and could ask questions of the user to evoke responses which would be further parsed and melded into the context of the current command. Some information could be gleaned from non-speech channels, specifically where a user was pointing. I continued to innovate speech user interfaces using recognition, text-to-speech synthesis, and recorded audio in a wide variety of projects. I built one of the first graphical user interfaces for audio editing, employing keyword recognition on voice memos in 1982 (“Intelligent Ear”). I built the first research-grade unified messaging system, which combined text and voice messages into a single inbox, with speech recognition over the phone for remote access, and a graphical user interface for desktop access in 1983 (“Phone Slave”). Again, Phone Slave included a parser which could accept fragmentary input and build up a complete representation of the user’s utterance with additional input. Especially in its follow up multi-user system, Conversational Desktop (1985), a number of contextual cues could also aid the dialog understanding, including date, time, calendar entries, and with whom a user was speaking over a teleconference channel. Along with my students we built the first system for real time spoken driving directions, including speech-accessible maps of Cambridge, Massachusetts in 1987 (“Back Seat Driver”). Back Seat Driver had to translate a route from a map form in which each block was a separate street segment with intersection information contained in nodes, into spoken directions, using words like we use with each other to tell the driver where

to go. We built some of the earliest speech-based personal assistants for managing messages, calendar, contacts, etc. (Conversational Desktop 1985, Chatter 1993, MailCall 1996). We built quite a few systems employing speech recognition in handheld mobile devices (ComMotion 1999, Nomadic Radio 2000, Impromptu 2001, and Symphony 2004, for example). We applied speech recognition to large bodies of everyday conversations captured with a wearable device and utilized as a memory aid (Memory Prosthesis 2004). We used speech recognition on radio newscasts to build a personalized version of audio newscasts (Synthetic News Radio, 1999) and also investigated adding speech recognition to a mouse-based window system a few years earlier.

9. In addition to this work running my academic research lab, I helped found and then design speech user interfaces for InTouch Systems, which built natural language based personal digital assistants accessible over the telephone. In this context I helped solve a number of problems in obtaining adequate speech recognition performance in the face of personal information, such as contact lists, containing words which were difficult for a system to understand. The system also maintained dialog context, such that while reading email, for example, commands related to email were anticipated, such as finding important messages or messages from some particular person. When a user was accessing a calendar, a different

grammar was invoked, emphasizing creation and retrieval of calendar entries by date or time.

10. I was later awarded the prestigious Association for Computing Machinery (ACM) Computer Human Interface (CHI) Academy membership specifically for those years of work pioneering speech user interfaces.

11. In the course of my research, I built a number of speech recognition client/server distributed systems, with the first being in 1985. Much of the initial motivation for a server architecture was that speech recognition required expensive digital signal processing hardware that we could not afford to put on each computer, so a central server with the required hardware was used. Later versions of the speech recognition server architecture allowed certain computers to perform specialized tasks serving a number of client computers providing voice user interfaces, either on screens or over telephone connections. Because of my early work with distributed speech systems, I served for several years in the mid-1990s with a working group on the impact of multimedia systems on the Internet reporting to the Internet Engineering Task Force (IETF) and later the Internet Activities Board (IAB). This work impacted emerging standards such as Session Initiation Protocol (SIP).

12. In my faculty position I taught graduate level courses in speech technology and user interaction design, and directly supervised student research and theses at the Bachelors, Masters, and PhD level. I oversaw the Masters and PhD

thesis programs for the entire Media Arts and Sciences academic program during my more senior years. I also served on the Media Laboratory intellectual property committee for many years.

13. My Curriculum Vitae is attached as Ex-1003.

14. I have been retained by counsel for Petitioner to provide my expert opinion in connection with the above-captioned proceeding as set forth herein.

B. INFORMATION CONSIDERED

15. In forming my opinions, I reviewed the following documents, in addition to others I have cited in my declaration. I also relied on my own knowledge of and experience in the field of speech recognition.

Exhibit No.	DESCRIPTION
1001	U.S. Patent 11,087,750 (“’750 Patent”)
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”)

U.S. Patent No. 11,087,750
Declaration of Mr. Christopher M. Schmandt

Exhibit No.	DESCRIPTION
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
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. 2020/0279576 (“Binder”)
1029	U.S. Patent Application Publication No. 2013/0006638 (“Lindahl”)

II. LEGAL STANDARDS

16. In forming my opinions and considering the subject matter of the '750 Patent and its claims in light of the prior art, I am relying on certain legal principles that counsel in this case explained to me. My understanding of these concepts is summarized below.

17. I understand that earlier publications and patents may act to render a patent unpatentable for one of two reasons: (1) anticipation, and (2) obviousness.

A. Anticipation

18. As explained to me by counsel, the claims of a patent are anticipated by a prior art reference if each and every element of the claim is found either explicitly or inherently in the reference. I understand that inherency requires a showing that the missing descriptive matter in the claim is necessarily present in the allegedly anticipating reference, and that it would have been so recognized by a person of ordinary skill in the art ("POSITA").

19. I understand that when a challenged claim covers several structures, either generically or as alternatives, the claim is deemed anticipated if any of the structures within the scope of the claim is found in the prior art reference.

20. Although anticipation typically involves the analysis of a single prior art reference, I understand that additional references may be used to show that the prior art reference has enabling disclosure (i.e., allows a POSITA to make the

invention without undue experimentation), to explain the meaning of a term used in the prior art reference, and/or to show that a characteristic is inherent in the prior art reference.

B. Obviousness

21. As explained to me by counsel, a claim is invalid as obvious if it would have been obvious to a person of ordinary skill in the art at the time the alleged invention was made. This means that even if all of the elements of the claim cannot be found in a single prior art reference that would anticipate the claim, a person of ordinary skill in the art who was aware of the prior art would have been able to come up with the claimed invention. This may be the case, for example, where the missing element represents only an insubstantial different over the prior art or a reconfiguration of a known system. I understand that in an obviousness determination, the person of ordinary skill in the art is presumed to have knowledge of all material prior art.

22. I understand that an obviousness analysis requires an understanding of the scope and content of the prior art, any differences between the alleged invention and the prior art, and the level of ordinary skill in evaluating the pertinent art.

23. I understand that when a product is available, design incentives and other market forces can prompt variations of it, either in the same field or a different one. If a person of ordinary skill in the art can implement a predictable variation,

obviousness likely bars its patentability. For the same reason, if a technique has been used to improve one device and a person of ordinary skill in the art would recognize that it would improve similar devices in the same way, using the technique would have been obvious.

24. I understand that whether a prior art reference renders a patent claim unpatentable as obvious is determine from the perspective of a person of ordinary skill in the art at the time of the alleged invention. I have been told that there is not requirement that the prior art contain an express suggestion to combine known elements to achieve the claimed invention, but a suggestion to combine known elements to achieve the claimed invention may come from the prior art, as filtered through the knowledge of one skilled in the art. In addition, I have been told that the inferences and creative steps a person of ordinary skill in the art would employ are relevant to the determination of obviousness.

25. I understand that one may consider, e.g., whether (1) the change was merely the predictable result of using prior art elements according to their known functions, or whether it was the result of true inventiveness; (2) there is some teaching or suggestion in the prior art to make the modification or combination of elements claimed in the patent; (3) the claimed innovation applies a known technique that had been used to improve a similar device or method in a similar way; (4) the claimed invention would have been obvious to try, meaning that the claimed

innovation was one of a relatively small number of possible approaches to the problem with a reasonable expectation of success by those skilled in the art; (5) the invention merely substituted one known element for another known element in order to obtain predictable results; (6) the invention merely applies a known technique to a known device, method, or product to yield predictable results; or (7) known work in one field of endeavor may have prompted variations of it for use in either the same field or a different one based on design incentives or other market forces that would have been predictable to a person of ordinary skill in the art.

26. I further understand that certain factors may support or rebut the obviousness of a claim. I understand that such secondary considerations include, among other things, commercial success of the patented invention, skepticism of those having ordinary skill in the art at the time of the invention, unexpected results of the invention, any long-felt but unsolved need in the art that was satisfied by the alleged invention, the failure of others to make the alleged invention, praise of the alleged invention by those having ordinary skill in the art, and copying of the alleged invention by others in the field. I understand that there must be a nexus—that is, a connection—between any such secondary considerations and the alleged invention. I also understand that contemporaneous and independent invention by others is a secondary consideration tending to show obviousness.

27. I am not aware of any allegations by the named inventor of the '750 Patent or any assignee of the '750 Patent that any secondary considerations tend to rebut the obviousness of the '750 Patent.

28. Additionally, I understand that in considering obviousness, it is important not to use the benefit of hindsight derived from the patent under consideration.

C. Level of Ordinary Skill in the Art

29. When interpreting a patent, I understand that it is important to identify the relevant art pertaining to that patent, as well as the level of ordinary skill in that art at the time of the claimed invention. The “art” is the field of technology to which the patent is related.

30. I am informed and understand that the POSITA is a hypothetical person who is presumed to know the relevant prior art. I understand that the actual inventor’s skill is not determinative of the level of ordinary skill. I further understand that factors that may be considered in determining level of skill include: (i) the types of problems encountered in the art; (ii) prior art solutions to those problems; (iii) the rapidity with which innovations are made; (iv) the sophistication of the technology; and (v) the educational level of active workers in the field. I understand that not all such factors may be present in every case, and one or more of them may predominate.

31. I understand that a POSITA is one who is presumed to be aware of all pertinent art, thinks along conventional wisdom in the art, and is a person of ordinary creativity.

32. A person of ordinary skill in the art at the relevant time, which I have been told to consider as March 12, 2013 (“POSITA”) would have had at least a four-year 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 higher level of education or skill might make up for less experience, and vice versa.

33. Based on my knowledge, skill, and experience, I have an understanding of the capabilities of a person of ordinary skill in the relevant art. For example, from my industry experience, I am familiar with what an engineer designing speech and natural language interfaces would have knowledge of. From teaching and supervising my post-graduate students, I also have an understanding of the knowledge that a person with this academic experience possesses. Furthermore, I possessed the knowledge of a POSITA myself at least as of 1980.

D. Claim Construction

34. I understand that the United States Patent and Trademark Office interprets claim terms in an *inter partes* review proceeding under the same claim

construction standard that is used in a United States federal court. I understand that under this standard, the meaning of claim terms is considered from the viewpoint of a POSITA at the time of the alleged invention.

35. I have been informed that claim terms are generally given their ordinary and customary meaning as understood by one of ordinary skill in the art in light of the specification and the prosecution history pertaining to the patent. I understand, however, that claims terms are generally not limited by the embodiments described in the specification.

36. I understand that in addition to the claims, specification, and prosecution history, other evidence may be considered to ascertain the meaning of claim terms, including textbooks, encyclopedias, articles, and dictionaries. I have been informed that this other evidence is often less significant and less reliable than the claims, specification, and prosecution history.

37. I have given all the claim terms their plain and ordinary meaning, as would have been understood by a person having ordinary skill in the art at the time of the alleged invention, which I understand is March 12, 2013, having taken into consideration the language of the claims, the specification, and the prosecution history of record.

III. SUMMARY OF OPINIONS

38. I understand that Petitioner challenges the validity of Claims 1-18 of the '750 Patent. For the reasons I discuss below, it is my opinion that Claims 1-18 of the '750 patent are rendered obvious by the prior art, according to the following grounds:

Ground	Summary
1	Claims 1-8, 11, 13-15, and 19 are obvious over Basye in view of Shenhav
2	Claims 9 and 16 are obvious over Basye in view of Shenhav and Parker
3	Claims 10 and 18 are obvious over Basye in view of Shenhav and Lebeau
4	Claims 12 and 17 are obvious over Basye in view of Shenhav and Conroy

IV. THE '750 PATENT

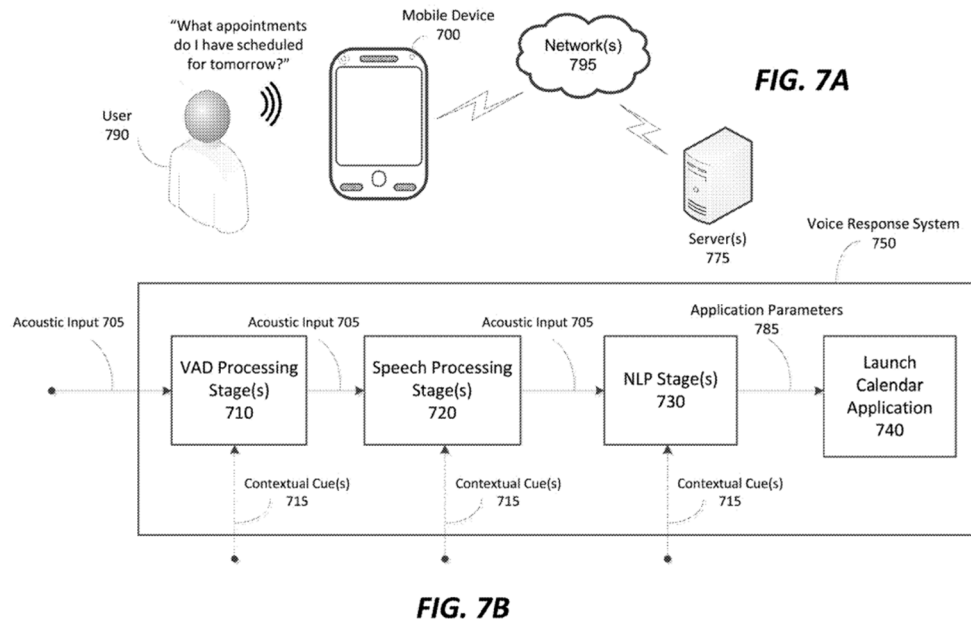
A. Overview of the '750 Patent

39. The '750 Patent is directed to techniques used by mobile devices to detect voice commands in received audio. Ex-1001, Abstract. The techniques disclosed and claimed by the '750 Patent allegedly enable mobile devices to detect voice commands in a way that minimizes the device's power consumption. *Id.*, 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.

40. The patent describes low power processing in the context of a mobile device operating in a “low power mode.” *Id.*, 2:29-41. The patent discloses that “[e]ntering a low power mode may include turning off the display, shutting down one or more processors (e.g., disabling the clock tree associated with one or more processors) and/or deactivating components that would otherwise draw power if left operational.” *Id.*, 5:14-18. By contrast, when the mobile device is 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.

41. An exemplary procedure for processing acoustic input is shown in Figures 7A and 7B:



42. As shown, acoustic input 705 is received by a mobile device 700 and proceeds through a number of processing stages. For example, “VAD [voice activity detection] processing stages 710 ... determine whether the acoustic input includes speech content or corresponds to non-speech acoustic activity.” Ex-1001, 30:15-17. Next, speech processing stages 720 “determine whether the acoustic input includes a voice command.” *Id.*, 30:60-62. After a voice command is detected, “NLP [natural language processing] stages 730 ... evaluate the semantic content of the acoustic input 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.

43. Further, “mobile device 700 is capable of communicating over network(s) 795” and the mobile device’s “voice response system 750 may elect to

have network resources perform one or more processing stages on the acoustic input.” Ex-1001, 32:7-11. Thus, as the patent discloses, “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.

44. The ’750 Patent also discloses that, “when a voice command is detected and the intent of the command ascertained, the further processing that is initiated will depend on the content of the command and/or the operating mode of the mobile device.” *Id.*, 32:14-18. Thus, “[w]hen the mobile device is operating in a low power mode, some commands may provoke a transition of the mobile device into an active mode, while other commands may be carried out without further waking up the mobile device.” *Id.*, 32:18-22.

45. An example of the latter scenario disclosed by the patent is a music application that a mobile device may execute while in a low power mode. In this case, certain application specific commands, such as “next track,” “previous track,” “repeat track,” or “pause music” “may be performed without having to exit a low power mode.” *Id.*, 32:24-27. Thus, in this scenario, acoustic input may be processed in a low power mode in order to detect a voice command, and “the voice command may be carried out without needing to further transition the mobile device into an active mode.” *Id.*, 32:27-32.

B. Prosecution History

46. The '750 Patent was filed on 8/16/2016 as a continuation of U.S. Application 13/795,933, which itself subsequently issued as U.S. Patent No. 11,393,461 (the “'461 patent”).

47. 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.

48. I understand, as informed by counsel, that the Lebeau reference (Ex-1014) was considered by the Examiner during the prosecution of the '750 Patent. However, the examiner did not apply Lebeau's teaching with Basye and Shenhav.

V. TECHNOLOGY BACKGROUND

A. It was well known to use contextual cues in speech recognition systems for power saving benefits.

49. In view of the various applications of speech recognition, there are various considerations that systems use beyond the acoustic input itself, that a POSITA would likely consider when designing a speech recognition system to run on a mobile battery powered device. I have summarized only a few considerations of the electronic device arrangements below.

50. For example, U.S. Patent No. 9,031,847 to Sarin (“Sarin”, Ex-1019) describes how a voice-controlled application can use contextual cues to assist in recognizing commands and using resources efficiently. Sarin teaches that contextual clues include “which applications or functions are active on a device, time of day, and location,” such that if the time and day and location indicate that the user is at work, “voice recognition relating to image capture operations can be turned off” so as to avoid unintended camera operations. Ex-1019, 11:32-47. Sarin also describes how the user’s recent activity, such as an active phone call, can be considered when determining whether voice recognition relating to image capture operations can be turned off, so as to avoid unintended camera operation. *Id.*, 11:47-52. This is beneficial because it saves battery by preventing voice recognition, which may require significant computation and therefore power, relating to image capture operations when it is highly unlikely that the user would invoke its mobile device’s camera since the user is at work or otherwise using their mobile device.

51. Similarly, U.S. Patent Application Publication No. 2012/0034904 to Lebeau (“Lebeau,” Ex-1014) titled “Automatically Monitoring for Voice Input Based on Context,” explains using contextual cues so as to evaluate the likelihood that a user would speak a verbal search request. Ex-1014 ¶¶3, 22 (“the mobile computing device can attempt to infer whether the current context indicates that the user has at least a threshold likelihood of providing voice input and, if so, monitoring

for voice input in response.”), 43 (“threshold likelihood of [the user] providing a voice-based request to the mobile computing device”). Lebeau explains:

current context associated with a mobile computing device (and/or with a user of the mobile computing device) can include a context external to the device, such as that represents an environment around the device, or a context internal to the device such as historical information about the device that is stored in the device. Context external to the device can include, for example, the physical location where the mobile computing device is located (e.g., home, work, car, etc., as determined by GPS in the device or other techniques), and motion of the mobile computing device (e.g., accelerating, stationary, etc.). Context that is internal to the device can include recent activity on the mobile computing device (e.g., social network activity, emails sent/received, telephone calls made/received, etc.). The current context for a mobile computing device (and/or its user) is separate from user input itself that would direct the device to listen for spoken input.

Ex-1014 ¶3.

52. Specifically, Lebeau’s system includes a context determination unit “that is configured to detect a current context associated with the mobile computing device,” indicating a current state of the surrounding environment. Ex-1014 ¶8. 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.

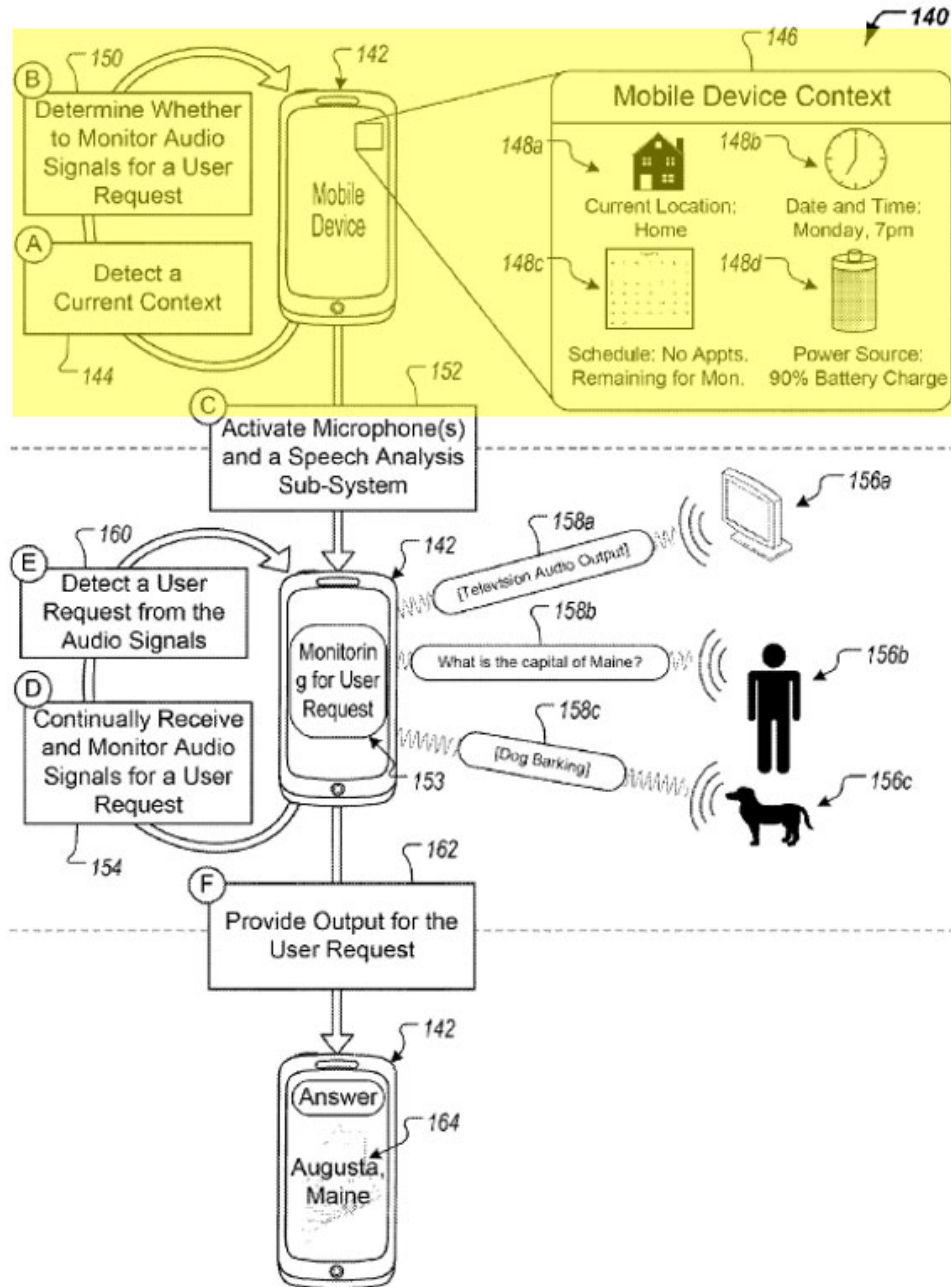


FIG. 1B

53. At step B, Lebeau explains that its device will “determine[] to proceed with monitoring ambient audio signals for a user request based on an inferred likelihood that the user will provide a user request and convenience to both the user

and the mobile computing device 142, as indicated by the context 146”. Ex-1014

¶43. The likelihood that the user will provide a user request,

can be inferred from, at least, the time (7 pm) and the user's schedule. Although it is evening, the user is likely to have not gone to bed yet (it is only 7 pm) and the user does not have any appointments for the remainder of the evening—the user's anticipated free time over the next several hours can indicate at least a threshold likelihood of providing a voice-based request to the mobile computing device 142. Monitoring for voice input can be convenient for the user based on, at least, the mobile computing device 142 being located at the user's home where the user may be more than an arm's length away from the mobile computing device 142 (e.g., the user may be moving around the house such that it may be more convenient for a user to simply speak his/her requests instead of having to locate the mobile computing device 142 to manually prompt the computing device 142 for each request). Additionally, monitoring the voice input can be convenient for the mobile computing device based on, at least, the battery having at least a threshold charge and based on a projection that monitoring will only last for a limited period of time (e.g., the mobile computing device 142 can forecast that the user will likely go to bed within a few hours).

Ex-1014 ¶43. Therefore, Lebeau uses the user's time, location, schedule, and the mobile device's battery status as contextual cues of when to monitor for the user's speech, which “help[s] conserve the amount of energy consumed by a mobile computing device, especially when the mobile computing device is using a portable power source, such as a battery.” Ex-1014 ¶10.

54. Another example that Lebeau provides is in the car-context, where the computing device “may determine that it is currently travelling in a car based on a detected high rate of speed at which the device is travelling (e.g., using any of a variety of motion sensors that are standard components of the device).” Ex-1014

¶22. Using the user's location based on "motion sensors," Lebeau can determine whether to monitor for voice input based on this current context. Ex-1014 ¶22. Like above, this "help[s] conserve the amount of energy consumed by a mobile computing device, especially when the mobile computing device is using a portable power source, such as a battery" by not monitoring for voice when the motions indicate that it would or would not be convenient for the user to speak a user request. Ex-1014 ¶10.

55. Sarin also notes that "other states and state transitions can apply where other device components (e.g., a gyroscope, an accelerometer) provide additional contextual cues (e.g., detecting the orientation of the device as being ready perform an operation such as capturing an image or recording video) for recognizing voice commands." Ex-1019, 12:16-24.

56. Using accelerometers in speech recognition systems is well-known. Like Sarin, Binder teaches a mobile device with a motion sensor such as an accelerometer. Ex-1028 ¶52. Binder further describes that "the context of the electronic device is used to help determine whether and how to operate the voice trigger," and lists a number of devices useful for detecting such context, including accelerometers. Ex-1028 ¶128. Binder explains that the context of the device can be used to adjust how and whether the voice trigger operates. Ex-1028 ¶129 (further explaining that the voice trigger is deactivated when the phone is in a predetermined

orientation, during predetermined time periods, when the phone is in a silent or do not disturb mode, in an enclosed space, or near other devices). The accelerometer can be used to determine that the device is face-up or face down so that the system knows to activate the voice trigger when the device is placed “face-up.” Ex-1028 ¶131; *see also* Ex-1028 ¶¶166-167 (“electronic device detects its orientation using light sensors (including cameras), microphones, proximity sensors, magnetic sensors, accelerometers, gyroscopes, tilt sensors, and the like.”). Like Lebeau’s vehicle example, Binder uses the accelerometer to determine that the device is in a vehicle by monitoring the motion and if it is determined that device is likely in a vehicle because it is traveling above 20 miles per hour, then it “is determined to be travelling along a road” and “then the voice trigger remains active and/or in a high-power or more sensitive state.” Ex-1028 ¶133. Specifically, Binder’s accelerometer detects “the motion of the electronic device and/or the external audio system 416,” which is then used to:

determine whether the voice trigger system 400 should monitor ambient sound using the on-device microphone 230 or an external microphone 418. Specifically, the difference between the motion of the device and the external audio system 416 provides information about whether the external audio system 416 is actually in use. For example, if both the device and a wireless headset are moving (or not moving) substantially identically, it may be determined that the headset is not in use or is not being worn. This may occur, for example, because both devices are near to each other and idle (e.g., sitting on a table or stored in a pocket, bag, purse, drawer, etc.). Accordingly, under these conditions, the voice trigger system 400 monitors the on-device microphone, because it is unlikely that the headset is actually being

used. If there is a difference in motion between the wireless headset and the device, however, it is determined that the headset is being worn by a user. These conditions may occur, for example, because the device has been set down (e.g., on a surface or in a bag), while the headset is being worn on the user's head (which will likely move at least a small amount, even when the wearer is relatively still). Under these conditions, because it is likely that the headset is being worn, the voice trigger system 400 maintains an active communication link and monitors the microphone 418 of the headset instead of (or in addition to) the on-device microphone 230. And because this technique focuses on the difference in the motion of the device and the headset, motion that is common to both devices can be canceled out. This may be useful, for example, when a user is using a headset in a moving vehicle, where the device (e.g., a cellular phone) is resting in a cup holder, empty seat, or in the user's pocket, and the headset is worn on the user's head. Once the motion that is common to both devices is cancelled out (e.g., the vehicle's motion), the relative motion of the headset as compared to the device (if any) can be determined in order to determine whether the headset is likely in use (or, whether the headset is not being worn).

Ex-1028 ¶139.

57. Using contextual cues of battery power or power status are frequently another contextual cue or consideration for power management in a speech recognition system for a mobile device. Lebeau is a reference that discusses considering the mobile device's battery. For example, in the car context explained above, Lebeau may determine that the user is currently in a car "based on the device being docked in a car-adapted mobile device dock (e.g., detecting a pin arrangement for a physical electronic connection between the mobile computing device and the dock). The mobile computing device can determine whether to monitor for voice input based on this current context." Ex-1014 ¶22; *see also id.* ¶28. Similarly,

Lebeau teaches that the mobile computing device will continue monitoring for voice input until the condition of the “battery for the mobile computing device running low on stored power (e.g., below 25% charge remaining in the battery).” *Id.* ¶24. Lebeau also teaches that the mobile device can be receiving power from a power cord that is plugged into an outlet. Ex-1014 ¶30. This contextual cue allows the device to determine whether the “potential convenience to the user of monitoring for voice input is outweighed by the inconvenience of the mobile computing device ... potentially draining the battery in a relatively short period of time.” Ex-1014 ¶31.

58. Battery status is also discussed as consideration in a speech recognition system disclosed in U.S. Patent Application Publication No. 2014/0163978 (“Basye,” Ex-1005). Basye explains that by keeping the computing device in a persistent active state, the first computing device may consume unacceptable amounts of energy to maintain the persistently active state. Ex-1005 ¶9. “Such energy demands may prove especially problematic for mobile computing devices that rely on battery power.” *Id.* Basye is thus directed to power management for speech recognition. *Id.* ¶10. Specifically Basye notes that a device may be provided power wirelessly or via a portable power supply, or may be powered using a wired power supply. *Id.* ¶43 The power supply unit communicates “the level of power that it can supply to the power management subsystem 100 (e.g., a percentage of battery life remaining, whether the power supply 218 is plugged into an electrical outlet,

etc.).” *Id.* Basye’s “on-device recognition selection element 514 may be automatically deselected (and on-device speech recognition capabilities automatically disabled) if the power supply 218 drops below a threshold power supply level (e.g., a battery charge percentage), as on-device speech recognition capabilities as implemented by the processing unit 202 and/or the application processing module 112 may require a relatively large power draw.” *Id.* ¶79. Basye teaches that its implementation is “particularly advantageous for computing devices that rely on battery power, all computing devices for which power management may be desirable can benefit from the principles of the present disclosure.” *Id.* ¶14.

59. U.S. Patent Application Publication No. 20130006638 to Lindalh (“Lindalh,” Ex-1029) is another reference that emphasizes the usefulness of context cues in speech processing. Lindalh recognizes that portable electronic devices generally have limited processing power and are not always activity connected to remote databases and services of interest and are often not “contextually aware,” which make it difficult to use for voice-based control functions. Ex-1029 ¶5. Lindalh’s user device is used to “receive voice commands” and if the device “has sufficient processing power, the voice commands may be partly or fully processed by the user device 12 (e.g., using a speech recognition engine...).” Ex-1029 ¶26. If the device does not have enough processing power, the voice commands are transmitted to a computer for further processing. *Id.* For example, Lindalh provides

the voice command of “find Italian restaurants near me,” and the device may transmit the voice command and contextual information to a computer due to “an inability to perform adequate speech processing due to a lack of available processing power.” Ex-1029 ¶36.

60. Thus, using context cues associated with the user or the device, allows for the device to improve the processing of the voice input, which leads to power savings.

B. Noise reduction and/or cancellation techniques were well known in order to better process speech input and speech recognition.

61. A POSITA in speech processing and recognition would have understood and applied noise reduction techniques well before 2013. For example, U.S. Patent Application Publication No. 20030179888A1 to Burnett et al. (“Burnett,” Ex-1026), explains that many noise suppression algorithms and techniques have been developed over the years and most of the noise suppression systems in use today for speech communication systems are based on a single-microphone spectral subtraction technique first developed in the 1970's and described in Ex-1026 ¶4 (citing S. F. Boll in “Suppression of Acoustic Noise in Speech using Spectral Subtraction,” IEEE Trans. on ASSP, pp. 113-120, 1979). Burnett further explains a “Pathfinder system,” which is a digital signal processing (DSP) based acoustic noise suppression system, which can use voice activity

detection information and received acoustic information to reduce or eliminate noise in desired acoustic signals “by estimating the noise waveform and subtracting it from a signal including both speech and noise.” Ex-1026 ¶41; *see also id.* ¶¶ 175, 193, Cl. 18.

62. Another noise suppression method that subtracts the background noise from a speech input is U.S. Patent Application Publication No. 20130332165 to Beckley et al. (“Beckley,” Ex-1024), explains how POSITAs were focused on improving speech recognition systems by reducing the background noise. Specifically, Beckley discusses how speech recognition systems, especially those with “always-on” speech recognition (referred to by Beckley as “ASR”), “have difficulty handling ambient noise, such as background conversation or other undesired noise.” Ex-1024 ¶6. Beckley explains that the “background conversation, for instance, may cause the system to recognize a command that was not intended by the user of the system, thus leading to a number of false positives and misrecognitions. For example, in a car environment, while the driver might be the issuer of voice commands to control the ASR system, the presence of passengers and ensuing conversation between them can greatly reduce the performance of the ASR system as their conversations may lead to commands that are false positives or misrecognitions.” *Id.*

63. Beckley seeks to improve speech recognition by: (i) obtaining a voice sample from a speaker; (ii) storing the voice sample of the speaker as a voice model in a voice model database; (iii) identifying an area from which sound matching the voice model for the speaker is coming; and (iv) providing one or more audio signals corresponding to sound received from the identified area to the speech recognition system for processing. *Id.* ¶7. Beckley further includes separating ambient noise from speech in the one or more audio signals and providing the resulting speech signal to the speech recognition system for processing. *Id.* ¶11. Specifically, Beckley receives a digital audio signal and may generate a “voice/speech reference signal” and a “noise reference signal.” *Id.* ¶88. Beckley’s noise reference refiner 220a may “reduce or effectively eliminate the residual desired audio from the noise reference signal 218 in order to generate a refined noise reference signal 222 a.” *Id.* Beckley provides an example that “the refined noise reference signal 222 a may be filtered and **subtracted** from the desired audio reference signal 216 in order to reduce noise in the desired audio.” *Id.* Beckley further explains that “sound from other areas than that of the area in which the identified speaker is located may be subtracted from sounds of the speaker’s area.” *Id.* ¶95. The resulting speech signal is then provided to the speech recognition system. *Id.* Thus, Beckley improves the quality of the speech input by removing the undesired audio, such as ambient noise.

64. Similarly, U.S. Patent Application Publication No. 20120296655A1 to Kristjansson et al. (“Kristjansson”) discusses the advantages of using “stored noise profiles.” Specifically, Kristjansson teaches that:

“speech recognition is used to transcribe utterances that may have been input in a wide variety of noisy environments, such as when the user is in a crowded room (e.g., a cafeteria), walking down the street, or in the presence of a radio or television. To filter this noise, some noise reduction algorithms require a sample of audio from the environment, without the user's speech. ***By predictively pre-recording audio before the use initiates the voice input (e.g., before the user presses a button), such a background recording becomes available for use by the noise reduction algorithm, thereby improving recognition accuracy.***”

Ex-1018 ¶42.

65. Kristjansson further explains that by “invoking the functionalities of the mobile device may include commencing a background audio recording of audio prior to a user's button press. The background audio may be used as input to the speech recognition engine, ***to estimate noise models for noise suppression, and/or estimate noise models and/or levels for speech endpoint detection.***” Ex-1018 ¶57.

66. Similarly, U.S. Patent Application Publication No. 20140350927 to Yamabe et al. (“Yamabe,” Ex-1025) explains how in the field of speech recognition, noise-signal suppression processing is performed prior to speech recognition processing, thereby suppressing unnecessary noise other than speech signals and improving the accuracy of speech recognition. Ex-1025 ¶5. Yamabe notes that typical noise reduction systems use noise spectrum detection, which involves

estimating noise spectra and “subtract[ing] a noise component form an input signal including noise, thereby reducing the noise.” Ex-1025 ¶6. Yamabe recognizes that such spectrum subtraction methods uses “stationary energy, such as white noise and air conditioners, as noise with relative ease,” and have a high suppression effect but a signal that is closer to a tone signal might be difficult to classify. Ex-1025 ¶7. Similarly, for sounds that have “a harmonic structure which is a characteristic of speech” it is difficult to estimate a noise spectrum. Ex-1025 ¶7. Yamabe provides an improvement to spectrum subtraction methods, by “extracting a peak spectrum” to be suppressed from the noise signal by “subtracting a signal in the time domain obtained by the frequency/time converting unit 207 converting the extracted peak spectrum having independence from the sound signals in the time domain collected by the sound collecting unit 201.” Ex-1025 ¶132. Yamabe thus discusses how noise reduction systems commonly used “spectrum subtraction” to reduce the noise in a speech signal.

67. U.S. Patent Application Publication 2020/0279576 to Binder et al. (“Binder,” Ex-1028) is another example of noise cancellation. For example, if a voice trigger includes noise, Binder implements a “sound cancellation functionality.” Ex-1028 ¶124. Binder explains “noise cancellation” that is performed by the audio digital signal processor, that “removes unwanted noise or sounds from the sound input prior to it being processed by the sound detectors.” *Id.* Binder gives

an example that sometimes the unwanted noise is background noise from the user's environment or just any "sounds above, below, or at predetermined amplitudes or frequencies." *Id.* For example, sounds above the typical human vocal range (e.g., 3,000 Hz) is filtered out or removed from the signal. *Id.* Binder explains that this process allows the sound input to be cleaner and more accurate. *Id.* ¶126.

68. U.S. Patent No. 8,694,316 to Pickering et al. ("Pickering," Ex-1023) describes a system with "an acoustic front end (AFE) for receiving an audio signal from an input device, removing noise, identifying the parts of the audio signal containing speech, and parameterizing sections of the identified speech." Ex-1023, 4:3-16. Specifically, the AFE is "responsible for: 1. pre-processing the audio signal to remove noise; 2. establishing which part of the audio signal contains valid speech; and 3. parameterizing successive sections of the signal to be matched by the recognizer." *Id.*, 6:6-10.

69. Noise cancellation and noise filtering is also noted in U.S. Patent Application Publication No. 2014/0006825 ("Shenhav," Ex-1006). Shenhav teaches a mobile device with an audio sensor module that may include a "variety of noise reducing or noise cancellation filters." Ex-1006 ¶40. Shenhav provides "removing background noise" as a filter used to generate a "filtered sound include[s] processing the sound input to only include those portions of the sound input that match audio frequencies associated with human speech." Ex-1006 ¶49.

70. Thus, as shown by at least these references, POSITAs were implementing noise reduction and/or cancelling techniques before 2013, in order to focus the sound input on the speech component.

C. Voice activity detection was well known to detect human speech.

71. As I explained above, voice activity detection (VAD) methods were well known to use in “signal processing systems to denoise acoustic signals.” Ex-1026, Abstract. Burnett describes a system that provides that “the VAD signal is processed independently of the noise suppression system, so that the receipt and processing of VAD information is independent from the processing associated with the noise suppression, but the embodiments are not so limited.” Ex-1026 ¶32. Burnett generally defines “VAD” as “a vector or array signal, data, or information that in some manner represents the occurrence of speech in the digital or analog domain.” Ex-1026 ¶34. Burnett’s invention is used with noise suppression systems like the Pathfinder Noise Suppression System, as described above. Burnett describes a VAD technique which includes evaluating “spectral information.” Ex-1026 ¶55.

72. Shenhav also describes using “spectral analysis of the sound signals.” Cl. 8. Specifically, Shenhav teaches comparing a spectral representation of the received sound to spectral representation of a predetermined wake-up phrase. Ex-1006 ¶¶38, 45. To do so, Shenhav applies a filter that may include “spectral equalization, or the like” to the sound signal, isolating a filtered “sound input that

matches audio frequencies associated with human speech.” Ex-1006 ¶¶48-49.

Shenhav therefore teaches using spectral analysis on the acoustic input to evaluate whether a spectrum of the acoustic input is indicative of voice activity (i.e. “human speech”), as opposed to other non-speech activity. After “only those portions of the sound input ... associated with human speech” are prepared, the filtered signal is further processed, to determine whether there is “a threshold probability that there is a match of the sound to a wake-up phrase.” Ex-1006 ¶50.

73. Basye also teaches determining whether audio input includes speech. Basye’s speech detection module applies “voice activity detection...based on various quantitative aspects of the audio input, such as the spectral slope between one or more frames of the audio input; the energy levels of the audio input in one or more spectral bands; the signal-to-noise ratios of the audio input in one or more spectral bands; or other quantitative aspects.” Ex-1005 ¶21.

74. Thus, voice activity detection techniques were well known to use spectral analysis to detect human speech before 2013.

VI. BRIEF DESCRIPTION OF THE APPLIED PRIOR ART

75. I have analyzed the prior art references used in the grounds below, and it is my opinion that they are analogous to the ’750 patent because they are within the same field of endeavor and reasonably pertinent to at least one problem addressed by the ’750 patent.

76. In particular, Basye (like the '750 patent) discloses a mobile computing device that detects and recognizes voice commands assisted by a remote server. Like the '750 patent, Basye is concerned with minimizing power consumption of its mobile computing device when it performs speech recognition, and, as such, discloses a “power management subsystem” that includes “selectively activated modules.” Ex-1005 ¶11. Basye discloses that these modules can be “implemented as dedicated hardware (such as an integrated circuit, a digital signal processor or other type of processor) that may be switched from a low-power, deactivated state with relatively lesser functionality, to a high-power, activated state with relatively greater functionality, and vice versa.” *Id.*

77. Shenhav, like the '750 patent, is also concerned with conserving device power. Similar to the '750 patent, Shenhav discloses a mobile device that utilizes a remote server to perform speech recognition. And, just as the '750 patent discloses, Shenhav's mobile device receives acoustic input and provides that acoustic input to the remote server while the mobile device is in a low power state.

78. Parker also discloses a computing device that listens for audio name information, i.e. the name of the device, where the device listens for the name information a power conservation mode. Upon detecting its name, the device switches to a higher power mode that permits speech recognition of more general

audio commands. Like the '750 patent, Basye, and Shenhav, Parker also relies on a remote server for speech recognition services.

79. Lebeau discloses a computing device that operates in two modes of operation so that the device may switch to allow for voice input that indicates a request to be performed. Lebeau uses contextual cues 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. Like the '750 patent, Basye, and Shenhav, Lebeau also relies on uses contextual cues when processing a voice input.

80. Conroy discloses a computing device with a CPU and audio subsystem which includes a digital signal processor. Conroy's digital signal processor is able to perform tasks with little or no involvement of the CPU so that the CPU is free to enter a low power state. Like the '750 patent, Basye, and Shenhav, Conroy's computing device is able to continue processing while the device in a low power mode, which extends the battery life of the device.

A. Basye (Ex-1005)

81. Basye is titled "Speech recognition power management." Ex-1005, Cover (54).

82. Basye discloses techniques for reducing power consumption in computing devices during speech recognition. Basye observed that conventional

speech recognition techniques are disadvantageous because computing devices may consume unacceptable amounts of energy by continuously powering the hardware components that implement the device's speech recognition functionality. *See id.* ¶9.

In particular, Basye discloses:

In ... current approaches [to speech recognition], the first computing device may be configured to remain in a persistently active state. In such a persistently active state, the first computing device may continuously maintain a network connection to the second computing device. The first computing device may also continue to power any hardware used to implement its own speech recognition capabilities. One drawback of these approaches, among others, is that the first computing device may consume unacceptable amounts of energy to maintain the persistently active state. ***Such energy demands may prove especially problematic for mobile computing devices that rely on battery power.***

Id.

83. To address these issues, Basye discloses a computing device (which can be a mobile device), that includes a “power management subsystem.” I have reproduced Figure 1 of Basye, which depicts the power management subsystem.

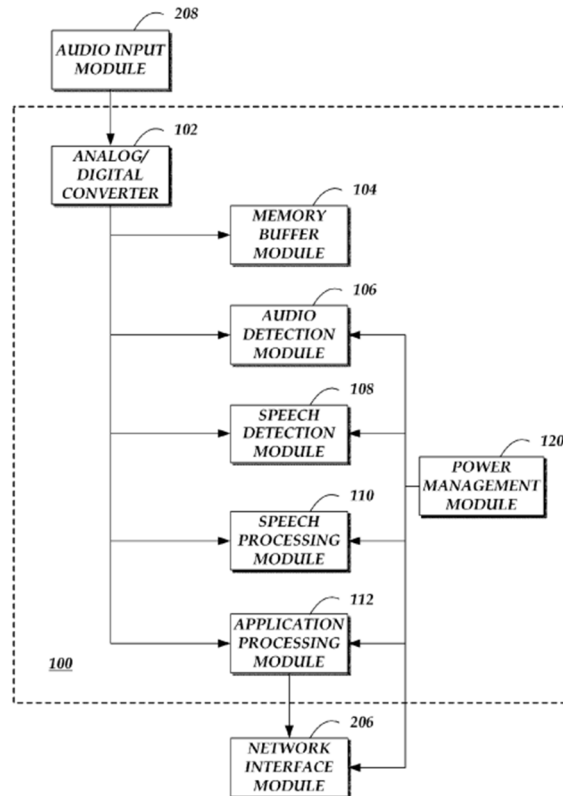


Fig.1.

Ex-1005, Fig. 1

84. In response to receiving audio input, the power management subsystem selectively activates one or more modules of the computing device as the audio input progresses through the subsystem. *Id.* ¶10.

85. The first modules depicted are the audio input module 208 and the analog/digital converter 102, which are responsible for receiving audio input from the environment of the computing device and digitizing that audio. As Basye discloses:

The analog/digital converter 102 may receive an audio input from an audio input module 208. The audio input module 208 is discussed in further detail below with respect to FIG. 2. The analog/digital converter

102 may be configured to convert analog audio input to digital audio input for processing by the other components of the power management subsystem 100. In embodiments in which the audio input module 208 obtains digital audio input (e.g., the audio input module 208 includes a digital microphone or other digital audio input device), the analog/digital converter 102 may optionally be omitted from the power management subsystem 100. Thus, ***the audio input module 208 may provide audio input directly to the other modules of the power management subsystem 100.***

Id. ¶16.

86. The next key module of the power management system is the audio detection module 106. When activated, audio detection module 106 determines an energy level of the audio input in order to determine if the audio is indicative of speech. In particular, Basye discloses:

The audio detection module 106 may process audio input to determine an energy level of the audio input. In some embodiments, the audio detection module 106 includes a low-power digital signal processor (or other type of processor) configured to determine an energy level (such as a volume, intensity, amplitude, etc.) of an obtained audio input and for comparing the energy level of the audio input to an energy level threshold. The energy level threshold may be set according to user input, or may be set automatically by the power management subsystem 100 as further discussed below with respect to FIG. 3. In some embodiments, the audio detection module 106 is further configured to determine that the audio input has an energy level satisfying a threshold for at least a threshold duration of time. In such embodiments, high-energy audio inputs of relatively short duration, which may correspond to sudden noises that are relatively unlikely to include speech, may be ignored and not processed by other components of the power management subsystem 100.

Id. ¶19.

87. Next, if the audio detection module determines that the audio input has an appropriate energy level, speech detection module 108 is activated. Speech detection module 108 is responsible for analyzing the audio input to determine whether the audio contains speech. To accomplish this, speech detection module 108 employs voice activity detection (VAD) techniques. According to Basye,

[S]peech detection module 108 may process audio input to determine whether the audio input includes speech. In some embodiments, the speech detection module 108 includes a low-power digital signal processor (or other type of processor) configured to implement one or more techniques to determine whether the audio input includes speech. In some embodiments, the speech detection module 108 applies voice activity detection (VAD) techniques. Such techniques may determine whether speech is present in an audio input based on various quantitative aspects of the audio input, such as the spectral slope between one or more frames of the audio input; the energy levels of the audio input in one or more spectral bands; the signal-to-noise ratios of the audio input in one or more spectral bands; or other quantitative aspects. In other embodiments, the speech detection module 108 implements a limited classifier configured to distinguish speech from background noise. The classifier may be implemented by techniques such as linear classifiers, support vector machines, and decision trees. In still other embodiments, the speech detection module 108 applies Hidden Markov Model (HMM) or Gaussian Mixture Model (GMM) techniques to compare the audio input to one or more acoustic models, which acoustic models may include models corresponding to speech, noise (such as environmental noise or background noise), or silence. Still other techniques may be used to determine whether speech is present in the audio input.

Id. ¶21.

88. Next, when speech is detected in the input audio, speech processing module 110 is activated. *Id.* ¶23. Speech processing module 110 processes the audio

input to determine whether a keyword is included in the speech. *Id.* ¶24. One type of keyword “that prompts the activation of one or more components” is referred to by Basye as a “wakeword.” *Id.* ¶10. In particular, Basye discloses:

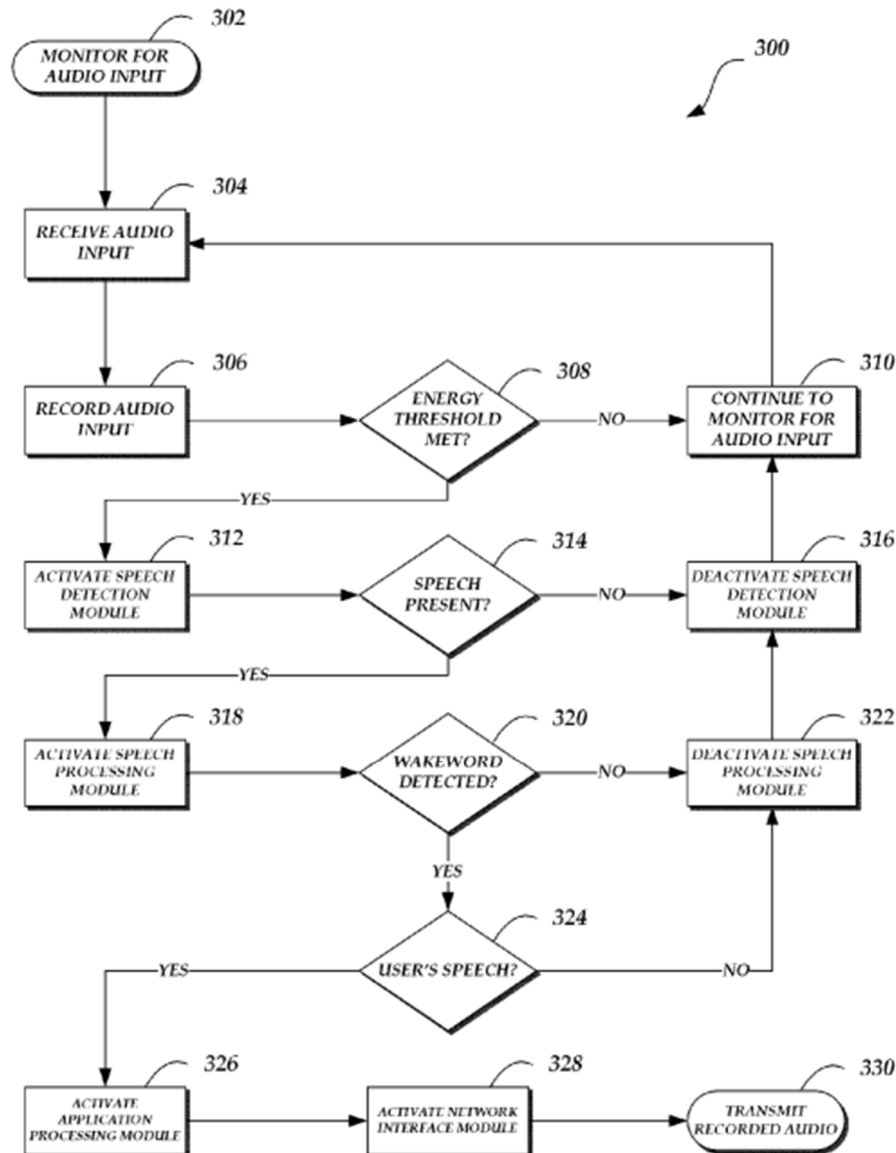
The speech processing module 110 may process the audio input to determine whether a keyword is included in the speech. In some embodiments, the speech processing module 110 includes a microprocessor configured to detect a keyword in the speech, such as a wakeword or sleepword. The speech processing module 110 may be configured to detect the keyword using HMM techniques, GMM techniques, or other speech recognition techniques.

The speech processing module 110 may be able to separate speech that incidentally includes a keyword from a deliberate utterance of the keyword by determining whether the keyword was spoken immediately before or after one or more other phonemes or words. For example, if the keyword is “ten,” the speech processing module 110 may be able to distinguish the user saying “ten” by itself from the user saying “ten” incidentally as part of the word “Tennessee,” the word “forgotten,” the word “stent,” or the phrase “ten bucks.”

Id. ¶¶24-25.

89. Next, if speech processing module 110 detects a keyword or wakeword, application processing module 112 and network interface module 206 may be activated. *Id.* ¶28. Basye discloses that 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, where speech recognition results (like a response to a query) may be returned to the application processing module 112. *Id.* ¶31.

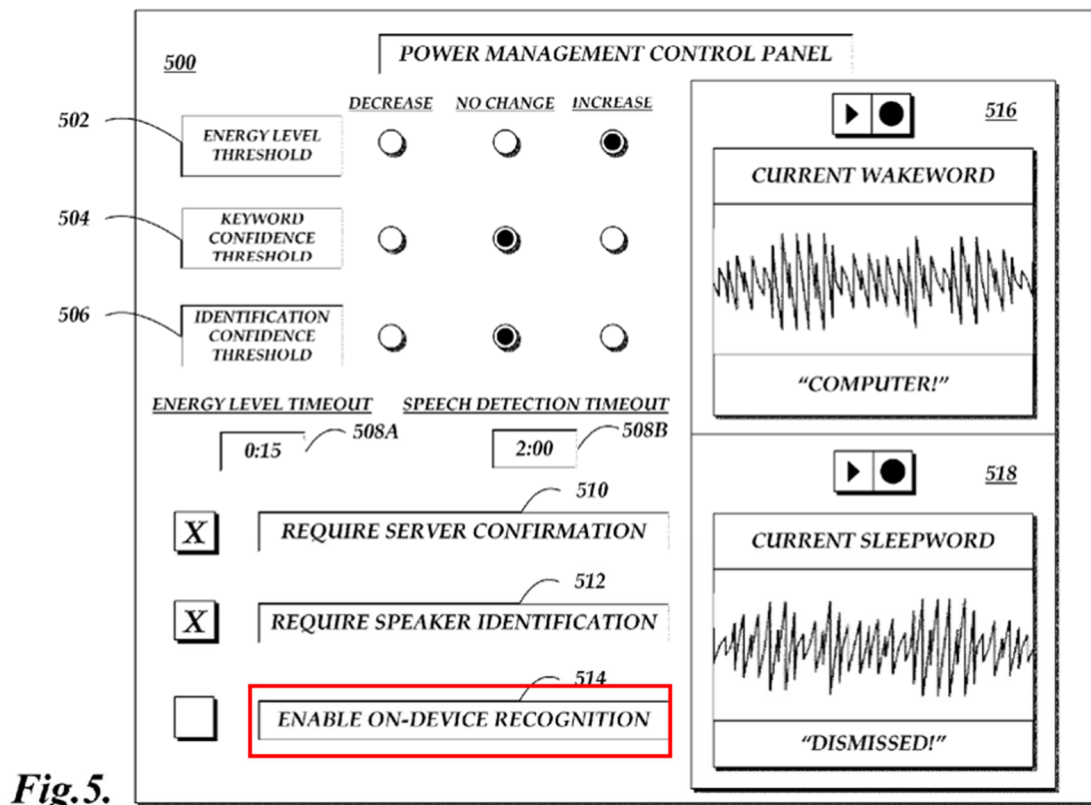
90. Basye discloses an exemplary process for the power management subsystem during the processing of audio input, shown below in Figure 3.



Ex-1005, Fig. 3

91. Although Basye discloses reliance on a speech recognition server to process audio input, the reference also discloses that its computing device can be configured via a user interface to perform on-device speech recognition, as opposed

to relying on the server. More particularly, Basye's user interface allows the user to "select whether the user computing device ... generates speech recognition results by itself, or whether audio inputs are routed to the speech recognition server ... for processing into speech recognition results." *Id.* ¶79. I have reproduced the user interface below in Figure 5, highlighting the selection box that allows the user to enable on-device speech recognition.



Ex-1005, Fig. 5 (annotated)

92. I also understand that Basye incorporates Gruber by reference in its entirety. *Id.* ¶13. I have been informed that when a document is incorporated by reference by another document, as Basye does with respect to Gruber, the

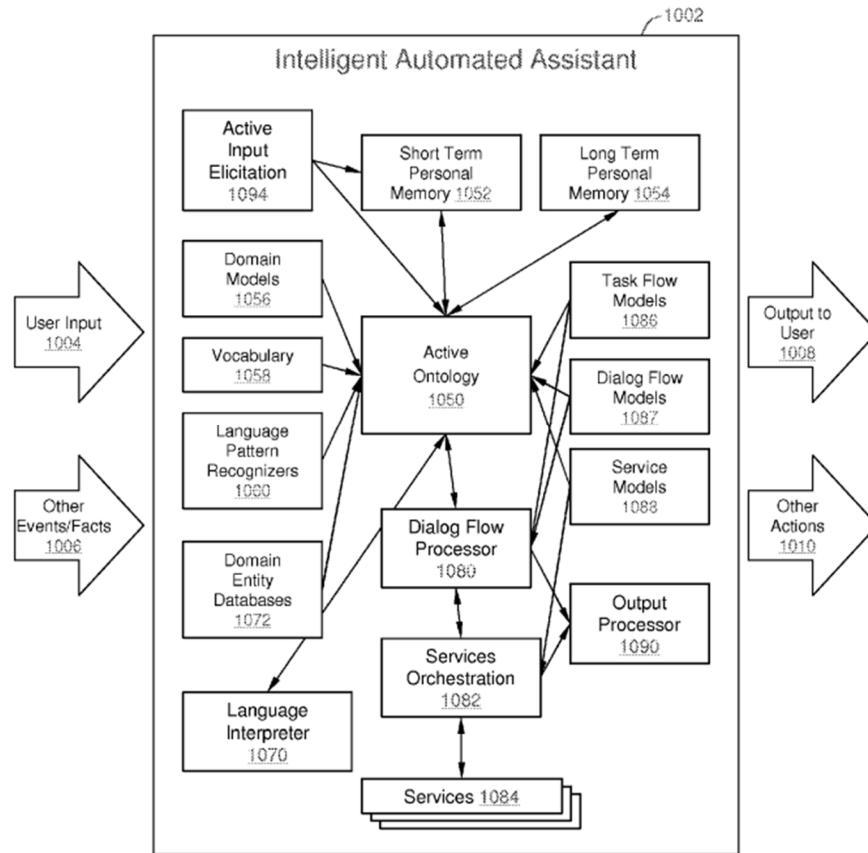
incorporated document (here, Gruber) becomes a part of the incorporating document (here, Basye).

93. 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.” Ex-1008, Abstract. More particularly, a person of ordinary skill in the art would recognize that Gruber’s intelligent assistant performs speech recognition. For example, Gruber discloses:

the intelligent automated assistant integrates a variety of capabilities provided by different software components (e.g., for supporting natural language recognition and dialog, multimodal input, personal information management, task flow management, orchestrating distributed services, and the like). Furthermore, to offer intelligent interfaces and useful functionality to users, the intelligent automated assistant of the present invention may, in at least some embodiments, coordinate these components and services. The conversation interface, and the ability to obtain information and perform follow-on task, are implemented, in at least some embodiments, by coordinating various components such as language components, dialog components, task management components, information management components and/or a plurality of external services.

Id. ¶9.

94. Basye refers to Gruber’s system (i.e., its intelligent assistant) as an “intelligent *agent* frontend.” Ex-1005 ¶13. Gruber sets out a high level diagram of its intelligent assistant in Figure 1.



Ex-1008, Fig. 1

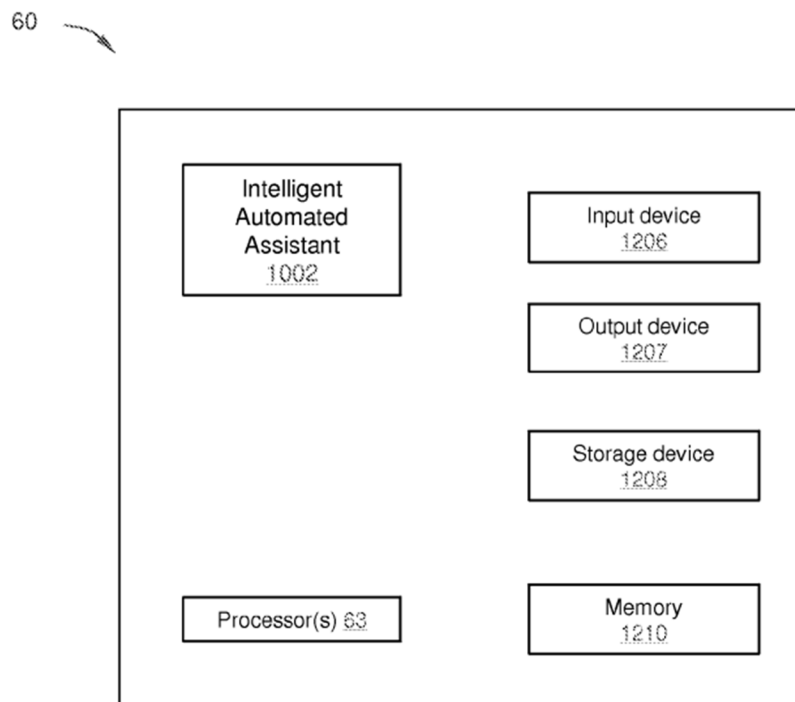
95. Gruber discloses that its intelligent agent can be implemented on a standalone computing system. Ex-1008 ¶83. In particular, Gruber discloses:

In one embodiment, the system of the present invention is implemented on a standalone computing system. Referring now to FIG. 4, there is shown a block diagram depicting an architecture for implementing at least a portion of an intelligent automated assistant on a standalone computing system, according to at least one embodiment. Computing device 60 includes processor(s) 63 which run software for implementing intelligent automated assistant 1002. Input device 1206 can be of any type suitable for receiving user input, including for example a keyboard, touchscreen, microphone (for example, for voice input), mouse, touchpad, trackball, five-way switch, joystick, and/or any combination thereof. Output device 1207 can be a screen, speaker, printer, and/or any combination thereof. Memory 1210 can be random-

access memory having a structure and architecture as are known in the art, for use by processor(s) 63 in the course of running software. Storage device 1208 can be any magnetic, optical, and/or electrical storage device for storage of data in digital form; examples include flash memory, magnetic hard drive, CD-ROM, and/or the like.

Id.

96. Gruber's assistant, as implemented on a standalone computer, is shown in Figure 4 below.



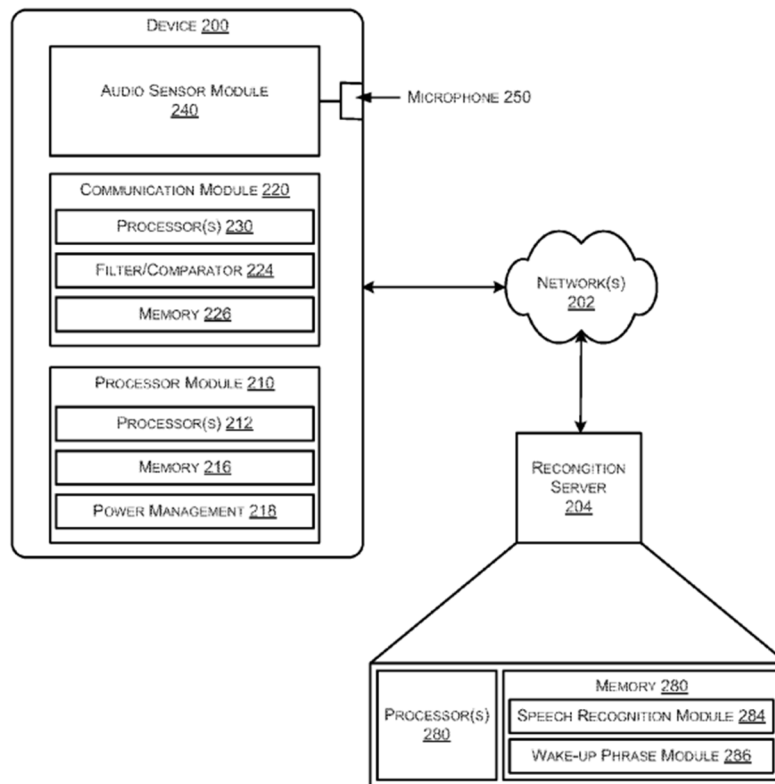
Ex-1008, Fig. 4

97. Thus, it is my opinion that Gruber's intelligent assistant (or "intelligent agent") may be implemented on Basye's application processing module 112 for performing on-device speech recognition. Ex-1008 ¶¶13, 29.

B. Shenhav (Ex-1006)

98. Shenhav is titled “SYSTEMS AND METHODS TO WAKE UP A DEVICE FROM A POWER CONSERVATION STATE.” Ex-1006, Cover (54).

99. Shenhav is generally directed to “waking up devices from power conservation states.” *Id.* ¶1. To achieve this result, Shenhav discloses a mobile device that communicates with a recognition server over a network, as shown in Figure 2. *Id.* ¶27.

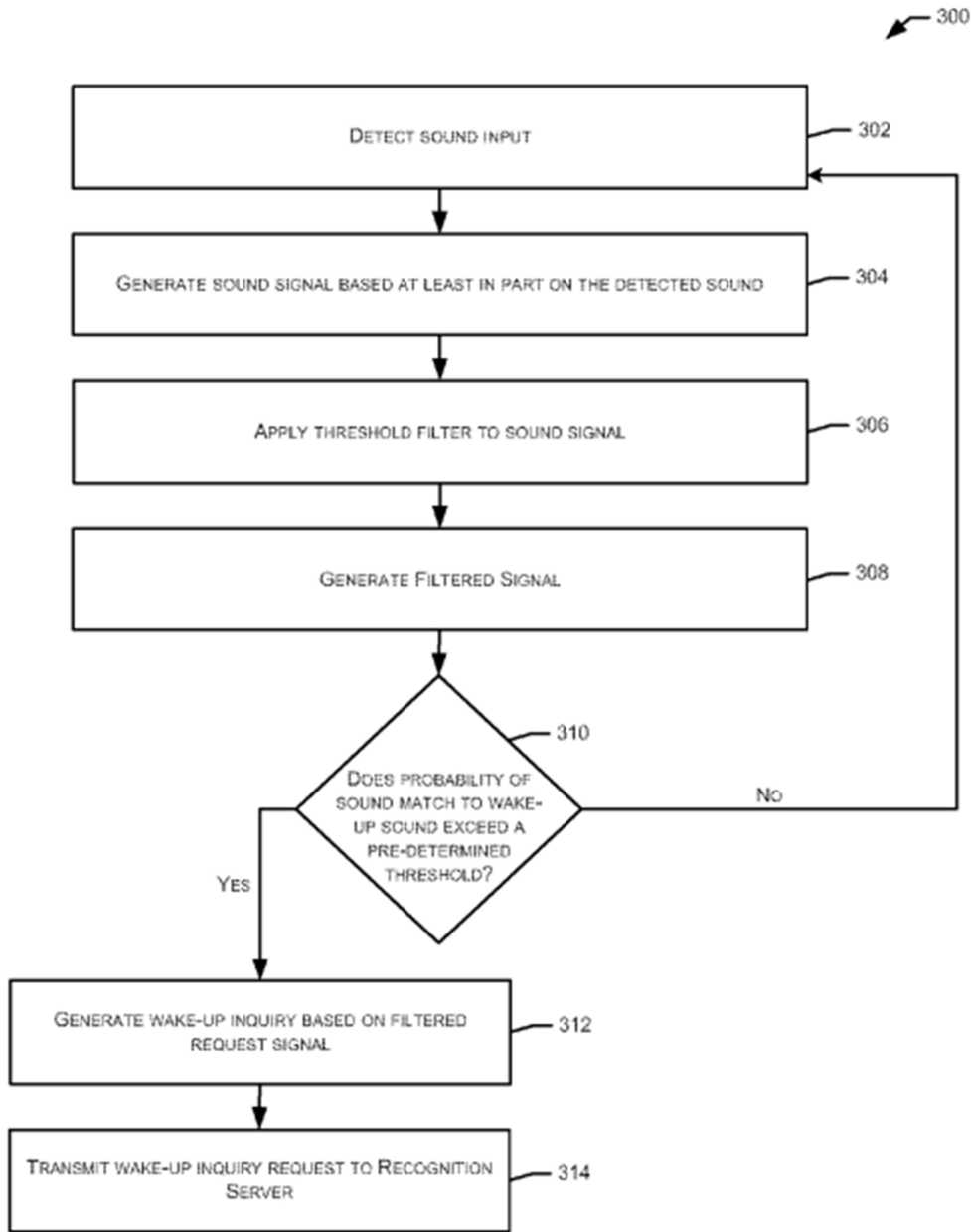


Ex-1006, Fig. 2

100. Shenhav discloses that its mobile device detects and performs initial processing on a sound signal while operating in a standby, or low power, state. *Id.* ¶14. The mobile device determines if the sound signal is indicative of one or more

particular wake-up phrases. *Id.* According to Shenhav, a wake-up phrase may be a short phrase, such as “wake up,” “awake,” or “phone,” that is intended to bring the mobile device out of a “stand by mode” into an “active mode.” *Id.* ¶45. In the standby mode, the mobile device’s main processor(s) are in a “partially [powered] or low power state.” *Id.* ¶33.

101. In Figure 3, Shenhav depicts the process carried out by the mobile device. In sum, 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.*



Ex-1006, Fig. 3

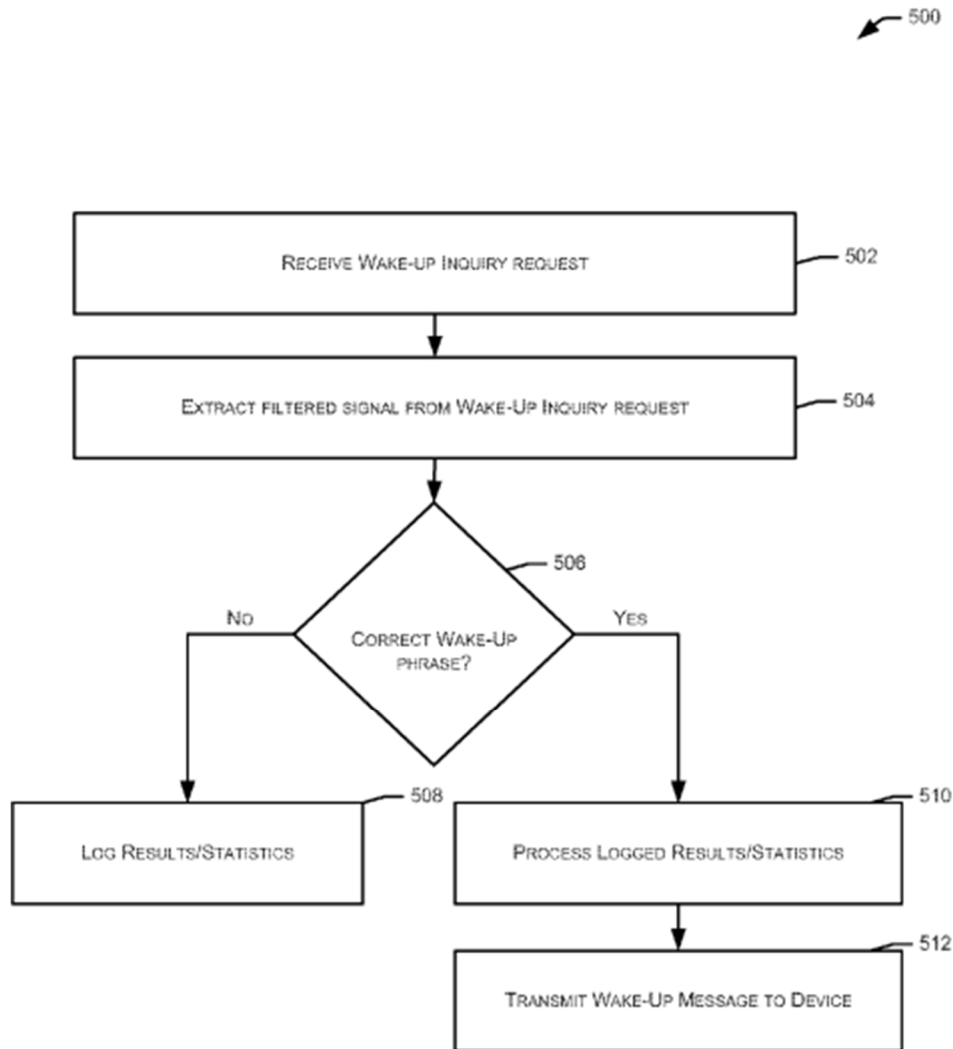
102. As described by Shenhav, “[a]t block 310, a determination may be made as to whether or not the filtered signal passes a threshold. This may be a threshold probability that there is a match of the sound to a wake-up phrase. This

process may be performed by the communications processors 230 and/or the filter/comparator module 224.” Ex-1006 ¶50.

103. Next, “[i]f ...the filtered signal representing the detected sound is found to not exceed a threshold probability of a match to a wake-up phrase, then the method 300 may return to block 302 to detect the next sound input.” *Id.* ¶50. However, “[i]f ... the detected sound is found to exceed a threshold probability of a match to a wake-up phrase, then at block 312, the filtered signal may be encoded into a wake-up inquiry request.” *Id.* Then, “[a]t block 314, the wake-up inquiry request may be transmitted to the recognition server 204.” *Id.*

104. Once the wake-up inquiry request has been sent to the recognition server, the server 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. *Id.* “The wake-up signal may prompt the [mobile] device to wake up from a sleep or stand by state to a powered state.” *Id.*

105. The server-side process is depicted in Figure 5.



Ex-1006, Fig. 5

106. In particular, Shenhav discloses that, after receiving the wake-up inquiry request at block 502,

Recognition server 204, at block 504, may extract the wake-up sound signal from the wake-up inquiry request by processing the contents of the request. In one aspect, the processors 260 may parse the one or more data packets of the wake-up inquiry request and extract the sound signal and/or the filtered sound signal therefrom. In certain embodiments, the recognition server 204 and the processors 260

thereon, may also extract information pertaining to the identification of the wake-up inquiry request for the mobile device 200.

Ex-1006 ¶57.

107. As shown in Figure 5, at block 506, the server determines if the sound signal corresponds to a correct wake-up phrase. *Id.* ¶58. Further, if at block 506, it is determined that the received sound signal does correspond to a wake-up phrase, then the *recognition* server “may proceed to transmit a wake-up signal to the mobile device 200 at block 512. The wake-up signal, as described above, may enable the processors 212 to awake into an on state from a stand by state.” *Id.* ¶60.

C. Parker (Ex-1007)

108. Parker is titled “NAME BASED INITIATION OF SPEECH RECOGNITION.” Ex-1007, Cover (54).

109. Parker discloses a client/server system that emulates a “conversation” with an end user. Ex-1007 ¶22. In particular, Parker’s system implements a method that involves:

[L]istening for audio name information indicative of a name of a computer, with the computer configured to listen for the audio name information in a first power mode that promotes a conservation of power; detecting the audio name information indicative of the name of the computer; after detection of the audio name information, switching to a second power mode that promotes a performance of speech recognition; receiving audio command information; and performing speech recognition on the audio command information.

Id., Abstract

110. A high level conceptual diagram of Parker's system is illustrated in

Figure 1:

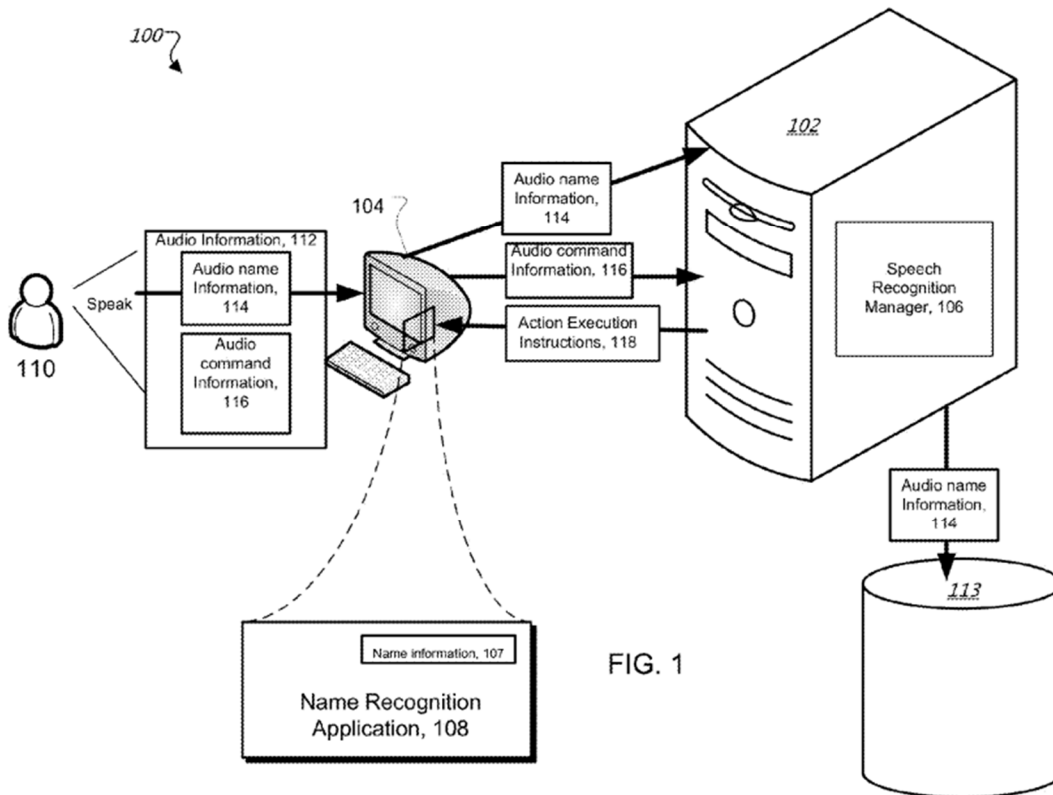
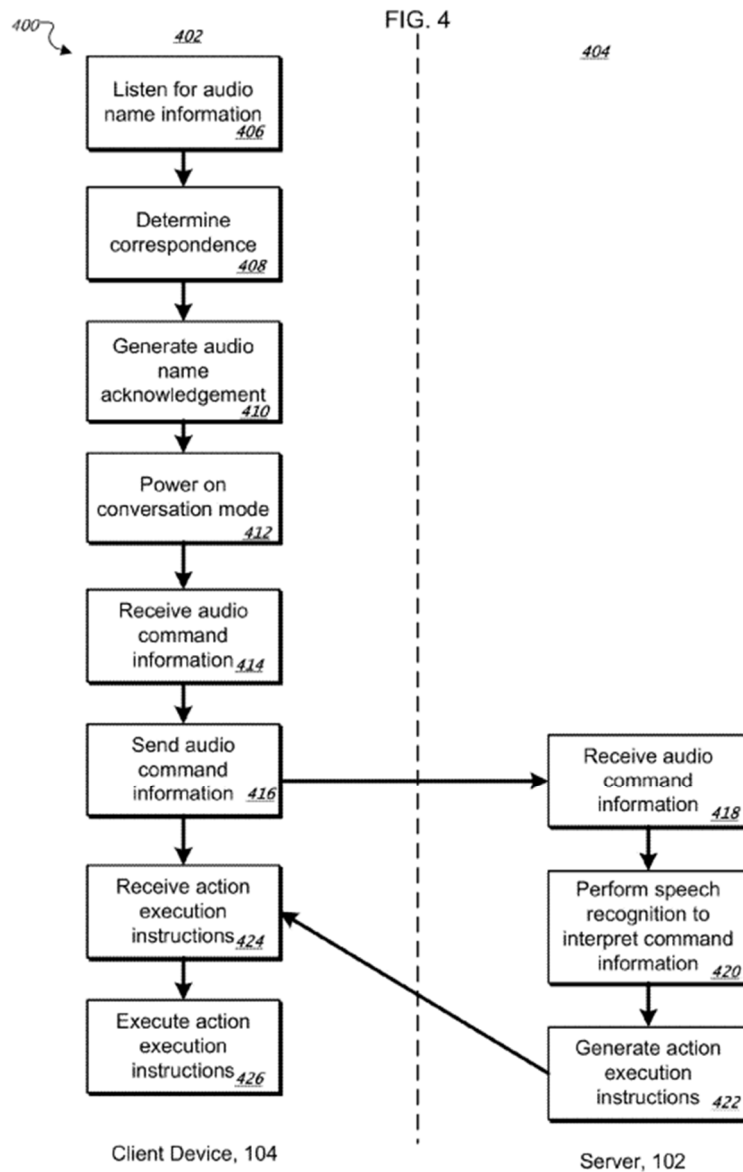


FIG. 1

Ex-1007, Fig. 1

111. As shown, the main components of Parker's system are a client device (designated 104) and a server (designated 102). *Id.* ¶¶34, 38. Client device 104 includes a name recognition application 108 that determines whether received audio matches the client device's name. *Id.* ¶34. If so, client device 104 proceeds to operate in a "conversation mode." *Id.* ¶36. When in conversation mode, client device 104 sends received audio commands to server 102 for processing. *Id.* Server 102

performs speech recognition on the audio commands. *Id.* ¶38. Based on its speech recognition, server 102 sends instructions back to client device 104 for client device 104 to perform. *Id.* Figure 4 of Parker depicts this process.



Ex-1007, Fig. 4

112. Regarding the interaction between client device 104 and server 102, Parker takes note of the time lag may be involved between the time a speaker utters

a voice command and the time the server processes the command. Specifically,

Parker discloses:

Name recognition application 108 powers on (412) conversation mode, for example, to begin performing speech recognition on audio command information 116 that is received by a microphone associated with client device 104. Name recognition application 108 receives (414) audio command information 116, for example, from user 110 of client device 104. ***Name recognition application 108 generates (not shown) an audio acknowledgement to notify user 110 of receipt of audio command information 116 by client device 104.*** Name recognition application 108 sends (416) the audio command information 116 to speech recognition manager 106 on server 102.

In the example of FIG. 4, audio command information 116 is processed by speech recognition manager 106 on server 102. ***Due to the time required for server 102 to process audio command information 116, a time lag may exist from when user 110 speaks audio command information 116 to when user 110 receives action execution instructions 118. In this example, the audio command acknowledgement provides user 110 with an assurance that the audio command information 116 is being processed.***

Id. ¶¶60, 70

113. Thus, Parker discloses the advantageous of immediate acknowledgement of receipt of a voice command to reduce confusion by an end user.

In particular, Parker discloses:

In an example, performance of name recognition locally on client device 104 (as opposed to going over a network to server 102) has multiple advantages, including, e.g., a faster name recognition speed, a reduction in consumption of power in system 100, and so forth. ***In this example, client device 104 acknowledges receipt of the command the user gives (e.g., the command being the second utterance by the user after name acknowledgement by client device 104), for example, immediately after receiving the command and before client device 104 has processed the command either locally or by sending the command***

over a network to a server for processing. By quickly acknowledging receipt of the command, client device 104 is able to process the command, which may take a few seconds, without the user being unsure as to whether client device 104 received the command.

Id. ¶93.

D. Lebeau (Ex-1014)

114. Lebeau is titled “AUTOMATICALLY MONITORING FOR VOICE INPUT BASED ON CONTEXT.” Ex-1014, Cover (54).

115. 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, monitoring for voice input. Ex-1014 ¶23.

E. Conroy (Ex-1017)

116. Conroy is titled “MEDIA PROCESSING METHOD AND DEVICE.” Ex-1017, Cover (54).

117. 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. Conroy’s audio subsystem handles all of the processing of audio data, so that the

CPU and the rest of the host subsystem may be configured to enter a low power state while the audio data is processed, by for example, a digital signal processor. *Id.* ¶14.

VII. CLAIM CONSTRUCTION

118. I have given all the claim terms their plain and ordinary meaning, as would have been understood by a person having ordinary skill in the art at the time of the alleged invention, which I understand is March 12, 2013, having taken into consideration the language of the claims, the specification, and the prosecution history of record.

VIII. DETAILED EXPLANATION OF GROUNDS

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

1. **A POSITA would have been motivated to combine Basye's and Shenhav's teachings and would have had a reasonable expectation of success.**

119. 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.

120. A POSITA also would have been motivated to combine the teachings of Shenhav with Basye because they are analogous art and teach improvements to the same type of devices. 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 from “stand by mode” into an “active mode”).

Each teaches using speech processing to determine a voice command. Ex-1005 ¶¶45, 74, Fig. 3; Ex-1006 ¶¶36-37.

121. Basye discloses a remote speech recognition server that confirms a wakeword was spoken by an end user. Ex-1005 ¶65. In Basye’s case, the computing device transmits to the speech recognition server audio input that includes both a wakeword and a separate voice command. Basye discloses:

With specific reference to FIG. 4A, an illustrative operation by which a wakeword may be confirmed is shown. The user 401 may speak the wakeword 502. ***The user computing device 200 may obtain the audio input that may include the user's speech (1) and determine that the wakeword 402 is present in the speech (2), substantially as discussed above with respect to FIG. 3. The audio input may also include a voice command or query.*** Responsive to determining that the speech includes the wakeword, the application processing module 112 and the network interface module 206 of the user computing device 200 may be activated (3) and the audio input transmitted (4) over the network 410 to the speech recognition server 420. The speech recognition server 420 may confirm (5) that the wakeword is present in the audio input, and may transmit (6) a confirmation to the user computing device 200 over the network 410.

Id. ¶65.

122. Shenhav teaches additional benefits applicable to Basye’s mobile device. 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 discloses maintaining 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.

123. To facilitate the transitioning of the main processor between the standby mode and an active mode (in which the main processor “wakes up”), Shenhav discloses a communications module. In particular, the communication module itself includes a processor, whereby

the processor 230 may perform alternate functions when the processor module 210 is turned off, powered down, in an energy conservation mode, and/or is in a standby mode. For example, when the processor module 210 is in a standby mode, or when it is completely turned off, *the communications module 220 may switch to a set of low power functions, such as functions where the communications module 220 may continually monitor for receipt of communications data, such as a sound indicative of waking up the mobile device 200 along with any components*, such as the processor module that may be in a power conservation mode. The communications module 220, filter/comparator module 224, and the *processor 230 may, therefore, be configured to receive a signal associated with a sound and process the received signal. In one aspect, the communications processors 230 and or the filter/comparator module 224 may be configured to determine 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.*

Id. ¶36.

124. Shenhav also observes that “the accuracy of speech recognition in certain environments, such as noisy environments, may ... be prone to error,” and

that the high level of processing bandwidth needed for speech recognition may not be available, especially on a mobile device in a power conservation state. *Id.* ¶4.

125. Accordingly, Shenhav discloses a technique for waking up the mobile device's main processor, in which the low power communications processor first 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, 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." *Id.* ¶¶36-37. As Shenhav discloses, its communications processor encodes the sound signal into a "wake-up inquiry request," which is then transmitted to the recognition server. *Id.* ¶51.

126. Further, when Shenhav's 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 response to the wake-up inquiry request sent by the mobile device, Shenhav's recognition server "extract[s] the wake-up sound signal from the wake-up inquiry request." *Id.* ¶57. Next, "[i]f ... it is determined that the received sound signal does correspond to a wake-up phrase, then the recognition server ... proceed[s] to transmit a wake-up signal to the mobile device" which "may *enable the processors ... to awake into an on state from a stand by state.*" *Id.* ¶60.

127. Based on Shenhav's teachings, a POSITA would have been motivated to achieve the potential 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-1005 ¶79 ("on-device speech recognition capabilities as implemented by ... the 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.

128. 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 the speech recognition module 110 is transmitted to the speech recognition server to confirm the audio input includes the wakeword without activating application processing module 112, and would have had a reasonable expectation of success in doing so.

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:

129. To the extent the preamble is limiting, Basye teaches² this element.

130. 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.

131. Basye describes the selective activation of these modules as “an *escalation of processing and/or power consumption*, as modules that are activated later ... may have relatively greater processing requirements and/or power consumption.”³ *Id.* ¶44. For 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

² I understand from counsel, that the term “teaches” includes both express teaching or those fairly suggested to a person of ordinary skill in the art.

³ All emphases/annotations/color are added unless noted otherwise.

configured to detect a keyword in the speech, such as a wakeword or sleepword.”

Id. ¶¶19, 21, 24. Furthermore, application processing module 112 may include “a dedicated microprocessor for implementing [a] speech recognition application.” *Id.*

¶30. Basye observes that on-device speech recognition as 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”).

132. Thus, in my opinion, a person of ordinary skill in the art would have understood that Basye’s device is in a “lower power mode” when any or all of audio detection module 106, speech detection module 108, and speech processing module 110 are activated, and application processing module 112 is **not** activated. On the other hand, when the application processing module 112 is also activated, a person of ordinary skill in the art would have understood that Basye’s computing device is in a “higher power mode” due to the application processing module’s relatively large power requirements.

133. Accordingly, in my opinion, a POSITA would have understood that Basye teaches a computing device that 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”).

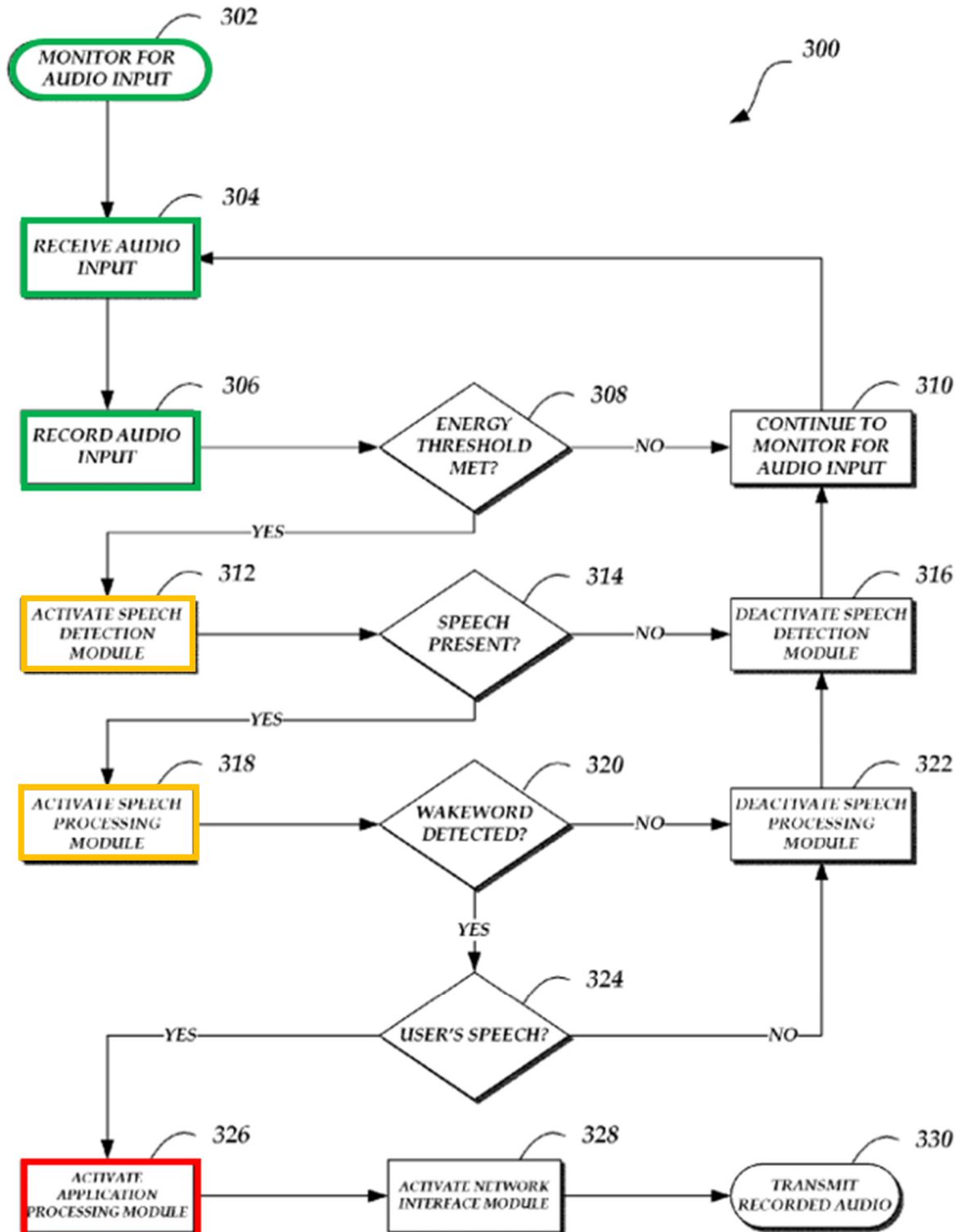
b. 1[a]: at least one processor configured to perform at least one processing stage on an acoustic input received from an environment of the device while the device is operating in the lower power mode, performing the at least one processing stage comprising:

134. Basye teaches this element.

135. 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. For example, Basye’s computing device includes “an *audio input module* 208” that “include[s] an audio input device, such as a microphone or array of microphones.” Ex-1005 ¶¶33, 40, Figs. 1, 2.

136. As shown below in Figure 3, audio input module 208 monitors for audio input, receives audio input, and records the received audio input to a memory buffer module (blocks 302, 304, and 306, highlighted green). Ex-1005 ¶45. Basye teaches using the audio input module to determine “if the user computing device 200 is in a relatively *noisy environment*,” so that the user may wish to increase the energy level threshold at which speech detection module 108 is activated (yellow), in order to decrease the chance of incorrectly detecting speech in the audio input. Ex-1005 ¶74. Basye’s audio input module 208 thus receives audio (acoustic) input (e.g. Figure 3, element 304) from the environment of Basye’s computing device because Basye

recognizes that the sound detected by the input module is from the environment of the mobile device, which may be noisy or quiet. Ex-1005 ¶74.



137. Basye teaches that “speech detection module 108” implements one or more techniques to determine whether the audio input includes speech, such as “voice activity detection (VAD) techniques.” Ex-1005 ¶21. Basye also teaches that “speech processing module 110 includes a *microprocessor* configured to detect a keyword in the speech.” *Id.* ¶24.⁴ Each of the speech detection module 108 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.

138. Further, these processing stages are consistent with the ’750 patent’s disclosure as to what constitutes a processing stage. For example, the ’750 patent discloses:

With respect to evaluating the content of acoustic input containing speech, act 220 may include one or more speech processing stages configured to learn something about the content of the acoustic input. For example, in embodiments that employ one or more *VAD processing stages, a speech processing stage (or multiple speech*

⁴ 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, it is my opinion that a person of ordinary skill in the art would have understood, for example, that speech detection 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.

processing stages) may be performed on acoustic input for which voice content has been detected via the one or more VAD processing stages. The speech processing stage(s) may evaluate the acoustic input to determine the content of speech contained therein to assess whether the user has spoken a voice command.

Ex-1001, 10:55-65.

139. Therefore Basye's computing device teaches a processor that performs at least one processing stage on an acoustic input received from an environment of the device, as claimed.

140. Additionally, Basye teaches that its processor processes an acoustic input while the device is operating in a lower power mode. As shown in Figure 3, "the power management module 120 *may activate the application processing module 112*" at block 326. Ex-1005 ¶49.

141. Moreover, Basye teaches that "speech detection module 108 includes a *low-power digital signal processor* ["DSP"]... configured to implement one or more techniques to determine whether the audio input includes speech," such as "voice activity detection (VAD) techniques." *Id.* Specifically, Basye provides that speech processing module 110 is activated after speech processing module 108:

The speech detection module 108 may communicate its determination as to whether speech is present in the audio input to the power management module 120. *If speech is present in the audio input, the power management module 120 may activate the speech processing module 110 (alternately, the speech detection module 108 may communicate directly with the speech processing module 110).* If speech is not present in the audio input, the power management module 120 may deactivate the speech detection module 108. Alternately, the

speech detection module 108 may communicate the score to the power management module 120, whereupon the power management module 120 may determine whether to activate the speech processing module 110 or deactivate the speech detection module 108.

Id. ¶23. Further, Basye's speech processing module 110 is activated prior to application processing module 112. In particular, Basye discloses:

The speech processing module 110 may communicate its determination as to whether a keyword is present in the speech to the power management module 120. ***If the keyword is present in the speech and the keyword is a wakeword, the power management module 120 may activate application processing module 112 and the network interface module 206 (alternately, the speech processing module 110 may communicate directly with these other modules).*** If the keyword is not present in the audio input (or the keyword is a sleepword), the power management module 120 may deactivate the speech processing module 110 and the speech detection module 108. Alternately, the speech processing module 110 may communicate the score to the power management module 120, whereupon the power management module 120 may determine whether to activate the application processing module 112 and the network interface module 206 or deactivate the speech processing module 110 and the speech detection module 108. In some embodiments, these activations and/or deactivations only occur if the speech processing module 110 determines that a particular user spoke the speech that includes the keyword.

Id. ¶28. Thus, the application processor is not activated (i.e. in a lower power state) in the earlier steps.

142. Therefore, when audio input module 208 monitors for audio input, receives audio input, and records the received audio input to a memory buffer module (blocks 302, 304, and 306, highlighted green) and speech detection module's DSP or speech processing module 110's ***microprocessor*** implements VAD

(highlighted yellow), application processing module 112 has not yet been activated (activation of that module highlighted in **red**, element 326 in Figure 3, above). Ex-1005 ¶45. Thus, as I discussed in §VIII.A.2.a, in my opinion, 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.

143. Accordingly, Basye teaches a computing device with a speech detection module that uses a digital signal processor and a speech processing module that uses a microprocessor (“at least one processor”) that, respectively, detects speech and a wakeword (“configured to perform at least one processing stage”) in audio input received by an audio input module that includes one or more microphones (“on an acoustic input received from an environment of the device”), where the audio input module receives the audio input and the speech processing module detects the wakeword in the audio input when the computing device’s application processing module is not activated (“while the device is operating in the lower power mode”).

- (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:**

144. Basye teaches this element.

145. Basye determines whether the acoustic input includes a specific word or phrase while in lower power mode. As discussed in §VIII.A.2.a, Basye's speech processing module 110 performs its processing while Basye's computing device is operating in a lower power mode. Speech processing module 110 processes "the audio input to determine whether a **keyword** is included in the speech." Ex-1005 ¶24. An example of a keyword that Basye provides is a "wakeword." *Id.*

146. For example, Basye discloses:

[S]peech processing module 110 may process the audio input to determine whether a keyword is included in the speech. In some embodiments, the speech processing module 110 includes a microprocessor configured to detect a **keyword in the speech, such as a wakeword or sleepword**. The speech processing module 110 may be configured to detect the keyword using HMM techniques, GMM techniques, or other speech recognition techniques.

Ex-1005 ¶24.

147. In my opinion, Basye teaches that the specific word indicates to activate a voice interface for subsequent voice input. 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. Basye also discloses that an example of a speech recognition application that

the application processing module implements is the intelligent agent disclosed in Gruber (U.S. Patent Application Publication No. 2012/0016678 (“Gruber,” Ex-1008), which I understand Basye incorporates by reference. *Id.*, ¶¶13, 29. For example, Gruber teaches that its “assistant ... 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 a speech recognition application implemented on application processing module 112, like Gruber’s intelligent agent, is an example of a voice interface. In my opinion, a person of ordinary skill in the art would have understood a “conversational interface” and a “voice interface” to be referring to the same thing in terms of user speech input.

148. My opinion regarding “conversational” and “voice” interfaces is informed at least in part by references that used similar terms prior to the time of the ’750 patent. As one example, U.S. Patent No. 9,634,855 (“Poltorak,” Ex-1011) disclosed:

An ***interface device*** and method of use, comprising ***audio*** and image inputs; a processor for determining topics of interest, and receiving information of interest to the user from a remote resource; an audiovisual output for presenting an anthropomorphic object conveying the received information, having a selectively defined and adaptively alterable mood; an external communication device ***adapted to remotely communicate at least a voice conversation with a human user of the personal interface device.***

Ex-1011, Abstract.

149. In summarizing the invention, Poltorak describes an exemplary voice interaction (i.e., a conversation) between a user and the device, which I set forth in part below:

The user might say: “I hate sitting here—I want to go on vacation somewhere fun!”. The device, recognizing the word “vacation” in the context of an open ended declarative, might respond: “early summer is a great time to go to Florida, before the hurricane season. Hilton Hotels are having a timeshare promotion like the one you went on last year. You can invite grandson Jimmy, who did well in school this year.” The user may respond: “that's a great idea. How much does it cost. And I don't want to sit in an endless timeshare sales pitch!” The device might then respond: “If you sit in the sales pitch, which is 90 minutes, you get \$300 off the hotel rate, plus it keeps you out of the sun midday. Besides, your friend Wendy Montclair owns a timeshare there and wrote good things about it in her blog. You always liked Wendy.” The user might respond: “I don't like her anymore. She's going out with Snidely Whiplash!” The device might then respond, “You're joking. Snidely Whiplash is a cartoon character from Dudley Do-Right. Besides, the timeshare you now own went up in value, and you can sell it at a profit to buy this one.” The user might respond, “I bought the last one to be near Harry. He's a good friend.” ***The conversational interface might respond:*** “I just checked; Harry Lefkowitz passed away last month at age 79. His obituary is in the Times. Would you like me to read it to you?”

As can be seen from this exchange, ***the conversational interface*** seeks to synthesize information, some of which can be gathered in real time based on the context of the conversation, and may optionally have commercial motivation.

Id., 15:36-64.

150. Further, as I discussed in Section VIII.A.2.a, Basye's speech processing module 110 performs its processing while Basye's computing device is operating in a lower power mode (i.e., when application processing module 112 is not activated).

However, once application processing module 112 is not 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.

151. Accordingly, Basye teaches a computing device having a speech processing module that 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 for the computing device 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, after a wakeword has been detected subsequent audio input is provided directly from the device's audio input module to the application processing module ("for provision of subsequent input").

(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;**

152. Basye teaches this element.

153. The '750 Patent explains that automatic speech recognition may be used to recognize at least one word or phrase in the acoustic input as:

Speech processing stages may include, but are not limited to, any one or combination of **automatic speech recognition (ASR)** processes such as limited vocabulary ASR (e.g., **evaluating the acoustic input**

to determine whether a user spoke an explicit voice trigger, evaluating the acoustic input to determine whether the acoustic input includes certain words characteristic of a voice command, etc.), unrestricted and/or large vocabulary ASR, etc.

Ex-1001, 10:66-11:6.

154. The '750 Patent further explains that:

ASR component 930 may be configured to perform limited vocabulary recognition and/or generally unrestricted recognition using one or any combination of suitable acoustic models, vocabularies/lexicons, grammars, language models, etc. to perform recognition, as aspects are not limited for use with any particular ASR implementation.

Ex-1001, 38:48-53.

155. Basye uses models in order to recognize at least one word in the acoustic input, while its device is in lower power mode in the same manner described by the '750 Patent. As discussed in §VIII.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. 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. As such, in my view, a person of ordinary skill in the art would have known that Basye's use of HMM and GMM techniques involves the comparison of audio input to keyword models to determine whether a keyword, or a

particular word or phrase, was detected in the same manner described by the '750 Patent.

156. A person of ordinary skill in the art would have understood HMM (Hidden Markov Model) and GMM (Gaussian Mixture Model) as known techniques for automatic speech recognition. For example, in describing an invention “related to machine recognition of human speech,” Espy-Wilson discloses in its background description of prior art:

Numerous *automatic speech recognition (ASR) procedures are known in the prior art*, some of which, including embodiments of the present invention, utilize knowledge of acoustic-phonetics. These procedures can be classified into three broad categories: (1) the acoustic-phonetic approach to recognition, (2) the use of acoustic correlates of phonetic features in the front-ends of dynamic *statistical ASR methods like Hidden Markov Models (HMMs)*, and (3) the use of phonetic features in place of phones as recognition units in the dynamic statistical approaches to ASR that use standard front-ends like Mel-Frequency Cepstral Coefficients (MFCCs).

Ex-1012, 1:21-22, 1:32-43.

157. Espy-Wilson further discloses that “[s]ome *prior art ASR systems* have utilized acoustic cues that are correlates of phonetic features to form the front-end in *HMM based ASR methods*.” *Id.*, 1:51-53.

158. Similarly, in describing an invention directed to “speech recognition systems and methods using an adaptive incremental learning approach,” U.S. Patent No. 10,573,304 (“Gemmeke,” Ex-1013) discloses, in its background section, that “*Automatic Speech Recognition (ASR) systems* are typically set up with *Hidden*

Markov Models (HMM), developed with continuous *Gaussian mixture (GMM)* emission densities and context-dependent phones.” Ex-1013, 1:19-22; *see also id.*, 1:47-48 (“Automatic Speech Recognition (ASR) systems are typically set up with Hidden Markov Models (HMM).”).

159. Further, as I discussed in Section VIII.A.2.a, Basye’s speech processing module 110 performs its processing while Basye’s computing device is operating in a lower power mode (i.e., when application processing module 112 is not activated).

160. Accordingly, Basye teaches a speech processing module that employs HMM, GMM or other speech recognition techniques (“performing automatic speech recognition (ASR)”) while the computing device’s 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 in the acoustic input”).

(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;

161. Basye teaches this element.

162. Basye determines whether the recognized word matches the specific word or phrase, while in lower power mode. As discussed in §VIII.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.

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. Basye also discusses speech detection module 108 applying HMM or GMM. *Id.* ¶21 (“the speech detection module 108 applies Hidden Markov Model (HMM) or Gaussian Mixture Model (GMM) techniques to compare the audio input to one or more acoustic models, which acoustic models may include models corresponding to speech, noise (such as environmental noise or background noise), or silence.”).

163. As such, in my view, a person of ordinary skill in the art would have known that Basye’s use of HMM and GMM techniques involves the comparison of audio input to keyword models to determine whether a keyword, or a particular word or phrase, was detected. For example, references available before the ’750 patent employ HMM techniques, such as U.S. Patent No. 6,594,630 (“Zlokarnik,” Ex-1016). Zlokarnik discloses an “apparatus for voice-activated control of an electrical device comprises a receiving arrangement for receiving audio data generated by user [sic].” Ex-1016, Abstract. More particularly, in describing the background to its invention, Zlokarnik discloses certain advantages of HMM:

Another advantage which distinguishes the HMM over DTW and ANN is that it does not require a defined starting or end point in order to

recognize a word. The *HMM uses qualitative means of comparing the features in an input stream to the stored templates* eliminating the need to distinguish the start and end points. *It uses a statistical method to match the sound that is being detected with any sound that is contained in it's [sic] templates* and then outputs a score which is used to determine a match.

Id., 3:17-25.

164. Thus, in my view, at a minimum, a person of ordinary skill in the art would have found it obvious based on Basye's description of its speech recognition server to implement this functionality (i.e., comparing the audio input to one or more keyword models stored in the non-transitory computer-readable medium drive 204 to determine if the detected keyword in the audio input signal matches a keyword model) in speech processing module 110.

165. Moreover, Basye teaches that its 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 teaches that its 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.

166. Thus, because both speech processing module 110 and speech recognition server 420 make a determination that a wakeword is present in an audio

input, and speech recognition server 420 accesses a data store that stores known wakeword models and compares audio input to those stored 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.

167. Indeed, in my view, implementing speech processing module 110 in this way is consistent with Basye's use of HMM or GMM techniques. More particularly, Basye's speech detection module 108, which also applies HMM or GMM techniques for the purpose of determining whether audio input includes speech, "compare[s] the audio input to one or more *acoustic models*, which acoustic models may include models corresponding to speech, noise (such as environmental noise or background noise), or silence." Ex-1005 ¶21. Further, these acoustic models are stored on non-transitory computer-readable medium drive 204, along with the wakeword models. *Id.* ¶36.

168. Accordingly, Basye teaches a speech processing module implemented on a computing device that 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").

- (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;**

169. Basye-Shenhav teaches this element.

170. Basye teaches requesting, 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 or phrase. Specifically, “[u]pon receiving an audio input determined by the user computing device 200 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.” Ex-1005 ¶60. Thus, after the computing device determines that the acoustic input includes the specific word, Basye’s computing device transmits the “audio input” that includes the wake word 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.

171. Basye 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. Thus, 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 and accesses a data store that includes models to identify a wakeword and a separate voice command.

172. A POSITA would have recognized that models can be used to identify a wakeword and a separate voice command. For example, Gruber also discloses the use of stored "models" in performing generalized voice command recognition.

Gruber states:

addition, in one embodiment, the statistical language models are tuned to look for words, names, and phrases that occur in the various models of assistant 1002 shown in FIG. 8. For example, in at least one embodiment the statistical language models are given words, names, and phrases from some or all of: domain models 1056 (e.g., words and phrases relating to restaurant and meal events), task flow models 1086 (e.g., words and phrases relating to planning an event), dialog flow models 1087 (e.g., words and phrases related to the constraints that are needed to gather the inputs for a restaurant reservation), domain entity databases 1072 (e.g., names of restaurants), vocabulary databases 1058 (e.g., names of cuisines), service models 1088 (e.g., names of service providers such as OpenTable), and/or any words, names, or phrases associated with any node of active ontology 1050.

Ex-1008 ¶312. Thus, in my opinion, because Basye's speech recognition server is configured to receive audio input that contains both a wakeword and a separate voice command, and because the speech recognition server accesses a data store that includes wakeword models, acoustic models, and language models, a person of ordinary skill in the art would have understood that the speech recognition server both confirms that a wakeword is included in the audio input (using the stored

wakeword models) and recognizes the separate voice command in the same audio input (using the stored language models).

173. I note here that Basye provides exemplary operations of its disclosed speech recognition system in Figures 4A-4C. Ex-1005 ¶55. Figure 4A depicts an example of the speech recognition server recognizing a wakeword.

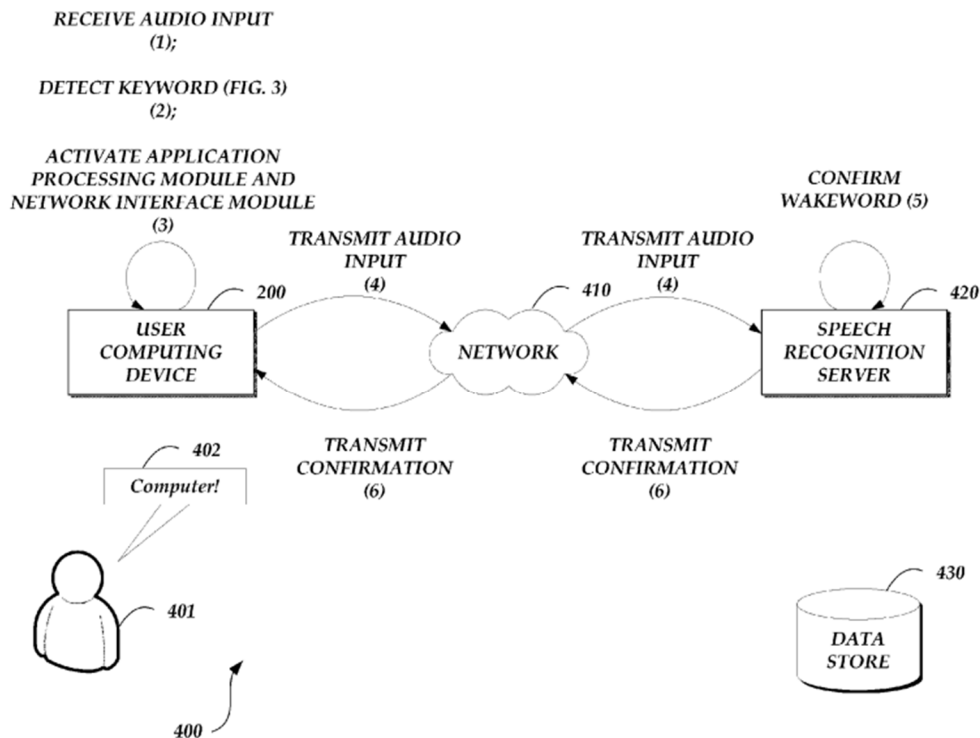


Fig.4A.

Ex-1005, Fig. 4A

174. Figure 4B depicts an example of the speech recognition server recognizing a voice command/agent query. Ex-1005 ¶66.

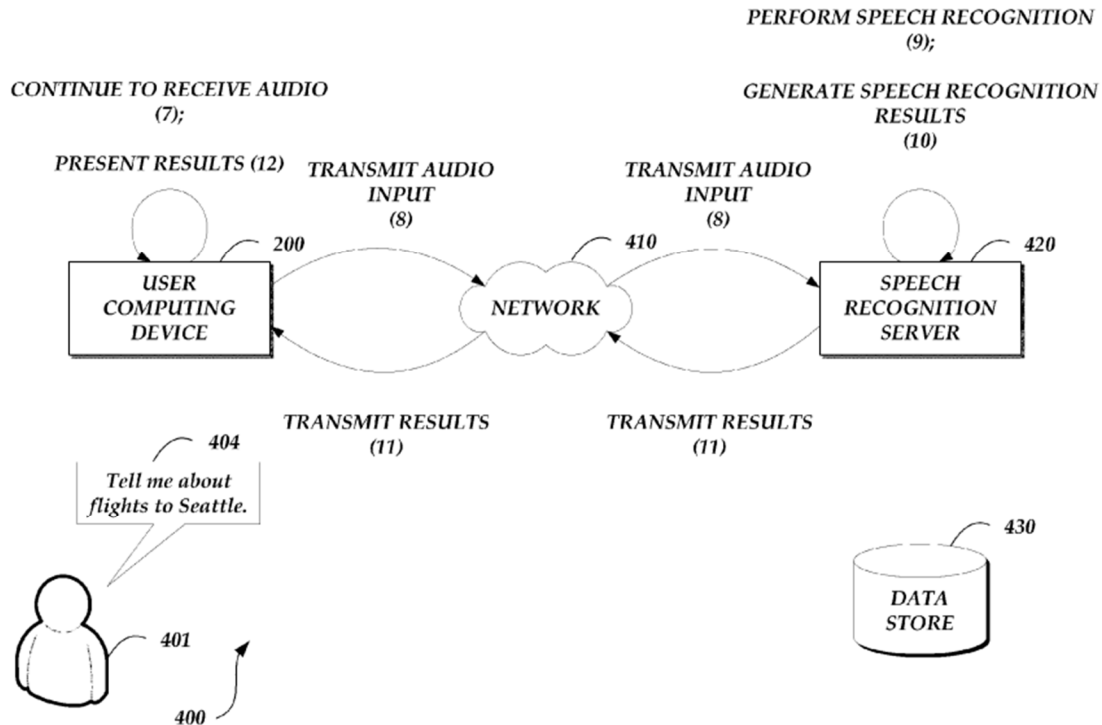


Fig.4B.

Ex-1005, Fig. 4B

175. While these scenarios in Figures 4A and 4B are provided as separate examples, in my opinion, this would not have changed the understanding of a person of ordinary skill in the art 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 the wakeword and the command. I reiterate here that Basye discloses that its speech recognition server receives audio input that may comprise both a wakeword and a separate voice command:

The user computing device 200 may obtain the audio input that may include the user's speech (1) and determine that the wakeword 402 is present in the speech (2), substantially as discussed above with respect to FIG. 3. The audio input may also include a voice command or query.

Responsive to determining that the speech includes the wakeword, the application processing module 112 and the network interface module 206 of the user computing device 200 may be activated (3) and the audio input transmitted (4) over the network 410 to the speech recognition server 420.

Ex-1005 ¶65.

176. Basye further confirms my opinion. More particularly, Basye describes wakeword confirmation and recognition of voice commands (e.g., a spoken query) as being performed on the same audio input. For example, Basye discloses:

In some embodiments, ***the speech recognition server 420 processes the audio input*** received from the user computing device 200 to confirm that a wakeword is present (e.g., by comparing the audio input to a known model of the wakeword), and transmits the confirmation to the user computing device 200. The speech recognition server 420 may further be configured to identify a user 401 that spoke the wakeword using known speaker identification techniques.

Ex-1005 ¶57.

177. Basye also notes that the same audio input can be processed to recognize a voice command:

The ***speech recognition server 420*** may ***process the audio input*** received from the user computing device 200 to determine speech recognition results ***from the audio input***. For example, ***the audio input may include a spoken query*** for an intelligent agent to process; speech to be transcribed to text; or other speech suitable for a speech recognition application.

Id. ¶58.

178. To the extent 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. 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. For example, Sarin notes this consideration: “Decisions on where to perform speech recognition tasks can be based on factors such as grammar size, nature of functions to be performed, resources available at local devices and remote servers (e.g., processing power, memory or storage capacity), and the quality of available communication links (e.g., latency in communication between a local device and a remote server). Ex-1019, 13:8-14. 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.

179. A POSITA would have found it obvious to have Basye’s computing device request that the server perform the voice recognition while Basye’s computing device remains in lower power mode. While Basye teaches that the application processing module can be activated to transmit the audio input, Basye also teaches that 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.

180. Moreover, while Basye teaches that the application processor module is deactivated when not used for command recognition, Shenhav explicitly teaches requesting that a server perform recognition while in a lower power mode. 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.

181. In addition to the reasons I explained in Section VIII.A.1, a POSITA would have been motivated to implement 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 main processor of a mobile device “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 where possible to do so.

182. Accordingly, the Basye-Shenhav combination thus teaches a computing device that, 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 or query (“to determine whether the acoustic input includes at least one voice command”).

- (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**

183. Basye-Shenhav teaches this element.

184. Basye teaches selecting to transition to a higher power mode in response to receiving an indication of a voice command. As discussed in the previous section, Basye transmits the acoustic input of the wake word and the voice command to the speech recognition server while in lower power mode (i.e., the application processing module is not activated). The “speech recognition server 420 may recognize speech (9) included in the audio input and generate speech recognition results (10) therefrom.” Ex-1005 ¶66, Fig. 4B. The speech recognition results are transmitted to the 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.*

185. 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 (e.g., via a display of the user computing device 200),” the computing device would need to select to proceed with activating application processing module 112, which, as a result, would select to transition the computing device from the lower power mode into a higher power mode. Ex-1005 ¶66.

186. For example, U.S. Patent No. 9,031,847 (“Sarin,” Ex-1019), explains that powering up a display of a mobile device typically requires a transition of the power state of the device, from a lower to a higher power state:

Exemplary mobile devices described herein can receive input, perform operations on data, and provide functionality to users while in several different states (e.g., locked or unlocked, active or low-power) or combinations of states. For example, in an unlocked state, a mobile device can provide full access to the functionality of the device. Typically, *a mobile device in an unlocked state is also in an active state, with the display powered on and other input/output devices available for use.* As another example, in a locked state, a mobile device can restrict access to functionality of the device while exhibiting behaviors such as displaying a lock screen (e.g., a screen that prompts a user for authentication information, such as a PIN or biometric data, to unlock the device), or displaying a blank screen or wallpaper. Depending on implementation, a mobile device in a locked state may still provide some functionality (e.g., image capture functionality) that can be accessed by users without unlocking the device. A locked mobile device may be in a locked/active state (e.g., with the display powered on and showing an authentication screen) or a *locked/low-power (or “standby”) state in which certain features of the mobile device are powered down (e.g., to reduce power consumption relative to an active state).* For example, in a low-power state, the display is typically dimmed or powered off, and other components (e.g., hardware buttons, keypads, trackballs) can be deactivated, as well.

Ex-1019, 9:41-65.

187. Moreover, this implementation detail would have further been obvious in view of Shenhav. Shenhav teaches to first determine “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 communications module transmits the sound 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. *Id.*

¶15. “The wake-up signal may prompt the [mobile] device to wake up from a sleep or stand by state to a powered state.” *Id.* 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.

188. Accordingly, the Basye-Shenhav combination thus teaches a computing device that, in response to receiving 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”) activates an application processing module 112 (“selects whether to remain in the lower power mode or to transition to a higher power mode”).

(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.

189. Basye-Shenhav teaches this element.

190. Basye-Shenhav responds to the voice command in either power mode based on the selection. As discussed in §VIII.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.

191. Furthermore, once activated, “application processing module 112 may cause presentation of the results (12) in audible format (e.g., via text-to-speech) or in visual format (e.g., via a display of the user computing device 200).” Ex-1005 ¶66. In my opinion, a POSITA would have understood that displaying speech recognition results on the display of the computing device is a response to a voice command, such as an intelligent agent query. Indeed, the voice command in the example of Figure 4B of Basye is “Tell me about flights to Seattle.” As such, the display of the speech recognition results (i.e., the actual flights to Seattle responsive to the query) qualifies as a response to the query (i.e., the voice command).

192. Accordingly, in my opinion, a POSITA would have understood Basye-Shenhav teaches a computing device that 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.”).

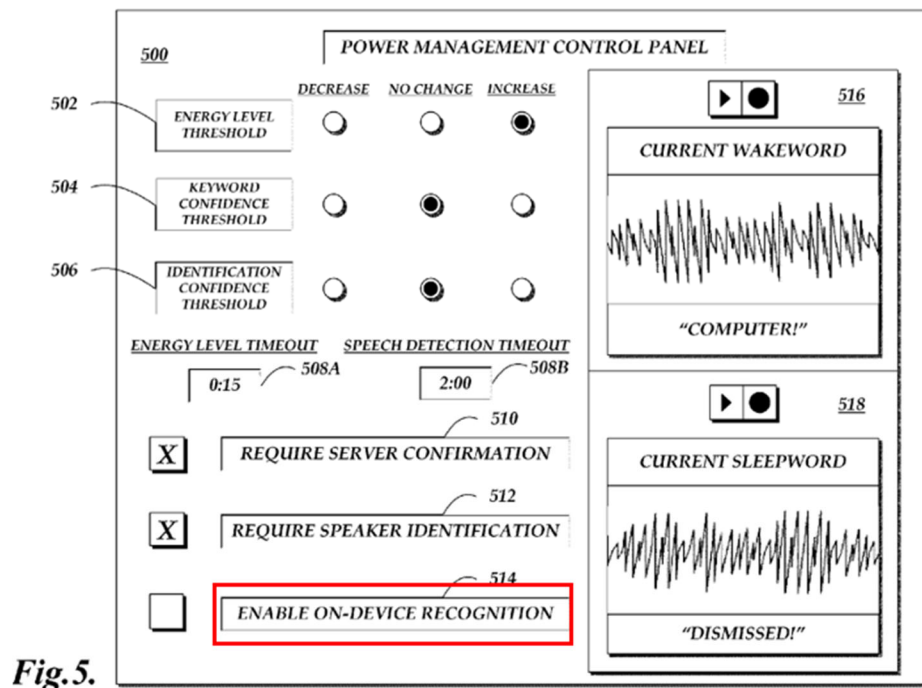
3. Dependent Claims 2-8 and 11-13

- 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.**

193. Basye-Shenhav teaches this claim. Basye teaches a first processor that performs a first processing stage. As I discussed in §VIII.A.2.b, Basye discloses a speech detection module 108, which includes a low-power digital signal processor configured to apply voice activity detection (VAD) techniques, a “processing stage” consistent with the ’750 patent. Additionally, as I explained in Section VIII.A.2.b(i), Basye’s speech processing module 110 processes “the audio input to determine whether a *keyword* is included in the speech.” Ex-1005 ¶24. This is completed before the second processing stage because 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. Moreover, as I explained above, in view of Basye’s flexible architecture (Ex-1005 ¶32), a POSITA would have understood, for example, that speech detection 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. Therefore Basye’s

low-power digital signal processor is the “first processor” that can perform VAD and keyword detection.

194. Basye further teaches a second processor configured to perform a second processing stage. Although, as discussed in §VIII.A.2.b(ii), Basye-Shenhav’s computing device relies on a remote server to perform speech recognition, Basye further teaches that its 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.



Ex-1005, Fig. 5

195. As shown in Figure 5 above, Basye provides a user interface 500 which performs on-device speech recognition (red):

The user interface 500 may also include an on-device recognition selection element 514, wherein the user 401 may select whether the user computing device 200 generates speech recognition results by itself, or whether audio inputs are routed to the speech recognition server 420 for processing into speech recognition results. The on-device recognition selection element 514 may be optionally disabled or grayed out if the user computing device 200 does not include on-device speech recognition capabilities. Further, the on-device recognition selection element 514 may be automatically deselected (and on-device speech recognition capabilities automatically disabled) if the power supply 218 drops below a threshold power supply level (e.g., a battery charge percentage), as on-device speech recognition capabilities as implemented by the processing unit 202 and/or the application processing module 112 may require a relatively large power draw.

Ex-1005 ¶79.

196. Relatedly, to perform on-device speech recognition, Basye discloses that “application processing module 112 includes a dedicated microprocessor for implementing the speech recognition application.” *Id.* ¶30. As I discussed above, an example of a speech recognition application that may be implemented by application processing module 112 is *Gruber’s* intelligent agent. *See id.* ¶13.

197. *Gruber* (which Basye incorporates by reference) teaches that its intelligent agent can perform natural language processing on voice input. Ex-1008 ¶410. *Gruber* discloses that its intelligent agent can perform natural language processing on voice input:

Referring now also to Figure 29, there is shown a screen shot

illustrating *natural language processing according to one embodiment. The user has entered (via voice or text) language input 2902 consisting of the phrase “who is playing this weekend at the filmore”. This phrase is echoed back to the user on screen 2901. Language interpreter component(s) 1070 component process input 2902 and generates a parse result.* The parse result associates that input with a request to show the local events that are scheduled for any of the upcoming weekend days at any event venue whose name matches “filmore”. A paraphrase of the parse results is shown as 2903 on screen 2901.

Ex-1008 ¶410.

198. Thus, Basye’s user computing device is configured to perform at least one second processing stage consistent with the ’750 patent, which also performs natural language processing on voice input. To note, the ’750 patent discloses that natural language processing is a type of processing stage:

Speech processing stages that may be utilized to evaluate the content of input also include one or more natural language processing stages to ascertain the semantic meaning of speech recognized using one or more ASR processing stages. NLP stages may be used either to evaluate whether speech contained in acoustic input corresponds to a voice command, or to determine the meaning of the voice command and/or intent of the user so that the voice command can be carried out.

Ex-1001, 20:3-11.

199. Moreover, a POSITA would have recognized that, that the application processing module 112 is “configured” to perform the at least one processing stage even when the Basye-Shenhav computing device routes audio input to a speech recognition server. That is, in my view, a person of ordinary skill in the art 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, application processing module 112, which still includes the bypassed speech recognition application, is nonetheless “configured” to perform speech recognition (i.e., at least one processing stage) in other instances.

200. Accordingly, in my opinion, the Basye-Shenhav combination thus teaches a computing device that includes a low-power digital signal processor that performs voice activity detection (“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 natural language processing on voice input (“a second processor configured to perform at least one second processing stage”).

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.

201. Basye-Shenhav teaches Claim 3.

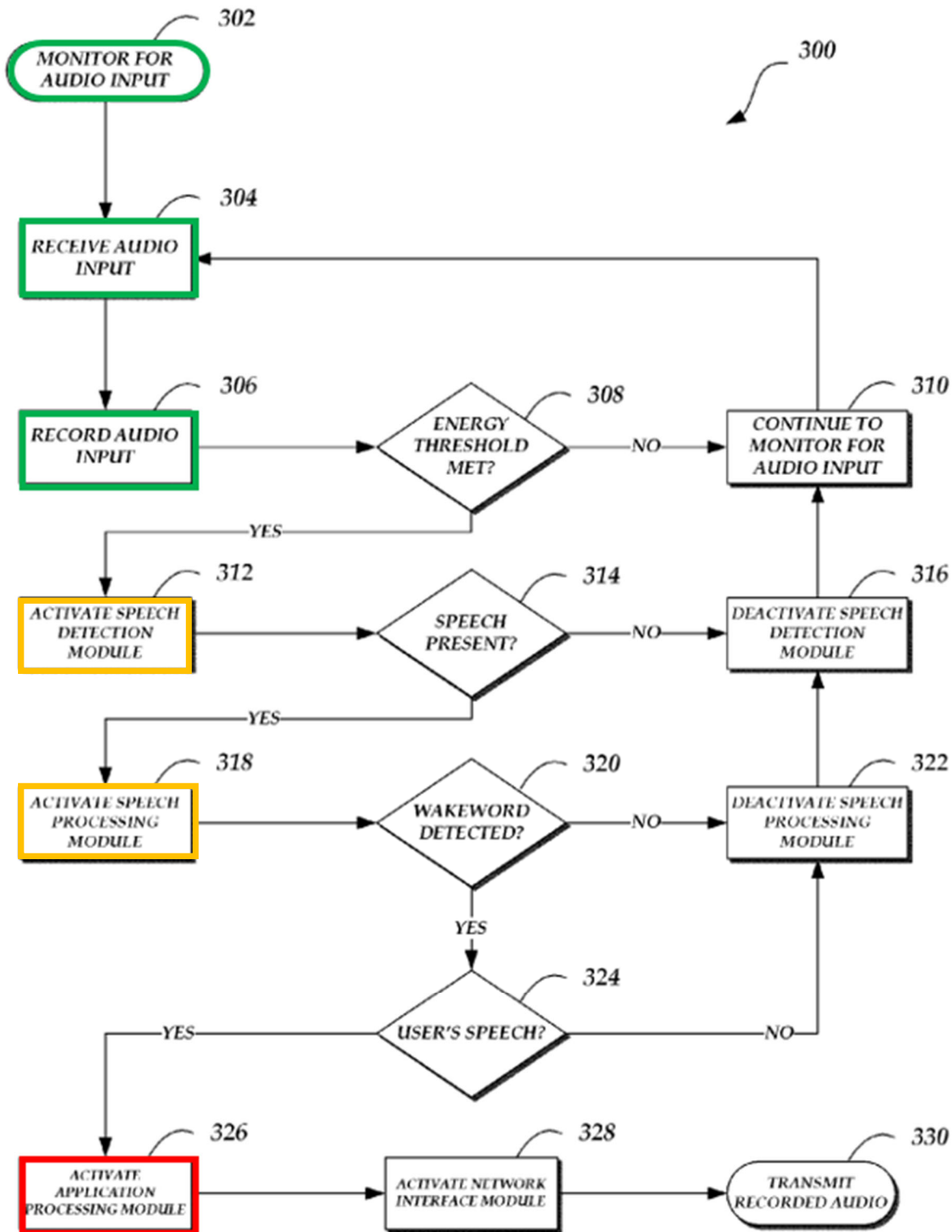
202. Basye’s first processor performs a first processing stage prior to the second processor being engaged for a second processing stage. As discussed in §VIII.A.2.b and the previous section, Basye-Shenhav’s device includes a low-power digital signal processor⁵ that performs a first processing stage of voice activity

⁵ As I explained above, in view of Basye’s flexible architecture (Ex-1005 ¶32), a POSITA would have understood, for example, that speech detection module 108 and speech processing module 110 could be implemented in separate processors (as

detection (Ex-1005 ¶49) and “determine[s] whether a **keyword** is included in the speech” (Ex-1005 ¶24), and an application processing module 112 that includes a microprocessor configured to perform a second processing stage of natural language processing.

203. 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).

disclosed in the exemplary embodiment), or, alternatively, within the same processor, such as a microprocessor or a special-purpose digital signal processor.



204. Accordingly, it is my opinion, the Basye-Shenhav combination teaches that its low-power digital signal processor that determines whether audio input

includes speech (“the first processor is configured to perform at least one first processing stage”) before the application processing module’s microprocessor is activated to perform natural language processing (“prior to the second processor being engaged to perform at least one second processing stage.”).

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.

205. Basye-Shenhav teaches this claim. As discussed in §VIII.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.

206. Accordingly, it is my opinion that a POSITA would have understood that the Basye-Shenhav combination teaches a speech detection module that includes a low-power digital signal processor (“the first processor is a digital signal processor”) that consumes less power than a microprocessor of the application processing module that requires a relatively large power draw (“configured to operate at a lower power than the second processor”).

- 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.**

207. Basye-Shenhav teaches Claim 5.

208. As discussed in §VIII.A.3, 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. With respect to its VAD processing, Basye teaches:

The speech detection module 108 may process audio input to determine whether the audio input includes speech. In some embodiments, the speech detection module 108 includes a low-power digital signal processor (or other type of processor) configured to implement one or more techniques to determine whether the audio input includes speech. In some embodiments, *the speech detection module 108 applies voice activity detection (VAD) techniques. Such techniques may determine whether speech is present in an audio input based on various quantitative aspects of the audio input, such as the spectral slope between one or more frames of the audio input; the energy levels of the audio input in one or more spectral bands; the signal-to-noise ratios of the audio input in one or more spectral bands; or other quantitative aspects.* In other embodiments, the speech detection module 108 implements a limited classifier configured to distinguish speech from background noise. The classifier may be implemented by techniques such as linear classifiers, support vector machines, and decision trees. In still other embodiments, the speech detection module 108 applies Hidden Markov Model (HMM) or Gaussian Mixture Model (GMM) techniques to compare the audio input to one or more acoustic models, which acoustic models may include models corresponding to

speech, noise (such as environmental noise or background noise), or silence. Still other techniques may be used to determine whether speech is present in the audio input.

Ex-1005 ¶21.

209. Therefore Basye specifically 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.* As I discussed in Section V.C, a POSITA would have understood that voice activity detection techniques include using spectral analysis to determine that there is speech included an audio input.

210. Accordingly, it is my opinion, that the Basye-Shenhav combination teaches that its speech detection module that includes a low-power digital signal processor which applies voice detection activity techniques to audio input (“the 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”).

- 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.**

211. Basye-Shenhav teaches Claim 6.

212. As I noted above in Section VIII.A.2.b(1)(a), the '750 Patent describes its implementation of automatic speech recognition as being used to recognize at least one word or phrase in the acoustic input as:

Speech processing stages may include, but are not limited to, any one or combination of automatic speech recognition (ASR) processes such as limited vocabulary ASR (e.g., **evaluating the acoustic input to determine whether a user spoke an explicit voice trigger, evaluating the acoustic input to determine whether the acoustic input includes certain words characteristic of a voice command**, etc.), unrestricted and/or large vocabulary ASR, etc.

Ex-1001, 10:66-11:6.

213. For example, the '750 Patent discloses:

Limited vocabulary ASR may be used to perform explicit voice trigger detection. For example, an exemplary speech processing stage may include 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-1001, 18:20-24.

214. Basye 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 §VIII.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 110 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 as described by the ’750 Patent. 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 described by the ’750 Patent.

215. 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 is based on a smaller, limited vocabulary by contrasting wakeword recognition with more general speech recognition, just like the ’750 patent. For example Salvador discloses:

A wake word recognition module 109 and a speech recognition module 110 may employ any number of conventional speech recognition techniques, such as use of hidden Markov models and Gaussian mixture models, to interpret voice input. For example, ***the 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***. In some implementations, the wake word recognition module 109 may employ a hidden Markov model that represents the wake word itself.

Ex-1009, 4:36-44.

216. A person of ordinary skill in the art would have recognized that Salvador's (and hence Basye's) wakeword detection is "limited vocabulary" ASR in the same sense that the '750 patent describes it. 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.*)").

217. Basye's keyword detection (and Salvador's wakeword detection) could be performed by Basye's first processor. As discussed in §VIII.A.3.a, Basye's first processor is speech detection module 108's *low-power digital signal processor*. Ex-1005 ¶21. Basye's power management subsystem has a flexible architecture, and leaves it as a design choice as to how its modules are implemented. Specifically, Basye teaches

The *modules of the power management subsystem 100 may be combined or rearranged* without departing from the scope of the present disclosure. The *functionality of any module* described above *may be* allocated among multiple modules, or *combined with a different module*. As discussed above, 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* or other dedicated microprocessing hardware. One or more modules may also be embodied in software implemented by a processing unit 202 included in a computing device, as discussed further below with respect to FIG. 2. Further, one or more of the modules may be omitted from the power management subsystem 100 entirely.

Ex-1005 ¶32. Therefore, based on Basye's disclosures, it is my opinion that a POSITA would have appreciated that the functionality of speech processing module

110 could be combined with the functionality of speech detection module 108 into a single module, and that single module can be implemented on one digital signal processor, because doing so was well known. For example, U.S. Patent Application Publication No. 2013/0110521 (“Hwang,” Ex-1010), shows that the digital signal processor could be a low power processor. Hwang discloses that its digital signal processor (DSP) can be included in a wireless communication device, which is shown in Figure 10, reproduced below.

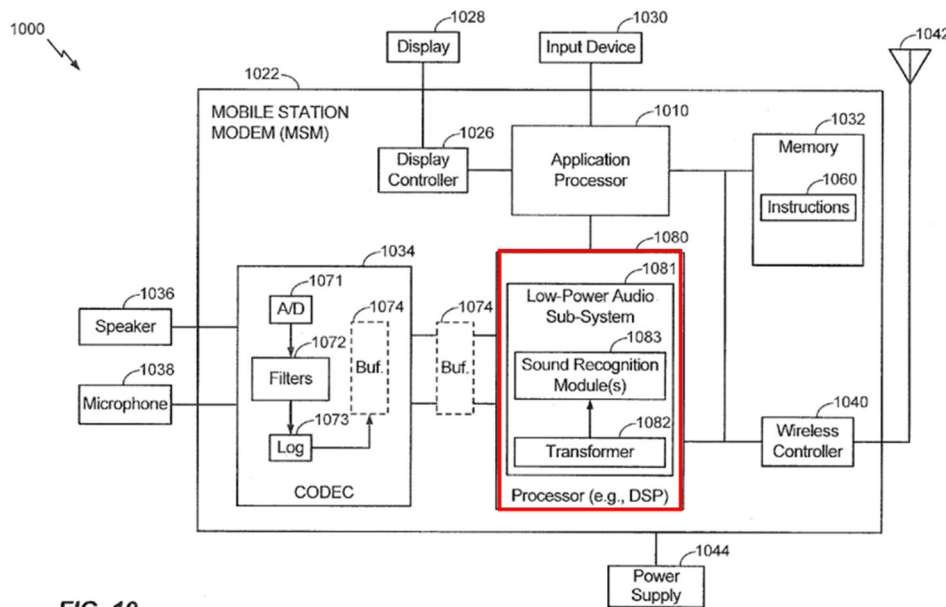


FIG. 10

Ex-1010, Fig. 10 (annotated).

218. As annotated in the figure, DSP 1080 includes what Hwang refers to as a “low-power audio sub-system (LPASS) 1081.” *Id.* ¶86. Within Hwang’s LPASS 1081 are sound recognition modules 1083, which include both target sound detection module 244 and continuous keyword detection module 243. *See id.* (“**The DSP 1080**

may include a transformer 1082 (e.g., the transformer 233 of Figure 2) and *one or more sound recognition modules 1083 (e.g., the sound recognition modules 241-245 of FIG. 2)* configured to perform one or more sound recognition operations.”). Thus, in my view, based at least on Hwang’s disclosure, it would have been apparent to a POSITA that both the detection of speech and the detection of a keyword could be performed on a single *low power digital signal processor*.

219. Moreover, Hwang notes that its digital signal processor carries out target sound detection and continuous keyword detection while the main application processor of the mobile device is not active. *See id.* ¶44 (“the [digital signal processor] may analyze the transformed audio feature data ... via one or more of the sound recognition modules 241-245, and may *determine whether to activate an application processor*.”). In my view, Hwang therefore confirms that a person of ordinary skill in the art was capable of implementing voice activity detection and keyword detection within a single digital signal processor while a mobile device is in a lower power state.

220. Moreover, 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. For example, this benefit is explained in other relevant references, such as U.S. Patent No. 9,998,977 (“Joshi,” Ex-1015). In particular, Joshi discloses

“[m]ethods, systems, and devices ... for wireless communication.” Ex-1015,

Abstract. Joshi observed that

[i]n addition to the host processor, the STA [e.g., a mobile device] may use a radio processor to facilitate roaming. But ***using two processors while roaming may consume excess power***. Accordingly, one of the processors (e.g., the host processor) may enter a sleep state during roaming and the other processor (e.g., the radio processor) may remain awake to facilitate the roaming. A processor (e.g., the host processor) that is in the sleep state may use low power and have limited functionality compared to the awake state. A STA or a processor of a STA may also have multiple different sleep states, e.g., corresponding to different limited functionality or levels of functionality. Thus, ***the STA may conserve power by using a single processor, fewer processors, or a limited portion of one or more processors during roaming and allowing the other processor to be in a sleep state***.

Id., 7:20-35.

221. Thus, in my opinion, a POSITA would have recognized that implementing both speech detection and speech processing functionality on a single low-power digital signal processor would lower power consumption, which furthers Basye’s goal of conserving device power.

222. Accordingly, in my opinion, the Basye-Shenhav combination teaches implementing 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* §VIII.A.2.b(i) (Element 1[a(i)])) using a vocabulary that is particular to a wakeword (“by performing limited vocabulary automatic speech recognition on the

acoustic input”) to determine that the audio input includes the wakeword in the vocabulary (“to determine whether the acoustic input includes a word or phrase from a limited set, the limited set including the specific word or phrase”).

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.

223. Basye-Shenhav teaches Claim 7.

224. Basye’s second processor configured to perform the automatic speech recognition also performs a natural language processing stage to understand the meaning of the voice command. As I discussed above in Sections VIII.A.3.b and c, Basye’s application processing module 112’s microprocessor is the claimed second processor. As I discussed in §VIII.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. In one example, Gruber’s natural language processing 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. Specifically, Gruber states:

Language input 202 is received, such as the string “who is playing this weekend at the filmore” in the example of FIG. 29. In one embodiment,

the input is augmented by current context information, such as the current user location and local time. In word/phrase matching 210, *language interpreter component(s) 1070* find associations between user input and concepts. In this example, associations are found between the string “playing” and the concept of listings at event venues; the string “this weekend” (along with the current local time of the user) and an instantiation of an approximate time period that represents the upcoming weekend; and the string “filmore” with the name of a venue. *Word/phrase matching 210 may use data from, for example, language pattern recognizers 1060, vocabulary database 1058, active ontology 1050, short term personal memory 1052, and long term personal memory 1054.*

Ex-1008 ¶412.

225. Accordingly, in my opinion, Basye-Shenhav’s second processor thus performs natural language processing 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”).

226. Moreover, as the claim is drafted, only one of ASR, a classification processing stage or a natural language processing stages is required to understand the meaning of the voice command as explained by the ’750 Patent itself:

Speech processing stages may include, but are not limited to, *any one* or combination of automatic speech recognition (ASR) processes such as limited vocabulary ASR (e.g., evaluating the acoustic input to determine whether a user spoke an explicit voice trigger, evaluating the acoustic input to determine whether the acoustic input includes certain words characteristic of a voice command, etc.), unrestricted and/or large vocabulary ASR, etc. *Speech processing stages may include one or more language processes that seek to understand whether words in the acoustic input are characteristic of a voice command and/or to*

understand the meaning of what the user spoke. For example, speech processing stages may include one or more statistical classifiers, natural language processing (NLP), or any suitable language techniques designed to categorize or understand the content to facilitate determining whether a voice command has been spoken.

Ex-1001, 10:66-11:15. Basye-Shenhav teaches to use natural language processing in the same manner as described by the '750 Patent.

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.

227. Basye-Shenhav teaches Claim 8.

228. 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 in the same manner as described by the '750 Patent. As discussed above in in Sections VIII.A.3.b and .c, 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.

229. 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. For example, as I explained above, Zlokarnik is an example that teaches an "apparatus for voice-activated control of an electrical device comprises a receiving arrangement for receiving audio data generated by user" (Ex-1016, Abstract), and the certain advantages of HMM. *Id.*, 3:17-25.

230. As I previously explained, Basye's wakeword detection is based upon a limited vocabulary comprising a limited set of words or phrases in the same manner as described by the '750 Patent. 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.

231. 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 perform 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. For example, U.S. Patent No. 9,020,459 to Waters et al. ("Waters," Ex-1027) explains that reducing device and system power dissipation without compromising performance are important goals in receivers, microprocessors such as digital signal processors (DSPs). Ex-1027, 1:46-54. Waters notes that these goals are "especially important in hand held/mobile applications where small size is so important, to control the cost and the power consumed while achieving excellent performance." *Id.* One such way is by using "power gating," which is turns "power on and off to at least one portion of the process" at different times." *Id.*, 1:67-2:5, 7:1-4.

232. Further, a POSITA would have been motivated to implement the recited techniques 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. Doing so would have further been advantageous when network connectivity is down, preventing the server from performing the ASR.

- 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.**

233. Basye-Shenhav teaches Claim 11.

234. Basye's second processing stage comprises evaluating a contextual cue, such as time of day and location of the device. As discussed above in in Sections VIII.A.3.b and c, Basye's application processing module 112's microprocessor is the claimed second processor that performs the second processing stage. As discussed in §VIII.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.

235. Gruber teaches that voice input to the method employs contextual information. More particularly,

Language input 202 is received, such as the string "who is playing this weekend at the filmore" in the example of FIG. 29. *In one embodiment, the input is augmented by current context information, such as the current user location and local time.*

Id. ¶412.

236. For example, to interpret the command “who is playing this weekend at the filmore,” the current local time of the user is used as context to interpret the string “this weekend.” *Id.* Further, Gruber discloses the use of ***location information*** “to detect when a user is at home.” *Id.* ¶142. In this example, Gruber teaches:

Location information coming from sensors or location-based systems. Examples include Global Positioning System (GPS) and Assisted GPS (A-GPS) on mobile phones. In one embodiment, location information is combined with explicit user input. In one embodiment, the system of the present invention is able to detect when a user is at home, based on known address information and current location determination. In this manner, certain inferences may be made about the type of information the user might be interested in when at home as opposed to outside the home, as well as the type of services and actions that should be invoked on behalf of the user depending on whether or not he or she is at home.

Id.

237. Gruber also teaches the use of time as contextual information:

Time information from clocks on client devices. This may include, for example, time from telephones or other client devices indicating the local time and time zone. In addition, time may be used in the context of user requests, such as for instance, to interpret phrases such as “in an hour” and “tonight”.

Id. ¶143.

238. Gruber also discloses the use of sensor data as contextual information:

Compass, accelerometer, gyroscope, and/or travel velocity data, as well as other sensor data from mobile or handheld devices or embedded systems such as automobile control systems. This may also include device positioning data from remote controls to appliances and game consoles.

Id. ¶144.

239. Further, Gruber discloses user interactions and device events as context:

Clicking and menu selection and other events from a graphical user interface (GUI) on any device having a GUI. Further examples include touches to a touch screen.

Events from sensors and other data-driven triggers, such as alarm clocks, calendar alerts, price change triggers, location triggers, push notification onto a device from servers, and the like.

Id. ¶¶145-146.

240. Gruber's contextual cues are similar to the contextual cues disclosed in the '750 patent. *See, e.g.*, Ex-1001, 21:1-22.

241. Accordingly, in my opinion, Basye's natural language processing 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 of the device").

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.

242. Basye-Shenhav teaches Claim 13.

243. Basye teaches that its device performs a default action when a specific task could not be ascertained from the acoustic input. Basye incorporates by reference 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, the virtual assistant “may prompt the user for confirmation [and/or request additional context information from any suitable source] before calling a service ... to perform a function.” *Id.* In my opinion, the above actions (e.g., prompting for confirmation of a command) are exemplary “default actions” when a specific task is not able to be discerned from the voice input.

244. The above actions are consistent with “default” actions disclosed in the ’750 Patent when a command or task cannot be ascertained. For example, the ’750 Patent discloses:

According to some embodiments, a default action may be performed when a voice command is detected, but where there is sufficient uncertainty with respect to the intent of the voice command and/or the voice command cannot be suitably mapped to a specific task or set of tasks. The default action may be, for example, an internet search using one or more words recognized in the acoustic input. Alternatively, ***the default action may include formulating a question to be rendered to the user to obtain further information and/or clarification regarding the voice command.*** It should be appreciated that any default action may be assigned to voice commands that have been detected, but for which a specific task or set of tasks are not identified (e.g., with a sufficient level of certainty), as the aspects are not limited in this respect.

Ex-1001, 37:30-45.

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. Gruber itself contemplates the above scenario. More particularly, Gruber teaches that the functionality of its intelligent agent can be "implemented on a distributed computing network, such as one having any number of clients and/or servers." Ex-1008 ¶84. Moreover, Gruber teaches that several components relating to speech recognition can be implemented on a server. For example, Gruber teaches:

Additional components may be implemented as part of server 1340, including for example:

language interpreter 1070;

dialog flow processor 1080;

output processor 1090;

domain entity databases 1072;

task flow models 1086;

services orchestration 1082;

service capability models 1088.

Id. ¶¶110-117. Thus, Gruber at the very least suggests that its intelligent agent requests clarification, not only when speech recognition is performed locally, but also when a server is relied upon to perform the speech recognition.

246. This would have been nothing more than applying a known technique (Gruber's technique of prompting an end user to confirm the meaning of an ambiguous voice command, where speech recognition occurs on a server) to a known device ready for improvement (the Basye-Shenhav computing device, which utilizes a server for speech recognition) to yield predictable results (a computing device that asks a user to confirm the meaning of voice commands that the server finds ambiguous).

247. Accordingly, Basye-Shenhav teaches a computing device that includes an intelligent agent which 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").

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:**

248. To the extent the preamble is limiting, Basye teaches this element. 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 that “one or more modules” of its power management subsystem “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.*

249. Although Basye teaches that its power management subsystem and software modules may be carried out by “one or more 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 digital signal processor. As such, a POSITA would have understood that Basye’s low power digital signal processor was capable of accessing non-transitory computer readable media as part of executing its software. More particularly, as I discussed, Hwang discloses a low power DSP that performs speech detection and keyword detection.

With respect to the DSP, Hwang states:

a non-transitory processor-readable medium includes instructions that, when executed by a processor, cause the processor to dynamically switch between operating in a first mode and operating in a second mode. A ratio of CODEC activity to processor activity in the first mode is greater than a ratio of CODEC activity to processor activity in the second mode. The instructions, when executed, also cause the processor to transition out of a lower-power state during a duty cycle and to analyze audio feature data that is extracted during

the low-power state. The instructions, when executed, further cause the processor to transition back into the low-power state.

Ex-1010 ¶10.

250. A POSITA would have recognized that it is Hwang's low power DSP, as opposed to a general purpose processor, that accesses the above mentioned non-transitory processor-readable medium. For example, Hwang discloses that

[i]t should be noted that the CODEC and *the DSP may support additional operating modes as well*. Typically, activity of the CODEC may be greater than or equal to activity of the DSP. The various operating modes may correspond to different ratios of CODEC activity to processor activity. Each operating mode may include different settings for the duty cycle of the CODEC (where 100% corresponds to always on), *the duty cycle of the DSP*, and/or how many frames of audio data are analyzed each time the processor wakes up. The details of the supported operating modes may be determined at design-time and/or manufacturing-time. Which particular operating mode is selected may be determined at run-time based on factors such as application context.

Id. ¶61.

251. Indeed, a POSITA would have been motivated to configure Basye's digital signal processor 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.

252. 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 Sections VIII.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. For example, the '750 patent discloses:

Entering a low power mode may include turning off the display, shutting down one or more processors (e.g., disabling the clock tree associated with one or more processors) and/or deactivating components that would otherwise draw power if left operational.

Ex-1001, 5:14-18.

253. Accordingly, in my opinion, Basye teaches a non-transitory memory ("at least one non-transitory computer readable medium") storing 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 activation of the application processing module ("while the device is operating in a low power mode").

b. 14[a]: performing at least one processing stage on the acoustic input, the at least one processing stage comprising:

254. Basye-Shenhav teaches this element for the reasons I discussed in §VIII.A.2.b.

c. 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;

255. Basye-Shenhav teaches this element for the reasons I discussed in §VIII.A.2.b(i)(a).

d. 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;

256. Basye-Shenhav teaches this element for the reasons I discussed in §VIII.A.2.b(i)(b). Additionally, Basye's wakeword is an example of a "a trigger word or phrase," in a similar way as the '750 patent. More particularly, the '750 Patent 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.

- e. **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;**

257. Basye-Shenhav teaches this element for the reasons I discussed in §VIII.A.2.b(iii)). While elements 14[a(iii)] and 1[a(ii)] use slightly different language as shown in the table below, the scope and meaning of these elements are not substantively different as used in '750 patent. For clarity, I have highlighted the differences between the claims in the table below. The specification uses “explicit trigger word or phrase” and “specific and predetermined word or phrase” interchangeably, and refers to them both as an “explicit voice trigger.” Ex-1001, 1:38-44. Accordingly, this claim element 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	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 at least one voice command ;	the acoustic input includes one or more voice commands ;
---	---

- f. 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**

258. Basye-Shenhav teaches this element for the reasons I discussed in §VIII.A.2.b(iii).

- g. 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.**

259. Basye-Shenhav teaches this element for the reasons I discussed in §VIII.A.2.b(iv). While elements 14[a(v)] and 1[a(iv)] use slightly different language as shown in the table below (I have highlighted the difference between the claims in the table below), Basye-Shenhav teaches that the voice command is “recognized by the at least one server” when the voice command can be an intelligent agent query, such as the command “[t]ell me about flights to Seattle.” As Basye teaches, the 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, this claim element is taught for the same reasons discussed above, including the discussion of server-side recognition.

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: 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.**

260. Basye-Shenhav teaches Claim 15 for the same reasons I explained in regard to Claim 6. *See* §VIII.A.3.e.

6. Independent Claim 19

- a. Elements 19[pre], 19[a]: A device comprising: 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:**

261. Basye-Shenhav teaches Elements 19[pre], 19[a], for the same reasons I described in Elements 1[pre] and 1[a]. *See* Sections VIII.A.2.a and VIII.A.2.b.

- b. 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;**

262. Basye-Shenhav teaches Element 19[a(i)], for the same reasons I described in Element 1[a(i(A))]. *See* Section VIII.A.2.b(i)(a).

- c. 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;**

263. Basye-Shenhav teaches Element 19[a(ii)], for the same reasons I described in Element 1[a(i(B))]. *See* Section VIII.A.2.b(i)(b).

- d. 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;**

264. Basye-Shenhav teaches Element 19[a(iii)], for the same reasons as described in Element 1[a(ii)]. *See* Section VIII.A.2.b(ii).

- e. 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**

265. Basye-Shenhav teaches Element 19[a(iv)], for the same reasons I described in Element 1[a(iii)]. *See* Section VIII.A.2.b(iii).

- f. 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.**

266. Basye-Shenhav teaches Element 19[a(v)], for the same reasons as described in Elements 1[a(iv)] and 14[a(v)]. *See* Sections VIII.A.2.b(iv), VIII.A.4.g. While Element 19[a(v)] differs from Element 14[a(v)], as 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.” *See, e.g.,* Ex-1001, 1:15-18 (“Many mobile communications devices ... are equipped with a voice response system ... that can recognize speech and *respond to voice commands to perform desired tasks.*”). For clarity, I have highlighted the differences between the claims in the table below. Accordingly, this claim element 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. A POSITA would have been motivated to combine Basye's, Shenhav's, and Parker's teachings and would have had a reasonable expectation of success.

267. 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.

268. A POSITA also would have been motivated to combine the teachings of Parker with Basye-Shenhav because they are analogous art and teach improvements to the same type of devices. Each teach 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.” *Id.* ¶38. The action execution instructions, Parker discloses, “include information specifying one or more actions to be performed by [the client device.]” *Id.* Thus, after receiving the action execution instructions, the client device performs the actions the instructions specify. *Id.*

269. Parker teaches additional benefits to Basye-Shenhav’s computing device. In particular, as I discussed in Section VI.C, Parker’s client device operates in what the reference refers to as a “conversation mode.” In Parker’s conversation mode, “command to be performed by client device 104. Name recognition application 108 may operate in conversation mode by sending audio command information 116 to server 102 for processing, as described in further detail below.” Ex-1007 ¶36. Once the audio command information is received by the server, the server performs speech recognition on it to generate action instructions for the client device to carry out. *See id.* ¶37. Parker observed that using a server may introduce delay into the speech recognition process:

In the example of FIG. 4, audio command information 116 is processed by speech recognition manager 106 on server 102. ***Due to the time required for server 102 to process audio command information 116, a time lag may exist from when user 110 speaks audio command information 116 to when user 110 receives action execution instructions 118.***

Id. ¶70.

270. Parker's system notes that an "audio command acknowledgement provides [the] user ... with an assurance that the audio command information ... is being processed." *Id.* To address potential uncertainty on the part of end users due to this time lag, an "audio command acknowledgement provides user 110 with an assurance that the audio command information 116 is being processed." *Id.* Thus, Parker discloses that its client device acknowledges receipt of an audio command promptly after receipt:

[C]lient device 104 ***acknowledges receipt of the command the user gives (e.g., the command being the second utterance by the user after name acknowledgement by client device 104), for example, immediately after receiving the command and before client device 104 has processed the command either locally or by sending the command over a network to a server for processing.*** By quickly acknowledging receipt of the command, client device 104 is able to process the command, which may take a few seconds, without the user being unsure as to whether client device 104 received the command.

Id. ¶93.

271. Parker's client device, moreover, is capable of providing either visual or audio notifications to the end user. For example, Parker discloses that the client device includes training module that generates an audio acknowledgement upon receiving audio name information:

Training module 208 generates (306) an audio acknowledgement of the received audio name information. In an example, training module 208 generates an audio acknowledgement that when rendered on a microphone associated with client device 104 notifies user 110 of the received audio name information.

Id. ¶63. *See also id.* ¶¶64, 54 (“The graphical user information may include a visual representation of a conversation between user 110 and client device 104, for example, when the graphical user interface is rendered on a display of client device 104.”).

272. Based on Parker’s teaching, a POSITA would have been motivated to implement Parker’s technique of immediately providing an acknowledgement that its client device received audio input into Basye-Shenhav’s computing device to reduce user uncertainty that their audio input was received and potentially conserve system resources. More specifically, in my view, a person of ordinary skill in the art would have understood that time lags between providing a voice command and receiving a response may result in user uncertainty as to whether the voice command was received by the device. In this scenario, a user may choose to repeat the command unnecessarily, perhaps multiple times, which, in my opinion, one of ordinary skill would recognize as further overburdening the device, as well as the network and server. Thus, Parker’s immediate acknowledgement technique would be a reasonable way to reduce user uncertainty and potentially conserve system resources because Parker’s visual or audio acknowledgement need not require additional computation and waking up the second processor (as compared to Basye’s text-to-speech responses. *See* Ex-1005 ¶66 (for application processing module 112 to “cause presentation of the results ... in audible format ... or in visual format (e.g.,

via a display of the user computing device 200),” the computing device would need to select to proceed with activating application processing module 112, which, as a result, would select to transition the computing device from the lower power mode into a higher power mode.).

273. A POSITA would have had a reasonable expectation of success in applying Parker’s teachings to the Basye-Shenhav computing device because Parker’s acknowledgement technique is compatible with the Basye-Shenhav computing device. For example, the Basye-Shenhav computing device includes speakers (Ex-1005 ¶42), which a POSITA would understand as enabling the device to provide the types of audio acknowledgements Parker provides. Ex-1007 ¶69 (“Name recognition application 108 generates ... an audio acknowledgement to notify user 110 of receipt of audio command information 116 by client device 104.”).

2. Dependent Claims 9 and 16

a. Claim 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 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.

274. Basye-Shenhav-Parker teaches Claims 9 and 16. 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 microprocessor of application module 112 (the claimed “second processor”). *See* Sections VIII.A.3.b, VIII.A.4.a (low power mode).

275. As I explained above, 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 discloses that its client device acknowledges receipt of an audio command promptly after receipt:

[C]lient device 104 *acknowledges receipt of the command the user gives (e.g., the command being the second utterance by the user after name acknowledgement by client device 104), for example, immediately after receiving the command and before client device 104 has processed the command either locally or by sending the command over a network to a server for processing.* By quickly acknowledging receipt of the command, client device 104 is able to process the command, which may take a few seconds, without the user being unsure as to whether client device 104 received the command.

Id. ¶93.

276. Parker’s client device, moreover, is capable of providing either visual or audio notifications. *Id.* ¶¶54, 64.

277. In addition to the reasons explained in Section VIII.B.1, a POSITA would have been motivated to implement Parker’s user indication into Basye-Shenhav’s computing device before activating the second processor because, “[b]y quickly acknowledging receipt of the command, [its] client device ... is 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.” Ex-1007 ¶93. In such a case, a user may choose to repeat the command unnecessarily, perhaps multiple times, further overburdening the device, as well as the network and server.

278. Accordingly, in my opinion, a POSITA would have understood Basye-Shenhav-Parker to teaches providing an audio acknowledgement to an end 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 microprocessor of the device's application processing module ("prior to activating the second processor or turning on a display of the device").

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

1. A POSITA would have been motivated to combine Basye's, Shenhav's, and Lebeau's teachings and would have had a reasonable expectation of success.

279. 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.

280. A POSITA also would have been motivated to combine the teachings of Lebeau with Basye-Shenhav because they are analogous art and teach improvements to the same type of devices. Each teach 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").

281. Lebeau offers additional benefits to Basye-Shenhav's computing device. Lebeau teaches that the mobile device will detect the mobile device's condition in order to determine whether to switch the device to monitor 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. Lebeau explains that a context includes whether the mobile device's battery is "running low on stored power (e.g., below 25% charge remaining in the battery)" (Ex-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.

282. 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.

283. A POSITA would have had a reasonable expectation of success in applying Lebeau's teachings to the Basye-Shenhav 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 (Basye's computing device considers power level to selectively activate or deactivate modules). The electronic components of Lebeau

with Basye-Shenhav's devices are similar and are 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.

2. Dependent Claims 10 and 18

a. Claim 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 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.

284. Basye-Shenhav-Lebeau teaches Claims 10 and 18.

285. Basye-Shenhav teaches performing a first processing stage while in the lower power mode. *See* Sections VIII.A.3.b (Claim 3) and VIII.A.4.a (Element 14[pre]). As described in Section VIII.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 notes to consider the mobile device's

power supply and whether it “drops below a threshold power supply level (e.g., a battery charge percentage).” Ex-1005 ¶79; *see also id.* ¶43 (“the power supply 218 communicates the level of power that it can supply to the power management subsystem 100 (e.g., a percentage of battery life remaining, whether the power supply 218 is plugged into an electrical outlet, etc.). In some embodiments, the power management subsystem 100 selectively activates or deactivates one or more modules based at least in part on the power level indicated by the power supply.”).

286. 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.

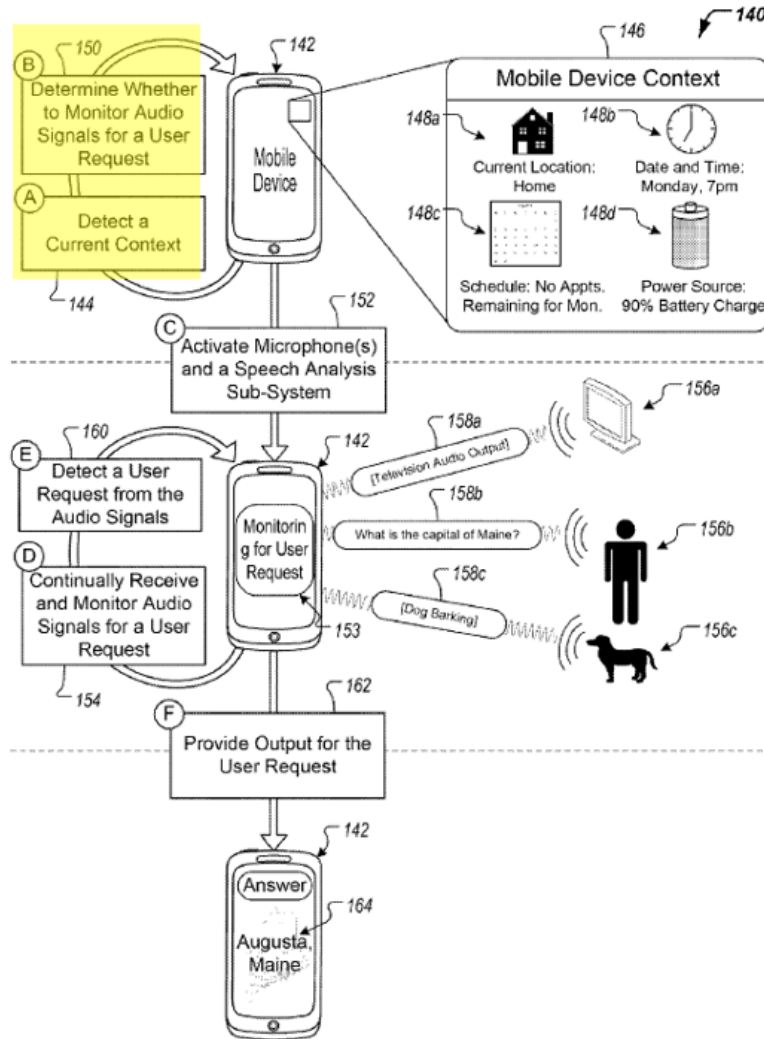


FIG. 1B

287. For example, Lebeau teaches to use the contextual cue of the time of day to determine “a threshold likelihood of [the user] providing a voice-based request to the mobile computing device.” *Id.* ¶43. Moreover, Lebeau explains that a context includes whether the mobile device’s battery is “running low on stored power (e.g., below 25% charge remaining in the battery)” (Ex-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 when evaluating the acoustic input, 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.

288. 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. This was a common way to use contextual cues, as I discussed above in Section V.A. For example, as I explained in

Section V.A, battery status is a well-known consideration in a speech recognition system and is “particularly advantageous for computing devices that rely on battery power, all computing devices for which power management may be desirable can benefit from the principles of the present disclosure.” Ex-1019 ¶14. Basye already notes that by keeping the computing device in a persistent active state, the first computing device may consume unacceptable amounts of energy to maintain the persistently active state. Ex-1019 ¶9. “Such energy demands may prove especially problematic for mobile computing devices that rely on battery power.” *Id.* Thus, Basye shows how it was well known to consider the battery status, such that “if the power supply 218 drops below a threshold power supply level (e.g., a battery charge percentage), as on-device speech recognition capabilities as implemented by the processing unit 202 and/or the application processing module 112 may require a relatively large power draw.” *Id.* ¶79.

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

1. A POSITA would have been motivated to combine Basye’s, Shenhav’s, and Conroy’s teachings and would have had a reasonable expectation of success.

289. 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.

290. 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, such as computing devices that operate in low power state in order to conserve battery *See, e.g.*, Ex-1005 ¶15; Ex-1006 ¶27; Ex-1017 ¶¶14, 23, 33.

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." *Id.* ¶¶31-34.

292. 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.

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

2. Dependent Claims 12 and 17

- a. Claim 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 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.

294. Basye-Shenhav-Conroy teaches Claims 12 and 17.

295. As I discussed in Sections VIII.A.2.b(ii) and VIII.A.2.b(iii), Basye teaches selecting to transition to a higher power mode in response to receiving an indication of a voice command. Basye transmits the acoustic input of the wake word and the voice command to the speech recognition server while in lower power mode (i.e., the application processing module is not activated). Specifically, Basye’s “speech recognition server 420 may recognize speech (9) included in the audio input and generate speech recognition results (10) therefrom.” Ex-1005 ¶66, Fig. 4B. While the speech recognition results are transmitted to the 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.*) and thereby waking 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.

296. 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-1017), which do just this. For example, Conroy teaches an electronic device, such as a media player or cell phone (Ex-1017 ¶23), with a “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 trigger 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 may be configured to enter a low power state while the audio

data is processed, by for example, a digital signal processor. For example, Conroy explains that:

the present invention are directed toward a multimedia device with an independent audio subsystem that handles all of the decoding, mixing, equalizing and routing of the audio signals, and then sends the output audio directly to peripheral output devices. ***Because the audio subsystem handles all of the processing of audio data, the CPU is not needed beyond the stage of routing audio data to the audio subsystem; therefore, the CPU and the rest of the host subsystem may be configured to enter a low power state while audio data is processed.***

Ex-1017 ¶14. *See also id.* ¶¶31 (“the host subsystem 300 includes a CPU 304, which controls and directs most functions of the device 100 other than audio processing. The audio subsystem 301, on the other hand, controls substantially all of the audio processing functions of the device 100.”), 33 (“Because the audio processing at step 212 can take place with little or no involvement of the CPU, the CPU is free to carry out other processing tasks or alternatively to enter a low power state which extends the useful battery life of the electronic device 100.”).

297. 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. *Id.* 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, such as playing music. Ex-1017 ¶¶14, 31, 33.

298. In particular, a POSITA would have understood that Basye-Shenhav would have received and transmitted the acoustic input from the user requesting to play music to the server for the server to determine the voice command, and in response to receiving an indication from the server that acoustic input includes a voice command, the application processor would start the music playback, and then select to remain in lower power mode to respond to the voice command of plying music, as taught by Conroy. Ex-1017 ¶¶14 (“***Because the audio subsystem handles all of the processing of audio data, the CPU is not needed beyond the stage of routing audio data to the audio subsystem; therefore, the CPU and the rest of the host subsystem may be configured to enter a low power state while audio data is processed.***”), 31 (“the host subsystem 300 includes a CPU 304, which controls and directs most functions of the device 100 other than audio processing. The audio subsystem 301, on the other hand, controls substantially all of the audio processing functions of the device 100.”), 33 (“Because the audio processing at step 212 can take place with little or no involvement of the CPU, the CPU is free to carry out other processing tasks or alternatively to enter a low power state which extends the useful battery life of the electronic device 100.”).

299. 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. Therefore it would have been obvious to

Declaration of Mr. Christopher M. Schmandt

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.

IX. SECONDARY CONSIDERATIONS

300. I am not aware of any evidence in the '750 patent's prosecution history or elsewhere supporting any secondary considerations arguments, or evidence of nexus of such alleged evidence to the Challenged Claims. *See generally* Ex-1004. To the extent Patent Owner asserts the existence of any secondary considerations in its responses, I reserve the right to address any such evidence.

X. CONCLUSION

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

Executed on: 13 October 2024

A handwritten signature in black ink, appearing to read 'Chris' followed by a stylized surname.

Christopher M. Schmandt