

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

KEA CLOUD, INC., HI AUTO, INC.

Petitioners

v.

VALYANT AI, INC.

Patent Owner

Inter Partes Review No.: To Be Assigned

U.S. Patent No. 10,592,706

DECLARATION OF BORIS GALITSKY, PH.D.

Table of Contents

I.	QUALIFICATIONS AND PROFESSIONAL EXPERIENCE.....	5
II.	SUMMARY OF OPINIONS.....	9
III.	RELEVANT LEGAL STANDARDS	9
A.	Person of Ordinary Skill in the Art.....	10
B.	Invalidity by Anticipation (35 U.S.C. § 102).....	10
C.	Invalidity by Obviousness (35 U.S.C. § 103).....	11
D.	Continuity of Disclosure (35 U.S.C. §§ 112(a), 119(e), and 120).....	12
IV.	STATE OF THE ART PRIOR TO THE ‘706 PATENT.....	12
A.	Automated Speech Recognition (“ASR”)	14
B.	Natural Language Processing (“NLP”)	18
C.	Automated Voice Ordering at a Drive-Thru.....	20
V.	THE ‘706 PATENT	21
A.	Specification of the ‘706 Patent	21
B.	File History of the ‘706 Patent	25
C.	Effective Filing Date of the ‘706 Patent	27
VI.	CHALLENGED CLAIMS AND CLAIM CONSTRUCTIONS.....	31
VII.	QUALIFICATIONS OF A PERSON OF ORDINARY SKILL	31
VIII.	OVERVIEW OF THE PRIOR ART	32
A.	Mahaffy – U.S. Pat. Appl. Pub. No. 2003/0018531.....	32
B.	Jasper – U.S. Pat. Appl. Pub. No. 2007/0007331	35
C.	Coleman – U.S. Pat. Appl. Pub. No. 2019/0108566.....	35
D.	Washington – U.S. Pat. Appl. Pub. No. 2017/0293905	37

IX. GROUND 1 – CLAIMS 1-11 ARE OBVIOUS OVER MAHAFFY IN VIEW OF JASPER.....	38
A. Motivation to Combine.....	38
B. Claim 1	42
C. Claim 2	62
D. Claim 3	62
E. Claim 4	66
F. Claim 5	76
G. Claim 6	78
H. Claim 7	79
I. Claim 8	80
J. Claim 9	81
K. Claim 10	82
L. Claim 11	84
X. GROUND 2A – CLAIMS 1-4 ARE OBVIOUS OVER COLEMAN	85
A. Claim 1	85
B. Claim 2	106
C. Claim 3	107
XI. GROUND 2B – CLAIMS 4-11 ARE OBVIOUS OVER COLEMAN IN VIEW OF WASHINGTON	109
A. Motivation to Combine.....	109
B. Claim 4	113
C. Claim 5	124

D. Claim 6	127
E. Claim 7	128
F. Claim 8	129
G. Claim 9	130
H. Claim 10	132
I. Claim 11	133
XII. OBJECTIVE EVIDENCE OF NONOBVIOUSNESS	134
XIII. CONCLUSION	135

I, Boris Galitsky, hereby declare as follows:

1. I am making this declaration at the request of Petitioners Kea Cloud, Inc. (“Kea”), Hi Auto, Inc. (“Hi Auto”), and E La Carte, Inc. d/b/a Presto (“Presto”) in the matter of Inter Partes Review of U.S. Patent No. 10,592,706 (“the ‘706 Patent”).

2. I am being compensated for my work in this matter at a rate of \$250/hour. My compensation in no way depends on the outcome of this proceeding.

3. In preparation of this declaration and in forming the opinions expressed below, I have considered the references cited throughout my declaration, the relevant legal standards, and my knowledge and experience based upon my work and study in this area as described below.

I. QUALIFICATIONS AND PROFESSIONAL EXPERIENCE

4. I have provided my full background in the curriculum vitae that is attached as EX 1030 to this declaration. In the paragraphs that follow, I highlight some of my relevant qualifications and background concerning the technology at issue in this proceeding.

5. I received my Masters of Science in Computer Science in 1991 and my Ph.D. in Computer Science in 1993, both from the Institute for the Information Transmission Problems in Moscow, Russia. EX 1030.

6. As reflected in my curriculum vitae, I have worked for private companies for the last 23 years developing software products for various industries, primarily focusing on the design and development of search and recommendation engines using natural language processing technology and artificial intelligence. EX 1030.

7. As reflected in my curriculum vitae, I have also authored six published books in the field of artificial intelligence, including on the topics of natural language processing and artificial intelligence. EX 1030.

8. I authored a two-part research monograph, entitled Artificial Intelligence for Customer Relationship Management, which provided a detailed discussion of automated speech recognition (“ASR”), natural language processing (“NLP”) and machine learning (“ML”), as applied to customer relationship management (“CRM”). The first volume was published in 2020. The second volume was published in 2021. EX 1030.

9. I also authored a book, entitled Developing Enterprise Chatbots, which discussed the applications of discourse analysis in text relevant assessment, dialogue management and content generation, which help to overcome the limitations of deep learning-based and data driven-based approaches to enterprise chatbots. EX 1030.

10. In 1999, I co-founded and was the CTO of iAskWeb, Inc. I was responsible for developing, designing and implementing search and

recommendation software that utilized ASR and NLP to understand user queries and make recommendations. I managed a development team of 20+ engineers, content managers and testers on the project, which resulted in software used by companies in the tax and financial industries. EX 1030.

11. In 2006, I was a Founding Scientist at Uptake.com. I was responsible for developing and designing search and recommendation software for use in the hotel industry. The search and recommendation software provided hotel recommendations based on NLP understanding of prior user experiences from customer reviews. EX 1030.

12. In 2008, I was a Consulting Scientist for several “.com” companies. My consulting work primarily involved the design and development of search and recommendation engines for various industries. EX 1030.

13. In 2010, I was a Lead Scientist at eBay, Inc. I was responsible for developing and designing product search and recommendation software that used NLP to understand user queries and recommend products based on those queries. I was also involved in a search engine optimization project that relied on NLP. EX 1030.

14. In 2012, I was a Chief Scientist at Become.com, which was acquired by Shopzilla / Connexity. I was responsible for designing product search and

recommendation software. I also managed a team of four software engineers on the project. EX 1030.

15. In 2014, I was the Head of Semantic Analysis at Elastica.net, which was acquired by Blue Coat / Symantec. My responsibilities included designing software for the ASR and NLP processing and classification of documents for enterprise system security and data loss prevention. EX 1030.

16. In 2015, I was the Director of Data Science at Ligadata Inc. I was responsible for designing search and recommendation software for the financial investment products. My responsibilities also included leading a small team of software engineers on the project. EX 1030.

17. In 2016, I was a Software Architect at Oracle Corp. I was responsible for designing and developing a task-oriented chatbot for customer support, called the “Oracle Digital Assistant.” The chatbot utilizes ASR and NLP to engage in natural conversation with users for customer support. While at Oracle, I provided technology leadership for a team of six software engineers. EX 1030.

18. I am currently the Principal Data Scientist at Relyance AI. My responsibilities included the design and development of software for NLP processing and organization of legal documents. EX 1030. I have held this position since January 2022.

19. As reflected in my curriculum vitae, I am also a named inventor on several patents in the field of natural language processing and artificial intelligence. My curriculum vitae includes a list of these patents. EX 1030.

II. SUMMARY OF OPINIONS

20. It is my opinion that:

- a. Claims 1-11 of the '706 Patent are obvious over Mahaffy (EX 1020) in view of Jasper (EX 1021).
- b. Claims 1-11 of the '706 Patent are obvious over Coleman (EX 1022) in view of Washington (EX 1023).

21. My opinions are based on the explanations and reasonings provided in further detail herein. While I believe I have provided such explanations and reasonings sufficient to fully support my opinions, I reserve the right to provide further details to the extent necessary to clarify or otherwise explain my opinions.

III. RELEVANT LEGAL STANDARDS

22. I am not a lawyer, and I am not providing any legal opinions, but I understand from counsel the following legal standards and that those standards are applicable with regard to the validity or patentability of a United States patent. I have applied this law to the facts in this matter in my analysis and in rendering my technical opinions.

A. Person of Ordinary Skill in the Art

23. I am informed that the term “person of ordinary skill in the art” (“POSA”) in the context of a patent validity challenge refers to a hypothetical person who is presumed to have knowledge of the relevant prior art at the time of a patented invention. I understand that many factors may influence the level of ordinary skill in the art, including: (1) the educational level of the inventor; (2) the type of problems encountered in the art; (3) prior art solutions to such problems; (4) the rapidity with which innovations are made in the relevant field; (5) the sophistication of the technology to which the patent is directed; and (6) the educational level of active workers in the field. I understand that a POSA is considered to be a person of ordinary creativity, not an automaton, meaning that a POSA may employ inferences and creative steps in his or her work.

B. Invalidity by Anticipation (35 U.S.C. § 102)

24. I understand that a patent claim is “anticipated,” and therefore invalid, if a single prior art reference discloses, either expressly or inherently, each and every element of the claimed invention in a manner sufficient to enable a POSA to practice the invention, thus placing the invention in possession of the public.

25. I understand that an anticipatory reference may “inherently” disclose a claim limitation if the explicitly missing matter is necessarily present in the apparatus or method described in the reference. The disclosure must be sufficient to

show that the natural result flowing from the operation of the apparatus or method disclosed in the reference would require the missing matter.

C. Invalidity by Obviousness (35 U.S.C. § 103)

26. I understand that even if a prior art reference fails to anticipate a patent claim, the claim may nonetheless be invalid as “obvious” if the differences between the subject matter claimed and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a POSA.

27. I understand that several factual inquiries underlie a determination of obviousness. These inquiries include the scope and content of the prior art, the level of ordinary skill in the art, and the differences between the claimed invention and the prior art.

28. I understand that a combination of familiar elements according to known methods may be obvious when it does no more than yield predictable results. I further understand that a simple substitution of one known element for another may be obvious to do when it does no more than obtain predictable results, and that the use of a known technique to improve similar devices in the same way may likewise be obvious. I also understand that making choices from a finite number of identified, predictable solutions, with a reasonable expectation of success, may be obvious to do. I also have been advised that common sense and ordinary creativity of one skilled in the art can be relevant to obviousness.

29. I have been additionally informed that certain secondary considerations may be relevant to a determination of whether an invention was obvious. Such secondary considerations may include: (a) whether there was a long-felt and long-unmet need for the invention; (b) whether the invention achieved unexpected results; (c) the commercial success of the invention; and (d) whether the invention was copied or praised within the industry.

D. Continuity of Disclosure (35 U.S.C. §§ 112(a), 119(e), and 120)

30. I am informed that, in order for a later filed patent to claim priority to an earlier filed patent application, there must be a continuity of disclosure between the later filed patent and the earlier filed patent application.

31. I am further informed that, in order for there to exist such a continuity of disclosure, the claims of the later filed patent must find adequate written description support in the disclosure of the earlier filed patent application. I understand that adequate written description support exists when a POSA could reasonably conclude from the disclosure of the earlier filed patent application that, as of the filing date of the earlier filed patent application, the inventor had possession of the subject matter claimed in the later filed patent.

IV. STATE OF THE ART PRIOR TO THE ‘706 PATENT

32. The ‘706 Patent generally relates to automating the traditional drive-thru voice ordering process, where an auditor switches to a human-assisted workflow

when the automated ordering process is “not functioning properly,” as is discussed in Section V below. Neither the concept of automated drive thru voice ordering, nor the concept of switching to a human-assisted workflow when the automated ordering process is “not functioning properly” were new in March 2017.

33. The speaker box drive-thru voice ordering process was introduced in the 1970s. The process is a familiar staple of American food service, and generally proceeds as follows: the customer arrives at the drive-thru and speaks their order into a microphone, an employee hears the order via a headset and enters the ordered menu items into the order processing system for preparation by the kitchen staff, the customer then pays for and receives their ordered menu items, and, finally, the customer drives away. EX 1040; EX 1022, ¶[0036]. See also, e.g., EX 1020, ¶[0005]; EX 1041.

34. In the early 2000s, speed, accuracy and service considerations prompted the fast-food restaurant industry to look for ways to outsource their order-taking functions and free up on-site employees for other tasks. By as early as 2009, the traditional drive-thru order taking was being outsourced to off-site order takers located at remote call centers. EX 1042.

35. By at least early 2016, automated speech recognition (ASR) and natural language processing (NLP) technology had been applied to automate the ordering taking process in fast-food restaurants. EX 1043.

A. Automated Speech Recognition (“ASR”)

36. ASR is sometimes called “speech recognition” or “speech-to-text,” and refers to the computer process by which human speech is converted into natural language text. EX 1050; EX 1051.

37. At a high level, most if not all ASR systems of the relevant time period followed the same basic approach to convert human speech into natural language text. Accordingly, when ASR or “speech recognition” is discussed in publications in the relevant time period, one of ordinary skill in the art would by default understand it to mean ASR following this basic approach, unless the discussion expressly referred to an outdated historical or outlier approach.

38. This basic approach to ASR relies on the Hidden Markov model and neural networks. It was developed around the mid 1980’s and was quickly adopted as the standard for ASR. EX 1052, pp. 12-18.

39. A detailed explanation of the Hidden Markov model and neural networks is not necessary for the purposes of my analysis because the claims of the ‘706 patent do not claim ASR to that level of detail. Instead, below I provide a general explanation of the basic approach as a POSA would understand it by the critical date. To the extent necessary, I reserve the right to provide a detailed explanation, and I am happy to provide such a detailed explanation upon request. Nevertheless, a POSA by the critical date would understand that most if not all ASR

systems of the relevant time period followed the same basic approach, described below, and that that approach was well known long prior to 2017.

40. The basic approach begins by sampling a recorded audio signal of the spoken words. That is, the waveform of the digitized speech signal is broken into a sequence of very short fragments. Each of the fragments is then converted into a corresponding vector, which is representative of the power spectrum of the fragment. That power spectrum is essentially the power of the signal as a function of frequency. These vectors are sets of numbers reflecting the power spectrum, and may be from 10 to 32 dimensions or more. EX 1053.

41. In order to convert the resulting sequence of vectors into text, groups of vectors are matched to one or more phonemes. A phoneme is a fundamental unit of speech reflecting perceptually distinct sounds. The English language has 44 phonemes, such as /b/, /a/ and /d/ in the word *bad*. This results in a text string or sequence of phonemes and pauses that represent the recorded speech. EX 1053.

42. The sequence of phonemes and pauses is then provided to a machine learning model, e.g., a neural network, which determines the most likely words or sequence of words (i.e., phrases) characterized by the phoneme sequence. This involves pattern matching the sequence of phonemes to the phonemes of known words or phrases in a corpus of known words and phrases. The pattern matching results in a text string of words corresponding to the recorded speech. EX 1053.

43. Because of variances in speech between speakers (and even between utterances of the same speaker), the machine learning model (e.g., neural network) is generally trained to recognize words and phrases in the context of the particular speech patterns of the speakers. For example, in a gardening business context, the machine learning model (e.g., neural network) may be trained to select “flower” over “flour” in the phrase “the flower pot display.” The machine learning model (e.g., neural network) may also be trained to consider semantic and contextual information, such as grammar. EX 1053.

44. At a high level, training the machine learning model (e.g., neural network) involves manually transcribing speech into words and phrases, and assigning those words and phrases to the ASR-determined phoneme sequences of the speech, in order to build the corpus of known words and phrases that is used to train the machine learning model. This is sometimes known in the art as a supervised learning approach.

45. The training also involves speaking to the ASR and correcting mistakes in the ASR’s transcription of the speech in a training feedback loop. Where corrections are made, the corrected phoneme sequence to known word association is stored as a reference for the neural network to pattern match against. Where corrections are not made, the phoneme sequence to known word association is reinforced. EX 1053. This feedback loop process is sometimes known in the art as

improved training or extended training of the ASR using expanded training datasets generated from the mistakes and corrections.

46. One of ordinary skill in the art would be readily familiar with the fact that most, if not all, ASR systems in use after the 1990's followed the same basic approach described above. Moreover, one of ordinary skill in the art would further understand that, within the relevant time periods, any computer that transcribes speech into text necessarily relies on the same basic ASR approach described above.

47. Dragon Naturally Speaking software is an example of ASR software that is known in the art to follow the basic approach described above. EX 1058; EX 1060. Dragon's Naturally Speaking was initially released in 1997 and is speech-to-text software that allows a user to dictate to a word processing program. EX 1060. It is trained to recognize user speech by comparing a corpus of text words with audio of a user speaking those words, so as to train the speech-to-text software to recognize user speech. EX 1058, pp. 21-24; EX 1060. Dragon Dictation is analogous software released in 2009 by the same company for Apple IOS platforms (whereas Dragon Naturally Speaking is for Windows). EX 1059. Dragon Dictation is known to be based on the same ASR approach as Dragon Naturally Speaking, and in relevant respects utilizes the same speech-to-text technology. EX 1059; EX 1060.

B. Natural Language Processing (“NLP”)

48. NLP refers to the computer process by which natural language text is converted into a formal representation, such as programming language or logical statements, that is directly readable by the computer. EX 1054; EX 1055.

49. The most common NLP approaches of the 2000’s were: a rules-based approach, a statistical model based approach, and a neural network based approach. A detailed explanation of these specific approaches is not necessary for the purposes of my analysis. It is sufficient for my analysis that all NLP systems of the relevant time period, at a high level, followed the same basic process to convert natural language text into a formal representation, which basic process was known well prior to 2016. EX 1055.

50. According to the basic NLP process, in the context of natural language interaction with users, computer understandable meaning is discerned from natural language text. The discerned meaning of the text is then matched to the most likely of a corpus of stored intents of the user, each of which are associated preprogrammed computer actions. The associated actions are then subsequently executed by the computer.

51. For example, Amazon’s Alexa (widely available since 2015) utilizes ASR to convert the spoken phrase: “Alexa, off” into a corresponding text string, and

then uses NLP to understand the text string as a command to turn off, which it then does. EX 1056.

52. One of ordinary skill in the art would be readily familiar with the fact that all NLP systems of the relevant time period discern computer actionable meaning from natural language text. Moreover, one of ordinary skill in the art would further understand that any computer that receives natural language text and performs some action responsive to the natural language meaning of the text is, at a high level, necessarily relying on the basic NLP process described above.

53. Accordingly, a person of ordinary skill in the art looking to the computer automation of a traditionally speech-based human interaction, such as for example a telephone operator interaction, a customer service interaction, or a food ordering interaction, would know that such automation would require the ASR and NLP procedures discussed above.

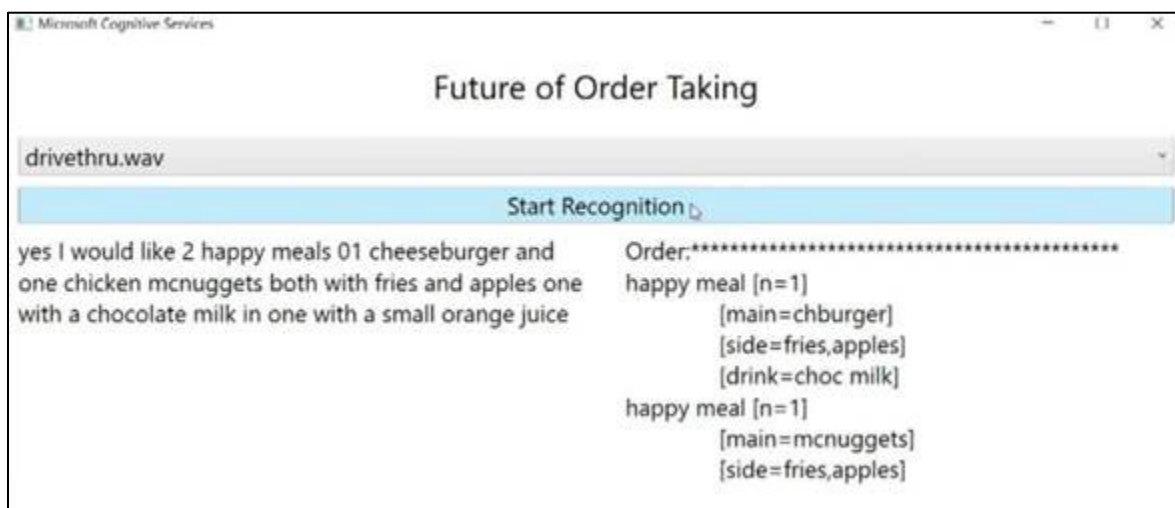
54. Indeed, such computer systems that automate traditionally speech-based human interaction using NLP were and are known in the art as a “conversational agents,” or sometimes as “automated online assistants” or “chatbots,” all of which were well-known by 2016. EX 1055.

C. Automated Voice Ordering at a Drive-Thru

55. As noted above, by at least early 2016, automated speech recognition (“ASR”) and natural language processing (“NLP”) technology had been applied to automate the ordering taking process in fast-food restaurants. EX 1043.

56. In the automated ordering process or workflow, instead of a human employee, a computer using ASR and NLP takes the customer’s verbal order and enters it into the restaurant’s order processing system to be fulfilled. EX 1020; EX 1022; EX 1043.

57. This would be done by converting the customer’s verbal order to natural language text via ASR, and that natural language text would have been further into a formal representation of the order via NLP. An example of this two-step process is illustrated in the below image of McDonald’s AI Drive Thru:



EX 1043. See https://www.youtube.com/watch?v=Y_rSb4Vb5zA.

58. As shown in the above image, the customer's audio order is recorded in the drivethru.wav file. The ASR converts the audio order to the natural language text shown in the left column. That natural language text is then converted into a formal representation of the order, shown in the right column, that can be understood by the restaurant's order processing system. EX 1043.

59. At the same time, it was also well known that ASR and NLP technology could, on occasion, break down or otherwise fail to accurately take customer orders. EX 1020; EX 1022. The natural response to this possibility is to monitor the automated workflow and to switch from the automated workflow to a human-assisted workflow when the automation was not functioning properly. EX 1020; EX 1022.

60. That automated processes sometimes fail is not a new phenomenon. Nor is the concept of reverting to pre-automated processes when automated processes fail. That natural response is as old as automation itself. For example, when a calculator fails, one naturally reverts to doing long division with pen and paper.

V. THE '706 PATENT

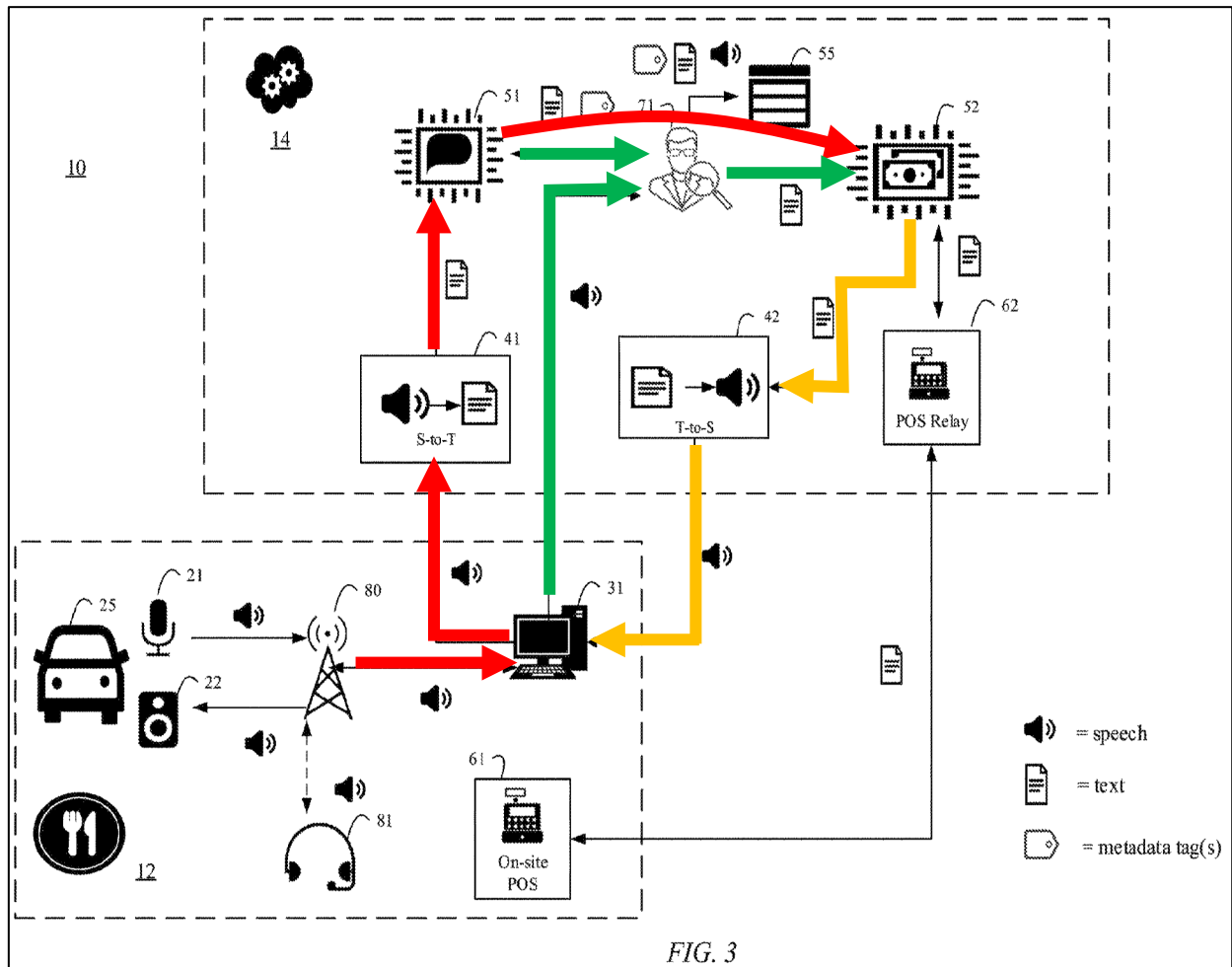
A. Specification of the '706 Patent

61. I have reviewed the specification and drawing figures of the '706 Patent. EX 1001. Below, I briefly discuss some of the concepts mentioned in the

patent that are referred to in the claims all of which, as discussed generally above and more specifically with reference to the prior art below, was well known by the critical date.

62. The '706 Patent describes an order processing system that automates the ordering function of a drive-thru restaurant by taking over the order processing function from an on-site employee. EX 1001, 12:46-56. A human auditor can abort or intervene in the autonomous workflow to pass the order processing back to an on-site employee. EX 1001, 12:48-53; 13:23-43.

63. The below annotated reproduction of FIG. 3 illustrates the workflows of the order processing system 10:



EX 1001, FIG. 3.

64. With reference to annotated FIG. 3 (above), the order processing system 10 includes an autonomous workflow (red) and a human-assisted workflow (green). EX 1001, FIGS 3-4. Both the human-assisted workflow and the autonomous workflow begin with a customer 25 speaking into a microphone 21 at the drive-thru order station 12 to place an order. EX 1001, 12:57-61. The customer audio is passed, via computer 31, to a speech-to-text module 41 that transcribes the audio into text. EX 1001, 7:29-44; 12:48-56; 13:7-16. The transcribed text is output

to a natural language processor (NLP) 51 that identifies order items (e.g., “one cheeseburger”) from the transcribed text. EX 1001, 13:17-23.

65. In the human-assisted workflow (green), the identified order items are passed to a human auditor 71 along with the customer audio for comparison. EX 1001, 12:48-53; 13:23-27. If the auditor 71 detects that the autonomous workflow is not functioning properly, or that the customer 25 is frustrated or otherwise not being adequately served by the autonomous workflow, the auditor can abort, or intervene in, the ordering processing and pass control to an on-site employee 81 who is at the same location as the customer. EX 1001, 13:33-43.

66. In the autonomous workflow (red), the auditor 71 only audits orders when the system detects an exception. EX 1001, 14:22-39; FIGS. 3-4. The autonomous workflow in FIG. 3 above (green), for example, does not involve the auditor because, in that example, no exception has been detected.

67. Once the order items are identified, whether by an exceptionless autonomous workflow or by the auditor stepping in, as the case may be, the identified order items are passed to a point of sale (“POS”) relay 62, which transmits the order to the on-site POS system 61 where an on-site employee 81 fulfills the order by delivering the ordered food to the customer 25. EX 1001, 13:53-56; 13:3-6; 17-39.

68. In an upselling subroutine (orange), the identified menu items may also be passed to a business processor 52 that outputs a text prompt (e.g., “would you like fries and a drink with that?”) to a text-to-speech module 42. EX 1001, 13:44-56. The text-to-speech module 41 converts the text prompt into audio that is output to the customer 25 via speakers 22 at the drive-thru order station. EX 1001, 13:44-56.

B. File History of the ‘706 Patent

69. It has been explained to me that the file history of the ‘706 Patent is relevant to understanding the ‘706 Patent. I have reviewed the file history of the ‘706 Patent. EX 1004.

70. I understand that Patent Owner (as applicant) filed a Track One request for accelerated examination. EX 1005. I further understand that, in the Track One request for accelerated examination, Patent Owner identified U.S. Patent Application Publication No. 2003/0018531 (“Mahaffy”) as the “most closely related” prior art to the claimed subject matter. EX 1005, pp. 3-4.

71. I also understand that Patent Owner acknowledged that most of the limitations of the original independent claims were taught by Mahaffy. EX 1005, p. 4. I still further understand that Patent Owner acknowledged that many of the dependent claims had “no features that would define over the references deemed

most closely recited,” if the original independent claims were unpatentable. EX 1005, p. 6.

72. I understand that, despite these acknowledgements by Patent Owner, the USPTO Examiner did not analyze, discuss, or rely on Mahaffy in rejecting the claims, but instead relied on various combinations of other U.S. patents and publications. EX 1006; EX 1008.

73. I further understand that Patent Owner distinguished over the Examiner-cited prior art as failing to teach “deactivating [an automated] customer order processing [by a remote operator] to pass customer control to an on-site employee.” EX 1007, p. 13-15; EX 1008, pp. 8-9.

74. I understand that the Examiner subsequently allowed the claims. EX 1009, pp. 8-9.

75. I have reviewed the rejections issued by the Examiner and the patents and publications relied on by the Examiner in making those rejections.

76. It is my opinion that none of the patents and publications relied on by the Examiner provided as strong an unpatentability ground as Mahaffy alone or in combination with Jasper, as described herein.

C. Effective Filing Date of the ‘706 Patent

77. I understand that the ‘706 Patent issued from U.S. Patent Application No. 16/274,170 (“the ‘170 Application”), which was filed on February 12, 2019. EX 1001.

78. I also understand that the ‘706 Patent claims priority to U.S. Patent Application No. 15/723,573 (“the ‘573 Application”) (EX 1010), filed on October 3, 2017, which in turn claims priority to U.S. Provisional Patent Application No. 62/478,454 (“the ‘454 Application”) (EX 1011), filed on March 29, 2017. In addition, I understand that the ‘706 Patent separately claims priority to U.S. Provisional Patent Application No. 62/633,240 (“the ‘240 Application”) (EX 1012), filed on February 21, 2018.

79. In evaluating the patentability of the challenged claims pursuant to Ground 1, I have been asked to assume that the Patent Owner is entitled to a March 29, 2017 priority date. Accordingly, my evaluation of the patentability of the challenged claims pursuant to Ground 1 is from the perspective of a POSA prior to March 29, 2017.

80. In evaluating the patentability of the challenged claims pursuant to Grounds 2A and 2B, I have been asked to assume that the Patent Owner is not entitled to priority before the filing of the ‘170 Application. Accordingly, my

evaluation of the patentability of the challenged claims pursuant to Grounds 2A and 2B is from the perspective of a POSA prior to February 12, 2019.

81. I have also been asked to consider whether there is continuity of disclosure between: (a) the ‘706 Patent and the ‘573 Application; (b) the ‘706 Patent and the ‘454 Application; and (c) the ‘706 Patent and the ‘240 Application.

82. I have been informed that, for continuity of disclosure to exist, the claims of the ‘706 Patent must find adequate written description support in the disclosure of the earlier filed patent application under consideration. I have further been informed that this analysis is to be on a claim-by-claim basis. That is, in order for continuity of disclosure to exist for a claim of the ‘706 Patent, the subject matter of that claim must find adequate written description support in the earlier filed patent application under consideration.

83. I further have been informed that adequate written description support exists where the subject matter of the claim is described in the earlier patent application in sufficient detail that a POSA can reasonably conclude that the inventor had possession of that subject matter as of the filing of the earlier patent application.

84. In my opinion, independent claims 1 and 4 of the ‘706 Patent are directed to methods for automated order-taking that involve a *human auditor* intervening in an ongoing order taking process. For example, claim 1 recites “alerting an auditor of detected exceptions in the order,” and “deactivating the order

processor with the auditor in response to the detected exceptions in the order.” EX 1001, Claim 1. Claim 4 similarly recites “providing an audio stream of an order ... to an auditor, the auditor located off-site,” “providing the generated order to the auditor,” “performing a comparison of the audio stream ... with the generated order by the auditor,” “errors detected in the order by the auditor,” and “deactivating the order processor with the auditor.” EX 1001, Claim 4.

85. Consistent with my opinion, the ‘706 Patent describes the *auditor* as someone who is connected to the artificially intelligent order processing system in order to ensure the fidelity of the order process and the recognition of speech by the artificially intelligent order processing system. EX 1001, 12:46-56. The ‘706 Patent explains that, “[i]f the auditor 71 detects that the artificially intelligent order processing system 10 is not functioning properly, or that the customer 25 is frustrated or otherwise not being adequately served by the artificially intelligent order processing system 10, then the auditor 71 can abort, or intervene in, the transaction” EX 1001, 13:33-40. Moreover, the figures of the ‘706 Patent depict the auditor (71) as a human, not as a computer. EX1001, Figs. 3, 4, 6. The specification of the ‘706 Patent further refers to the auditor as “who,” not “what”. EX1001, 13:13-14.

86. In my opinion, none of the parent applications describe this *human auditor* intervening in the order taking process, or anything reasonably similar.

87. The '573 Application describes a computer-generated conversational holographic assistant capable of being used in a fast food restaurant to take customer orders. EX 1010. The human auditor of the '706 Patent is not described. EX 1013 is a redline comparison of the text of Specification of the '706 Patent and that of the '573 Application, which illustrates the disparate disclosures.

88. The '454 Application describes a computer-generated holographic assistant capable of being used in a fast food restaurant to take customer orders. EX 1011. The human auditor of the '706 Patent is not described. EX 1014 is a redline comparison of the text of Specification of the '706 Patent and that of the '454 Application, which illustrates the disparate disclosures.

89. The '240 Application describes methods for improving speech-to-text generation accuracy using video-speech matching. EX 1012. The human auditor of the '706 Patent is not described. EX 1013 is a redline comparison of the text of Specification of the '706 Patent and that of the '240 Application, which illustrates the disparate disclosures.

90. Based on the foregoing, it is my opinion that a POSA cannot reasonably conclude that the inventor of the '706 Patent had possession of all the subject matter of independent claims 1 and 4 as of: (a) the filing date of the '573 Application, (b) the filing date of the '454 Application, or (c) the filing date of the '240 Application.

91. Accordingly, the subject matter of claims 1 and 4 of the ‘706 Patent lacks adequate written support in (a) the ‘573 Application, (b) the ‘454 Application, or (c) the ‘240 Application. Continuity of disclosure is therefore lacking between: (a) the ‘706 Patent and the ‘573 Application; (b) the ‘706 Patent and the ‘454 Application; and (c) the ‘706 Patent and the ‘240 Application.

VI. CHALLENGED CLAIMS AND CLAIM CONSTRUCTIONS

92. I have been asked to review claims 1-11 of the ‘760 Patent. I discuss the specific limitations of each of those claims in Sections IX to XI below, where I discuss my analysis of the claims compared to the prior art.

93. In order to properly evaluate these claims, I understand that the terms of the claims must first be construed. For the purposes of this declaration, I have applied the plain and ordinary meaning of each of the terms in claims of the ‘706 Patent as they would have been understood by a POSA as of the critical date. In my opinion, these plain and ordinary meanings are the same regardless of whether their meanings are considered in 2017 (for Ground 1) or 2019 (for Grounds 2A and 2B). Nevertheless, I reserve my right to offer my opinions under any alternative construction proposed by the Patent Owner or adopted by the Board.

VII. QUALIFICATIONS OF A PERSON OF ORDINARY SKILL

94. I have reviewed the ‘706 Patent, those references cited in the ‘706 Patent, as well as the prior art documents relied on in this proceeding. Based on this

review and my knowledge of automated speech-recognition (“ASR”) and natural language processing (“NLP”) systems from the relevant time period, including my work on such systems, it is my opinion that a POSA would have had at least a bachelor’s degree in computer science or computer engineering and at least 2 years of experience with speech-recognition and natural language processing technology. It is my further opinion that a POSA would also have a general knowledge of the fast-food ordering process.

95. I understand that this determination is made at the time of the invention, which I have been asked to assume is around March 29, 2017 for Ground 1 and around February 12, 2019 for Grounds 2A-2B.

96. As also discussed above, I am familiar with the level of knowledge and the abilities of a POSA at the time of the claimed invention under either date. Based on my qualifications and experience, it is my opinion that I meet this level of knowledge and skill presently, and did so in at least March of 2017 and February 2019.

VIII. OVERVIEW OF THE PRIOR ART

A. Mahaffy – U.S. Pat. Appl. Pub. No. 2003/0018531

97. I understand that U.S. Patent Application Publication No. 2003/0018531 by Mahaffy (“Mahaffy”) was published on January 23, 2003. EX

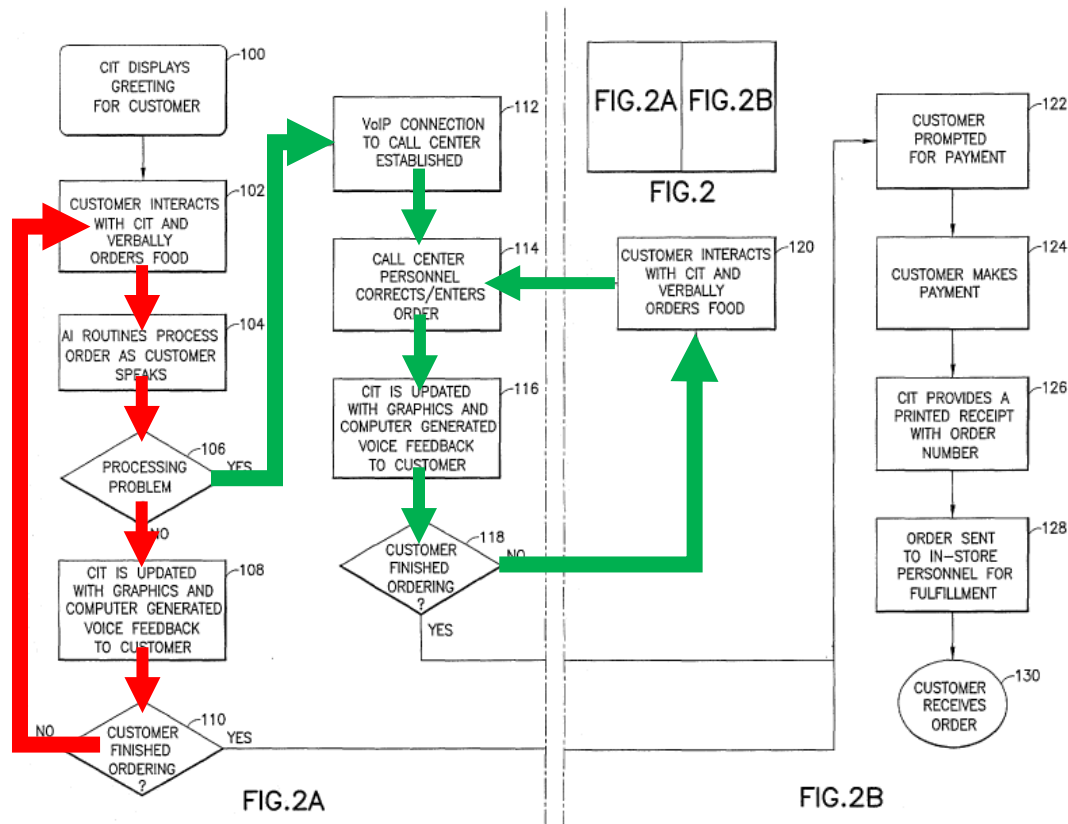
1020. It has been explained to me that Mahaffy is considered prior art, since it was published before the earliest effective filing date of the '706 Patent.

98. Mahaffy generally relates to a system and method for automating the traditional fast-food voice ordering process. EX 1020, ¶¶[0015]-[0018].

99. Mahaffy describes the traditional non-automated fast-food operation model as involving a point-of-sale (“POS”) register operated by a cashier behind the counter to enter a customer’s order into the register. The order is then communicated to employees who prepare and assemble the order. EX 1020, ¶[0005]. The customer provides payment to the cashier, and the food order is delivered to the customer. EX 1020, ¶[0005].

100. Mahaffy describes replacing the on-site employee with a customer interaction terminal (“CIT”) that interacts with the customers to take voice orders. EX 1020, ¶¶[0016]-[0021]; [0030]-[0033]; [0050]. The CIT 12 is coupled to a computer system 14 running software that permits the CIT 12 to interact with the customer and process, via artificial intelligence (“AI”) routines, customer orders spoken into the microphone 22 of the CIT 12. EX 1020, ¶¶[0030]-[0033]; [0044]; [0050].

101. FIGS. 2A-2B, reproduced below with annotations, illustrate aspects of the ordering process described by Mahaffy.



102. The red arrows on the left identify an automated workflow. In the automated workflow, when the customer order is verbally provided to the CIT 12, the AI routines utilize speech recognition and natural language processing to identify menu items and add them to an order list. EX 1020, ¶¶[0030]-[0033]; [0044]; [0050].

103. When the AI routines are unable to satisfactorily complete a fast food order, remotely located human operators can intervene to complete, correct, or verify the order via remote operation of the CIT. EX 1020, FIGS. 2-3; ¶¶ [0018]. This intervention by the remotely located human operators is shown by the green arrows in the center of the figure.

B. Jasper – U.S. Pat. Appl. Pub. No. 2007/0007331

104. I understand that U.S. Patent Application Publication No. 2007/0007331 by Jasper (“Jasper”) was published on January 11, 2007. EX 1021. It has been explained to me that Jasper is considered prior art, since it was published before the earliest effective filing date of the ‘706 Patent.

105. Jasper generally relates to a system and method for outsourcing the traditional fast-food voice order taking process. Jasper describes a drive-thru point-of-sale (“POS”) location 70 where customers are connected, via a VoIP connection, with a remote order taking agent 12a (e.g., a human at a call center or an AI) that takes the customers’ verbally submitted orders. EX 1021, ¶¶[0045]; [0047]; [0054]-[0055]; FIG. 1. In failure situations, an alternate remote agent 12a may be selected (thus switching from AI to human, or vice versa) to step in and take the order, or the system may defer to the personnel at the store to take the order via a restaurant point-of-sale system (“RPOS”) 30. EX 1021, ¶¶[0059]-[0062]. Jasper also describes that the remote order taking can also be selectively used to handle overflow orders, such as those orders which exceed the restaurant’s or chain’s own order taking capacity, e.g., lunch or dinner rush or during staffing shortages. EX 1021, ¶[0063].

C. Coleman – U.S. Pat. Appl. Pub. No. 2019/0108566

106. I understand that U.S. Patent Application Publication No. 2019/0108566 by Coleman (“Coleman”) was effectively filed on October 5, 2017

and was published on April 11, 2019. EX 1022. It has been explained to me that Coleman is considered prior art, since it was effectively filed before the permissible effective filing date of the '706 Patent.

107. Coleman generally relates to a system for automated drive-thru ordering. FIG. 1, reproduced below with annotations, illustrates aspects of the system described by Coleman.

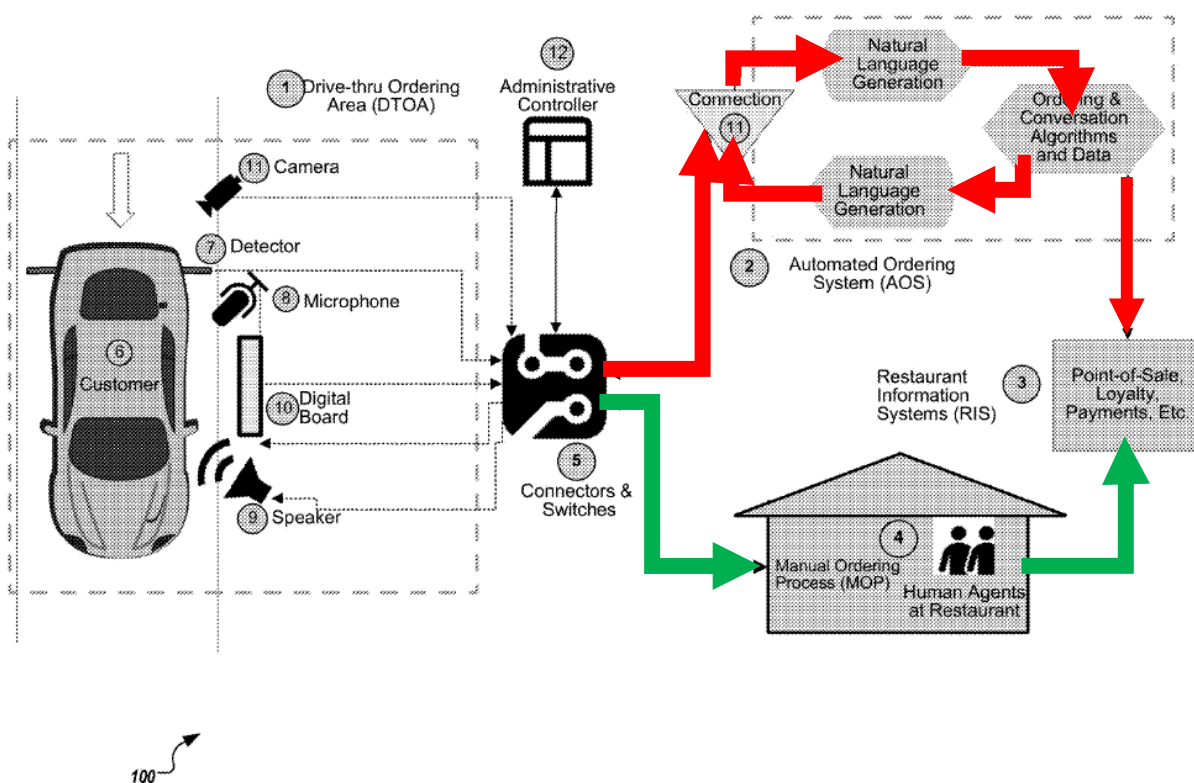


FIG. 1

EX 1022, FIG. 1.

108. Coleman describes the system 100 as including a drive-thru ordering area 1 that has a microphone 8 into which a customer 6 speaks an order, which is

entered into a Restaurant Information System (“RIS”) 3 for processing. EX 1022, ¶¶[0029]; [0034]-[0036]; [0080].

109. In an automated workflow (red arrows top right of figure), an automated order system (AOS) 2 uses a natural language understanding (NLU) component to convert the spoken order into transcribed text and intent so as to process the customer’s order. EX 1022, ¶¶[0034]-[0035]; [0070]-[0071]. In a human-assisted workflow (green arrows bottom right of figure), the customer’s spoken order is processed by a manual order process system (“MOP”) 4, in which human agents take the order via interfacing with the customer using, e.g., headsets. EX 1022, ¶¶[0035]-[0036]; [0068]. Coleman describes rerouting the voice ordering process from the AOS to the MOP if the automated workflow is not satisfactorily progressing to a completed order. EX 1022, ¶¶ [0021]-0024]; [0036]; [0071]-[0077].

D. Washington – U.S. Pat. Appl. Pub. No. 2017/0293905

110. I understand that U.S. Patent Application Publication No. 2017/0293905 by Washington (“Washington”) was effectively filed on April 8, 2016. EX 1023. It has been explained to me that Washington is considered prior art, since it was effectively filed before the earliest effective filing date of the ‘706 Patent.

111. Washington generally relates to an order processing system 100 that facilitates the order-taking process at a drive-thru restaurant via a speech-to-order (“STO”) system 106. EX 1023, ¶¶[0021]-[0022]. The speech-to-order system 106 converts customer speech into an order that is displayed to an operator. EX 1023, ¶¶[0030]-[0037]; [0039]-[0045]; [0068]-[0073]. The operator also listens to the customer speech and can confirm or make corrections to the order. EX 1023, ¶¶[0038]; [0068]-[0070]; [0020].

IX. GROUND 1 – CLAIMS 1-11 ARE OBVIOUS OVER MAHAFFY IN VIEW OF JASPER

A. Motivation to Combine

112. It is my opinion that a POSA considering Mahaffy would have looked to Jasper to improve the ordering process of Mahaffy.

113. More particularly, both Mahaffy and Jasper describe improvements in the conventional quick-service (i.e., fast-food) ordering process where customers can either order at the counter or in a drive-thru. When at the counter, in both references, customers speak their orders to face-to-face on-site employees. *E.g.*, EX 1020, ¶[0005]; EX 1021, ¶[0003]. When at the drive-thru, in both references, customers speak into an order box that is set up with a microphone and speaker to allow two-way communication with the on-site employee. *E.g.*, EX 1021, ¶[0005]. Either way, in both references, the on-site employee enters the customer’s order into a point-of-sale (POS) terminal, such as an integrated order input/cash register. The

customer's order is then transmitted from the POS terminal to the kitchen staff in both references. The kitchen staff fulfills the order while the cashier (e.g., on-site employee) accepts payment from the customer in both references. *E.g.*, EX 1020, ¶[0005]; EX 1021 ¶[0003].

114. Mahaffy and Jasper each recognize that “delays and inefficiencies in order intake and fulfillment” can “become[] the critical bottleneck in the overall workflow,” which “lead[s] to reduced production, long customer queues and ultimately customer dissatisfaction.” EX 1021, ¶[0010]; *see also* EX 1020, ¶[0006] (“The primary bottleneck to serving greater numbers of customers is in the processing of the orders,” and “profits for a fast food restaurant can be increased by decreasing the transaction time for the orders, such that more orders can be entered in a given time frame.”). Each also recognizes that merely having more on-site employees taking orders is not a viable solution to the bottleneck problem, whether because of a lack of on-site space (EX 1020, ¶[0006]) or because of a lack of local labor (EX 1021, ¶[0006]).

115. Mahaffy's solution to the bottleneck problem is to automate the ordering process with a customer-interaction-terminal (CIT) equipped with speech-recognition AI routines to take the customer orders autonomously. EX 1020, ¶¶[0003], [0015]-[0018], [0021], [0030]-[0033], [0044], [0050]. This solution effectively provides more order takers (i.e., AI-based CITs) than the on-site space or

labor pool otherwise allows. Thus, the number of behind-the-counter cashiers may be reduced, providing additional room for order fulfilment. *Id.*, ¶ [0011].

116. Still, Mahaffy acknowledges shortcomings in its solution where the AI may fail. EX 1020, ¶¶[0018], [0034], [0039], [0047], [0050]. The problem of AI failure is also recognized by others in the art. See EX 1022, ¶[0021] (explaining that, while NLP services continue to improve, customer voice level, speech patterns, accents, etc., may preclude the automated system from smoothly completing the ordering process). Moreover, as discussed in Section IV.C. above, the problem of automation failing is as old as automation itself.

117. Accordingly, as a fallback, Mahaffy reverts to a call-center-based approach with human operators at a call site intervening to take over the order processing from the AI routines. EX 1020, ¶¶[0018], [0039], [0047], [0050].

118. Similar to Mahaffy, Jasper also uses a call-center-based approach to address the bottleneck problem. In Jasper, a VoIP connection is established between at least some of its customers at the drive-thru and specialized order-taking agents at a call center who take customer orders. EX 1021, ¶¶[0022], [0045], [0047], [0054]-[0055], FIG. 1.

119. Jasper, however, identifies a problem with the call-center based approach not explicitly recognized in Mahaffy but which is just as likely in Mahaffy, in that the customer-agent connection may fail, which can waste time and lead to

customer frustration. EX 1021, ¶[0059]. Jasper also identifies a solution to this problem: if the connections between the remote order-taking agent and the customer fail for any reason, the order-taking responsibility is rerouted to on-site personnel, who can take the customer's order via the conventional process known for decades. EX 1021, ¶[0059].

120. A POSA would have found it obvious to modify Mahaffy to include the same failsafe measures of Jasper for its own call-center-based connections—with the motivation coming directly from Jasper.

121. More particularly, while Mahaffy does not recognize the potential for connection failures in its call-center failsafe, Mahaffy does not disclose any reason its call center would be less prone to failures than the call center used in Jasper, leading to a similar waste of time and customer frustration to that noted by Jasper. Because Jasper identifies an existing problem in an aspect of Mahaffy, as well as that problem's solution, and because the ultimate goals of any fast-food ordering system are to decrease ordering time and increase customer satisfaction, a POSA would be motivated by Jasper to similarly address the same problem in Mahaffy, and to thereby similarly avoid the resulting wasted time and customer frustration. Moreover, while Mahaffy suggests reducing on-site cashiers, Mahaffy contemplates some behind-the-counter cashiers remain on-site. Thus, applying Jasper's revert-to-

the-on-site-employee failsafe measure to Mahaffy is little more than deferring to the existing on-site workflow when the off-site outsourcing fails.

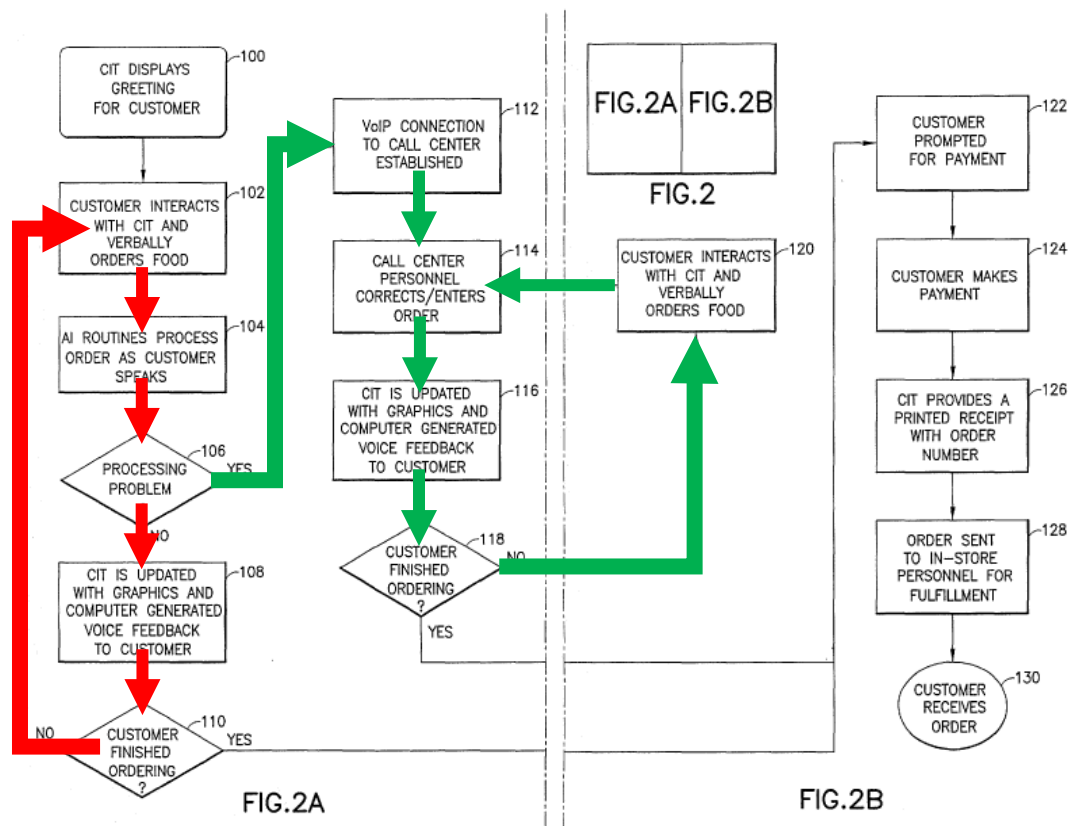
122. A POSA would have a reasonable expectation of success in adding Jasper's revert-to-the-on-site-employee failsafe measure to Mahaffy, because a POSA would recognize that it would merely require application of known techniques from Jasper to improve a similar system from Mahaffy in the same way by combining known elements with predictable results.

B. Claim 1

[1-Pre] – A method comprising the steps of:

123. I understand that the preamble of a claim may or may not be considered a limitation of the claim. Regardless of whether the preamble is determined to be a limitation of this claim, it is my opinion that Mahaffy describes to a POSA the subject matter of the preamble.

124. In particular, with reference to FIGS. 2A-2B, reproduced below with annotations, Mahaffy describes a method for automating the traditional fast-food voice ordering process. EX 1020, ¶¶[0015]-[0018].



125. The red arrows identify an automated workflow, in which a customer order is verbally provided to a CIT 12, and AI routines utilize speech recognition and natural language processing to identify menu items and add them to an order list. EX 1020, ¶¶ [0030]-[0033]; [0044]; [0050]. When the AI routines are unable to satisfactorily complete a fast food order, remotely located human operators can intervene to complete, correct, or verify the order via remote operation of the CIT 12 in a human-operated workflow (i.e., the green arrows). EX 1020, FIGS. 2-3; ¶¶ [0018].

126. I provide further details concerning the method introduced by the preamble in the sections that follow.

*[1a] – providing an audio stream of an order
of a customer positioned on a site*

127. Mahaffy discloses “providing an audio stream of an order of a customer,” in the form of “a customer order [that] is verbally provided into the microphone 22 of the CIT at 102.” EX 1020, ¶[0033].

128. The “sounds (words) spoken into the microphone 24 of the CIT and/or recorded by the computer system 14” are reproduced for the operator 40 and therefore form an audio stream. EX 1020, ¶[0031].

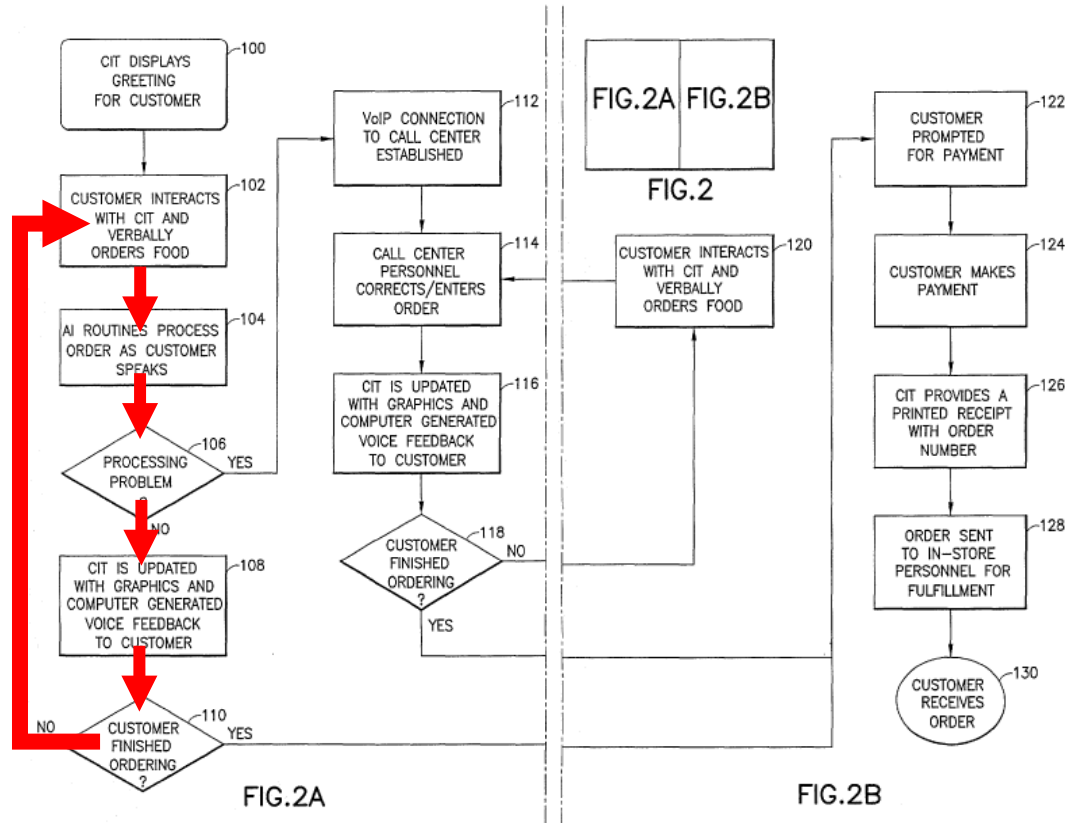
129. The customer is “positioned on site” because the CIT 12 “may be placed ... along a take-out drive-through route ... or in any other suitable location within or relative to a fast food restaurant enabling customer interaction with the CIT.” EX 1020, ¶[0029].

130. It is therefore my opinion that the limitations of [1a] are disclosed by Mahaffy.

*[1b] – providing an order processor having
a speech recognition module trained using
artificial intelligence programs*

131. Mahaffy discloses an “order processor having a speech recognition module” in the form of a computer system 14 with artificial intelligence (AI) routines that parse customers’ utterances to determine what the customer wants to order. EX 1020, ¶[0033].

132. More specifically, the CIT 12 is coupled to the computer system 14 running software that permits the CIT 12 to interact with the customer and process, via the AI routines, customer orders spoken into the microphone 22 of the CIT 12. EX1020, ¶¶[0030]; [0033]. In the automated workflow (red), shown below in annotated FIGS. 2A-2B, when the customer order is verbally provided, the AI routines utilize speech recognition and natural language processing to identify menu items and add them to an order list:



EX 1020, FIGS. 2A-2B, ¶¶[0034], [0037]-[0038], [0040],[0044].

133. In the automated workflow of Mahaffy, the AI routines are embodied in software of the computer system 14, which is “adapted to permit each CIT 12 to

‘interact’ with a customer and process ... customer orders spoken into the microphone.” EX 1020, ¶[0030]. These AI routines “recognize and parse the speech of the customer ... such that individual menu items ordered are properly added to the customer’s order list ... after each menu item is ordered.” EX1020, ¶[0044]. Mahaffy provides the example of the AI routines recognizing the words “three” and “SuperBurgers” from the customer’s utterance, which the AI routines interpret as an order for three SuperBurgers. EX 1020, ¶[0040].

134. A POSA would understand that Mahaffy’s description of the AI routines recognizing and parsing the customer’s verbal order so as to add recognized menu items to the customer’s order list describes the application of speech-recognition or ASR to the customer’s verbal order. Mahaffy indeed describes several speech-recognition systems in connection with describing its recognizing and parsing of the customer’s verbal order. EX 1020, ¶ [0037]-[0038]. For example, Mahaffy identifies Dragon’s Naturally Speaking software as exemplary speech-to-text software for its system. EX 1020, ¶[0038]; EX 1058.

135. A POSA would also understand that the speech recognition module of Mahaffy would be “trained using artificial intelligence.” More particularly, the ‘706 Patent describes training its speech-to-text module as providing sample audio and corresponding human-transcribed text to a neural network so as to build the pattern recognition program that is the speech-to-text module. EX 1001, 9:40-11:33. This

is what is known as a ‘supervised learning’ approach, where the training data consists of labeled examples, e.g., audio is labeled with its corresponding phonemic transcription.

136. Mahaffy similarly describes using known speech-to-text software that is also trained using, at a high level, the same supervised learning approach. EX 1020, ¶[0038]. In particular, Mahaffy identifies Dragon’s Naturally Speaking software as exemplary speech-to-text software. EX 1020, ¶[0038]. Dragon’s Naturally Speaking is speech-to-text software that is trained to recognize user speech by comparing a corpus of text words with audio of a user speaking those words, so as to train the speech-to-text software to recognize user speech. EX 1058, pp. 21-24; EX 1060. In particular, training the software involves recording a user reading known text. EX 1058, pp. 21-24. A POSA would understand that this indicates the Dragon’s Naturally Speaking software also follows a supervised learning approach, in which the recorded audio is compared to the known text so as to generate the labeled training data, i.e., audio labeled in accordance with the text

137. In the context of generating labeled training data for training speech-to-text software, there is no meaningful difference between recording a user reading known text and manually transcribing a user’s recorded speech, because both result in a corpus of known text words and corresponding audio of a user speaking those words.

138. Also, as noted above, the ‘706 Patent identifies its speech-to-text module, that is “trained using artificial intelligence,” as Dragon Dictation software. EX 1001, 7:37-41. The Dragon Dictation software is known to be based on the Dragon Naturally Speaking software, and in relevant respects utilizes the same speech-to-text technology. EX 1059; EX 1060. A POSA would thus be aware that, in relevant respects, the two software products utilize the same approach to ASR. A POSA would therefore understand that the speech recognition module of Mahaffy (i.e., Dragon’s Naturally Speaking) would be similarly “trained using artificial intelligence” as the Dragon Dictation software identified by the ‘706 Patent.

139. Mahaffy also describes that its AI routines are optimized based on conditioning data collected at the POS, which conditioning data is used to train the AI routines on a continuous and regular basis. EX1020, ¶¶ [0034]; [0037]. A POSA would understand this as referring to training using the supervised learning approach discussed above, and consequently that the AI routines are “trained using artificial intelligence,” for the same reasons.

140. In addition, at a basic level, a POSA would understand that most if not all ASR systems at the time, including the one of Mahaffy, would be developed by providing corpus of sample audio that has been assigned text words or phrases to a neural network to train the neural network’s ability to pattern match against incoming audio, as described in Section IV.A. above.

141. In particular, as explained in Section IV.A. above, ASR neural networks at the time were trained, at a high level, by manually transcribing speech into words and phrases, and assigning those words and phrases to the ASR determined phoneme sequences of the speech, in order to build a corpus of known words and phrases. The training also generally involved speaking to the ASR and correcting mistakes in the ASR's transcription of the speech in a training feedback loop.

142. Where corrections are made, the corrected phoneme sequence to known word association is stored as a reference for the neural network to pattern match against (i.e., the training dataset is updated). Where corrections are not made, the phoneme sequence to known word association is reinforced. A POSA would be readily familiar with the fact that most, if not all, ASR systems at the time were trained according to this general approach.

143. This is also precisely how the '706 Patent describes its speech-to-text module 41 is trained. EX1001, 9:40, et seq. Sample audio and its assigned human-transcribed text is provided to a neural network so as to build the pattern recognition program that is the speech-to-text module. EX1001, 11:14-33. There is no difference between this and what a POSA would understand from Mahaffy's teachings.

144. Indeed, both the Dragon Naturally Speaking and the Dragon Dictation software programs were known to utilize this same basic ASR training process. EX 1058; EX 1059; EX 1060.

145. A POSA would therefore understand that Mahaffy's AI routines are trained in accordance with the same basic ASR training process, and are thus "trained using artificial intelligence."

146. It is therefore my opinion that the limitations of [1b] are disclosed by Mahaffy.

[1c] – converting a word or words in the audio stream to text using the speech recognition module

147. As discussed in [1b] above, Mahaffy discloses a "speech recognition model" in the form of the AI routines.

148. As also discussed in [1b] above, Mahaffy further discloses that its speech recognition model "converts a word or words in the audio stream to text" by its AI routines that "recognize and parse the speech of the customer ... such that individual menu items ordered are properly added to the customer's order list ... after each menu item is ordered." EX 1020, ¶¶[0044]; [0040].

149. For the reasons discussed in [1b] above, a POSA would understand that Mahaffy's description of the AI routines recognizing and parsing the customer's verbal order so as to add recognized menu items to the customer's order list is a

description of the AI routines applying speech-recognition to the customer verbal order in order to recognize menu items from the customer's verbal order.

150. A POSA would further understand that Mahaffy's AI routines for recognizing a menu item from a customer's speech would involve speech-to-text conversion of the audio stream to text. As discussed in [1b], Mahaffy indeed suggests using a known speech-to-text software (*e.g.*, Dragon's NaturallySpeaking) that was known to convert words in the audio stream to text. EX 1020, ¶[0038]; EX 1058; EX 1059.

151. Additionally, as explained in Section IV.A. above, a POSA would understand that most if not all speech-recognition routines at the time, including the AI routines of Mahaffy, work by converting incoming an audio stream into a text-based phoneme sequence (the sequence of sound units, *e.g.*, *th*, *p*, *d*, etc.) for pattern matching against known words or word sequences. For example, the audio "three SuperBurgers" may be converted to the phoneme sequence "th-r-ee--s-oo-p-u-r--b-u-r-g-u-r", which may be pattern-matched against known words to generate the text "three super burgers." Once identified, the known words or word sequences are pattern matched to known intents or meanings. For example, the text "three super burgers" may be pattern matched to the known intent of adding three SuperBurgers to the order.

152. There is no difference between this and what a POSA would understand from Mahaffy's teachings.

153. It is therefore my opinion that the limitations of [1c] are disclosed by Mahaffy.

[1d] – processing the text communication with the speech recognition module to identify a word or words in the text of the converted audio stream according to a previous spoken word training

154. As discussed in [1b] and [1c] above, Mahaffy discloses a “speech recognition module” and “text of the converted audio stream.”

155. Mahaffy discloses the remainder of this limitation in the form of its AI routines that “recognize and parse the speech of the customer ... such that individual menu items ordered are properly added to the customer's order list ... after each menu item is ordered.” EX 1020, ¶¶[0044]; [0040].

156. Mahaffy also discloses the use of “spoken-word training” by explaining that its parsing of sentences uses grammar, preferably learned automatically from a corpus of utterances parsed manually. EX 1020, ¶[0042]. Mahaffy further discloses that the “AI routines are preferably optimized based on conditioning data collected from conversation over a period of time AI routine training and optimization is preferably performed on a continual basis” EX1020, ¶[[0034].

157. A POSA would understand that this constitutes processing text communication according to a previous spoken word training.

158. As discussed in Section IV.A. and in [1b] above, a POSA would understand that the neural networks of speech-recognition modules are trained according to spoken word training to perform the pattern matching. That is, the neural networks “trained” by being provided with a corpus of sample audio of words and phrases that are manually parsed into their phonemes against which the neural networks perform the pattern matching.

159. The ‘706 Patent similarly describes the training of its speech-to-text 41 module as sample audio and its corresponding human-transcribed text being provided to a neural network so as “to build a pattern recognition program.” EX 1001, 11:14-33. There is no difference between this and what a POSA would understand from Mahaffy’s teachings.

160. It is therefore my opinion that the limitations of [1d] are disclosed by Mahaffy.

[1e] – providing a natural language processor having order assembly capabilities and exception detection capabilities, wherein the natural language processor receives recognized text from the order processor and creates or modifies an order based upon the recognized text

[1f] – generating an order with the natural language processor

161. Mahaffy discloses a “natural language processor having order assembly capabilities” by the computer system 14 with AI routines that parse customers’ utterances to determine what the customer wants to order. EX 1020, ¶¶[0033]; [0050].

162. Mahaffy further discloses “order assembly capabilities,” “creates or modifies an order,” and “generating an order” because “the AI routines are able to recognize and parse the speech of the customer ... such that individual menu items ordered are properly added to the customer’s order list (the transaction) after each menu item is ordered.” EX 1020, ¶[0044]; [0050].

163. A POSA would understand that Mahaffy’s AI routine adding a recognized menu item from a customer’s speech to the customer’s order list would involve “receiv[ing] recognized text” corresponding to the menu item.

164. In particular, Mahaffy discloses that “as the AI routines process the words of the customer’s order, the task is to recognize the customer’s request at a level sufficient to process his order,” e.g., by recognizing menu items such as “three SuperBurgers,” which “analysis is the direct result of the speech recognition.” EX 1020, ¶[0040].

165. Moreover, as discussed in Section IV.A. above, a POSA would understand that most if not all speech-recognition at the time, including the module of Mahaffy, works by converting incoming an audio stream into a text-based

phoneme sequence (the sequence of sound units, e.g., *th*, *p*, *d*, etc.) for pattern matching against known words or word sequences. For example, the audio “three SuperBurgers” may be converted to the phoneme sequence “th-r-ee--s-oo-p-u-r--b-u-r-g-u-r”, which may be pattern-matched against known words to generate the text “three super burgers.”

166. Once identified, the known words or word sequences are pattern matched to known intents or meanings. For example, the text “three super burgers” may be pattern matched to the known intent of adding three SuperBurgers to the order. There is no difference between this and what a POSA would understand from Mahaffy’s teachings.

167. Mahaffy further discloses “exception detection capabilities” in the form of “the ability of the AI routines of the computer system 14 to recognize when the result of the speech recognition is correct and, when it is incorrect.” EX 1020, ¶[0039]. Mahaffy explains that “all current speech recognition systems use an underlying probabilistic model [that] estimates how likely [the] particular word sequence is to have generated the acoustic signal heard.” EX 1020, ¶[0047]. Mahaffy indeed describes “problem[s] with the order processing ... preferably determined via a probabilistic calculation”. EX 1020, ¶[0047].

168. It is therefore my opinion that the limitations of *[1e]* and *[1f]* are disclosed by Mahaffy.

[1g] – alerting an auditor of detected exceptions in the order, the auditor located off-site and connected to the order processor via a data connection

169. Mahaffy discloses an “auditor located off-site” by “human operators 40 [that] work at the call center,” which is “preferably located on different premises than the CIT.” EX 1020, ¶[0031].

170. Mahaffy also discloses that the human operators are “connected to the order processor via a data connection” or “a network connection,” which may be a VoIP connection, created between the CIT 12 (and the computer system 14) and the call center 16. EX 1020, ¶¶[0047]; [0050]; FIGS. 2A-2B.

171. As discussed in *[1e]* above, Mahaffy also discloses “detected exceptions in the order.”

172. Mahaffy further discloses “alerting an auditor of detected exceptions in the order” in the form of the VoIP connection that is created in response to “a problem with the order processing at 106,” and via which VoIP connection the operator is “able to correct, verify or complete” the customer’s order in the human-assisted workflow. EX 1020, ¶¶[0047]; [0050]; FIGS. 2A-2B.

173. A POSA would understand that the act of establishing the VoIP connection would alert the operator 40 of detected exceptions in the order.

174. More particularly, a VoIP is a multimedia transmission protocol for making audio and/or video ‘phone calls’ over the internet. EX 1061. Thus,

according to the VoIP protocol, the operator 40 would receive and need to answer an incoming VoIP call from the CIT 12 / computer system 14 in order to establish the VoIP connection. Since Mahaffy explains that the VoIP connection is only established in response to the detected exceptions, EX1020, ¶¶[0047]; [0050], the operator would not receive an incoming VoIP call (i.e., the alert) unless exceptions in the order processing were detected. The incoming VoIP call thus necessarily alerts the operator of detected exceptions in the order.

175. It is therefore my opinion that the limitations of *[lg]* are disclosed by Mahaffy.

*[1h] – deactivating the order processor with
the auditor in response to the detected
exceptions in the order*

176. Mahaffy discloses this limitation in the form of the operator completing the order without reverting to the automated workflow, after having switched to the human-assisted workflow. EX 1020, FIGS. 2A-2B; ¶[0047]. Cf., EX 1020, FIGS. 3A-3B; ¶[0050].

177. More particularly, Mahaffy describes switching from the automated workflow to the human-assisted workflow if at any time during the customer's order placement there is an exception detected in the AI routine order processing, i.e., when the AI of the computer system 14 is unable to satisfactorily complete a fast food order. EX 1020, ¶¶[0039]; [0047]. As discussed in *[lg]* above, the human-

assisted workflow is initiated by establishing a VoIP data connection between the CIT 12 / computer system 14 and a human operator 40 located at an off-site call center 16. EX 1020, ¶¶[0047]; [0031]. Moreover, once the VoIP connection is made, the operator 40 may be utilized to complete the order with the customer without reversion to the AI routines. EX 1020, ¶¶[0047]; [0050]; FIGS. 2A-2B.

178. A POSA would understand that completing the order with the human-assisted workflow to the exclusion of the automated-workflow corresponds to “deactivating the order processor” in that the AI routines are not reverted to for the order taking transaction, which Mahaffy discloses as an alternative option. EX 1020, ¶ [0050]; FIGS. 3A-3B.

179. In addition, a POSA would understand that such deactivation of the AI routines occurs naturally “with the auditor,” in the course of establishing the VoIP connection, including the operator’s acceptance of the VoIP connection, and then the operator’s completing of the order without reversion to the AI routines.

180. As discussed in *[Ig]* above, the VoIP connection, which deactivates the automated-workflow in favor of the human-assisted workflow, is established when the operator 40 answers the incoming VoIP call from the CIT 12 / computer system 14. A POSA would therefore understand that the AI routines are deactivated by the human operator 40 answering the VoIP and then completing the order without

reversion to the AI routines, and that such deactivation is “in response to” the detected exceptions in the order.

181. It is therefore my opinion that the limitations of [1h] are disclosed by Mahaffy.

*[1i] – prompting an on-site employee to
engage the customer and complete the order*

182. Mahaffy does not appear to expressly disclose this limitation.

183. However, as discussed in [1g] to [1h], Mahaffy discloses that when the AI of the computer system 14 is unable satisfactorily complete a fast food order, a VoIP data connection is established via which the human operators 40 can intervene “to correct, verify or complete” the order by communicating with the customer through the CIT 12. EX 1020, ¶¶[0047]; [0050].

184. Jasper describes an analogous order-taking system in which customers at a drive-thru POS 70 are connected, via a VoIP connection, with a remote order taking agent 12a (e.g., at a call center). EX 1021, ¶¶[0047]; [0054]-[0055]; FIG. 1. The remote order taking agent 12a can be a human agent or an artificial intelligence based interactive voice response system. EX 1021, ¶[0045]. As the remote order taking agent 12a hears the order components, the agent inputs the order components into the remote agent terminal 42, causing the generation of order data identifying the ordered order components and the transmission of that order data to the store or system interface 80. EX 1021, ¶[0055].

185. Both Mahaffy and Jasper thus describe similar circumstances where VoIP connections are made between on-site customers and remote order taking agents.

186. Jasper further discloses “prompting an on-site employee to engage the customer and complete the order” by a fail-safe mode according to which, in failure situations, the system 1 may select an alternate remote agent 12a to handle the order or defer to the personnel at the store to take the order. EX 2021, ¶¶[0059]; [0061]-[0062] (describing failure situations occurring mid-order). Accordingly, “the order taking responsibility reverts back to the store and employees, with headsets 18, 20 ... [that] allow an employee to establish contact with the customer to receive the customer’s order. The employee may then enter the order into the RPOS system, picking up where the remote order taking agent 12a left off based on the order data transmitted thus far.” EX 1021, ¶[0061].

187. Jasper provides the example that, if the fail-safe mode is activated when two cheeseburgers have already been recorded as ordered, the on-site employee then “determines, from the POS terminal, what has been ordered thus far, and says ‘I have 2 cheeseburgers for you, would you like anything else?’ The customer can then finish the order with the employee.” EX 1021, ¶[0062].

188. Jasper further discloses the “prompting” in the form of the incomplete order data being sent to the POS terminal of the on-site employee for order

completion. EX 1021, ¶[0062]. This would naturally prompt the on-site employee to engage the customer to complete the order. As in the above example described by Jasper, seeing on the POS terminal that 2 cheeseburgers had been ordered thus far would naturally prompt the employee statement/query that engages the customer for order completion.

189. A POSA would have found it obvious to modify Mahaffy in view of Jasper for at least the reasons discussed in Section IX.A. above.

190. More particularly, a POSA would have found it obvious to modify Mahaffy to include the failsafe mode and measures of Jasper so that, in the event the VoIP connection between the call center operator 40 and the CIT 12 fail, the order taking reverts to an on-site employee to complete the order, as suggested by Jasper. A POSA would recognize that this is a simple combination of known elements with predictable results.

191. A POSA would have further found that prompting on-site employees to engage customers and complete orders is little more than reverting to the traditional fast food operational model described in Section IV above. Such a reversion is nothing more than a predictable and natural business response to automation breaking down. Indeed, in such a situation, unless and until the on-site employee is prompted to complete the customer's order manually, the on-site employee would not be aware of the need to complete the order manually.

192. It is therefore my opinion that a POSA would have found the limitations of [1i] obvious based on the combined teachings of Mahaffy and Jasper.

C. Claim 2

wherein the order is processed in real time

193. I understand that Claim 2 depends from Claim 1, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

194. Mahaffy further discloses “the order is processed in real time” by describing that “the customer then interacts at 206 with the CIT 12 in real-time, verbally ordering food.” EX 1020, ¶[0050]. Then, once the order is complete and payment is made, the customer is directed to the order pick-up area where he is provided with the order. EX 1020, ¶[0048].

195. Moreover, Mahaffy discloses that its order processing occurs in the context of the traditional drive-thru process. EX 1020, ¶[0021]. A POSA would understand that drive-thru orders are processed in real time. EX 1040, p. 2; EX 1041, p. 7 (“[drive-thru restaurants] strive fulfill an order in three minutes or less.”).

196. It is therefore my opinion that the limitations of Claim 2 are disclosed by Mahaffy.

D. Claim 3

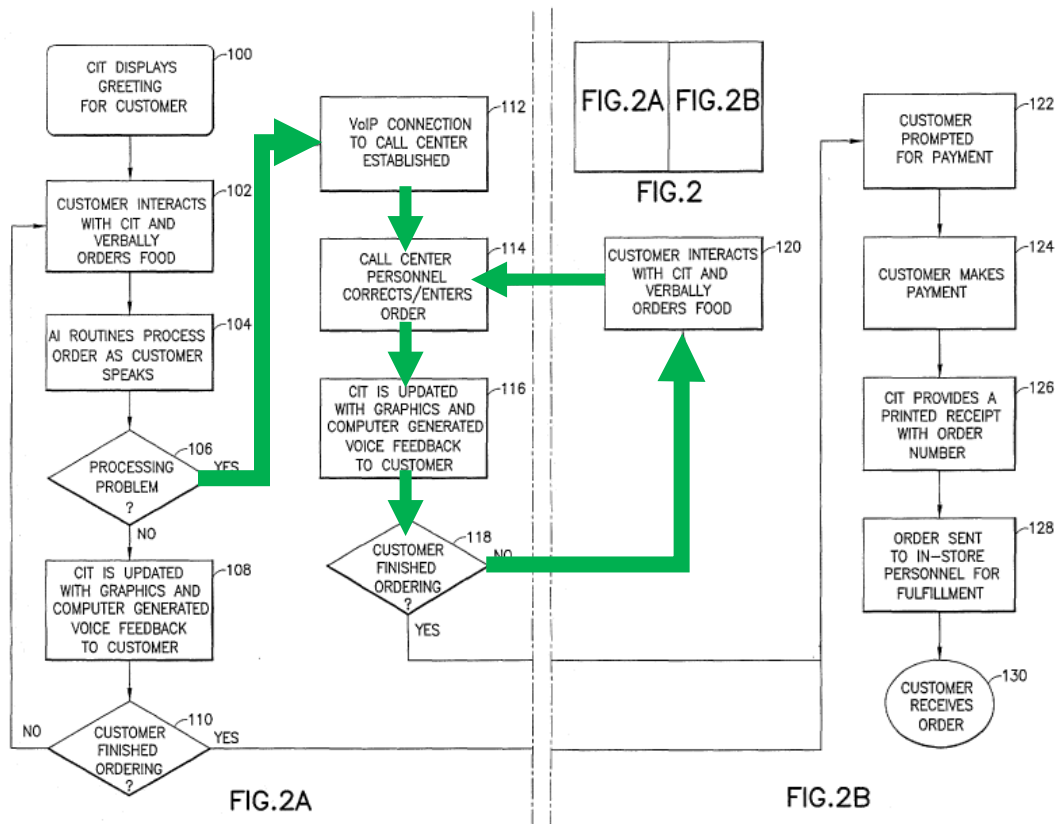
wherein the alert to the auditor prompts the auditor to switch the order processor from an

*autonomous mode to an auditor-assist mode
prior to deactivating the order processor*

197. I understand that Claim 3 depends from Claim 1, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

198. Mahaffy discloses an “autonomous mode” in the form of the automated workflow discussed in [1b], above. In particular, when the customer order is verbally provided, the AI routines utilize speech recognition and natural language processing to identify menu items and add them to an order list. EX 1020, ¶[0044].

199. Mahaffy discloses an “auditor-assist mode” in the form of the human-assisted workflow discussed in [1g] and [1h] above. EX 1020, ¶[0047]. In particular, in response to “a problem with the order processing at 106,” a VoIP connection is created via which an operator is “able to correct, verify or complete” the customer’s order. EX 1020, ¶¶[0047]; [0050]. This human-assisted workflow (green) is shown in annotated FIGS. 2A-2B below:



EX 1020, FIGS. 2A-2B.

200. As further discussed in [1g] above, Mahaffy discloses “the alert to the auditor” by the incoming VoIP call to the operator 40 when an exception is detected. EX 1020, ¶¶[0039]; [0047]. The operator answering the VoIP call establishes the VoIP connection that “switch[es] the order processor from an autonomous mode [(i.e., the automated workflow)] to an auditor-assist mode [(i.e., the human-assisted workflow)],” by permitting the operator to “correct, verify or complete the customer order.” EX 1020, ¶[0047].

201. The incoming VoIP call thus “prompts the auditor to switch the order processor from an autonomous mode to an auditor-assist mode,” because the

operator 40 would receive and need to answer the incoming VoIP call from the CIT 12 / computer system 14 in order to establish the VoIP connection.

202. As discussed in [1h] above, Mahaffy additionally discloses “deactivating the order processor” by, after switching to the human-operated workflow, completing the order processing with the operator “without reversion to the AI routines.” EX 1020, ¶[0047]); FIGS. 2A-2B. Mahaffy also discloses that the automated workflow may continue to process additional items for the order after the operator 40 corrects the order. EX 1020, ¶[0050]; FIGS. 3A-3B. A POSA would understand that these two approaches to completing the order stem from a common event: the operator has intervened to correct the AI routine.

203. It would have therefore been obvious to a POSA to give the operator control as to which approach to follow once the initial correction to the AI generated order is made, namely, to either (a) revert to the automated workflow to complete the order or (b) to complete the order without reverting to the automated workflow. The AI routines would thus not be deactivated until it is decided that the order processing is to be completed without reverting back to the automated workflow.

204. A POSA would recognize that modifying Mahaffy so that the operator controls whether to continue to process the order or to revert back to the AI routines merely combines known elements of Mahaffy to achieve a predictable result. The

predictable result is the ability to either continue to manually process the order or revert back to the AI routines.

205. Moreover, doing so would be nothing more than an obvious design choice, since there are a limited number of possible options. That is, whether to complete the order with the AI or the operator can be either decided before the operator intervention, during the operator intervention, or after the operator intervention. Thus, a POSA would have found it obvious to make the decision during or after the operator intervention.

206. It is therefore my opinion that a POSA would have found the limitations of Claim 3 obvious based on the combined teachings of Mahaffy and Jasper.

E. Claim 4

[4-Pre] – A method comprising the steps of:

[4a] – providing an audio stream of an order of a customer positioned on a site

[4b] – providing an order processor having a speech recognition module trained using artificial intelligence programs

[4c] – converting a word or words in the audio stream to text using the speech recognition module

[4d] – processing the text communication with the speech recognition module to identify a word or words in the text of the converted audio stream according to a previous spoken word training

[4e] – providing a natural language processor having order assembly capabilities and exception detection capabilities, wherein the natural language processor receives recognized text from the order processor and creates or modifies an order based upon the recognized text

[4f] – generating an order with the natural language processor

207. The majority of claim 4 tracks *[1-Pre]* and *[1a]* to *[1f]*, and its limitations are satisfied by Mahaffy for the same reasons as explained above. The differences in Claim 4 are discussed below.

[4g] – providing an audio stream of a customer order to an auditor, the auditor located off-site and connected to the order processor via a data connection

208. Aside from the “providing” clause, *[4g]* tracks *[1g]*, and its limitations are satisfied by Mahaffy for the same reasons as explained above.

209. Mahaffy further discloses “providing an audio stream of the customer order to an auditor” in the form of the VoIP connection “permitting the transmission of the customer’s voice order quickly and inexpensively to an operator 40 at the off-site call center 16. Additionally or alternatively, ... the customer’s recorded verbal order is sent from the memory of the computer system 14 to the audio speaker 42 directed at the operator 40.” EX 1020, ¶[0047].

210. It is therefore my opinion that the limitations of *[4g]* are disclosed by Mahaffy.

[4h] – providing the generated order to the auditor

211. Mahaffy discloses this limitation by describing that “each operator is provided with ... a display 46,” and “the display 46 permits the operator to see the customer’s order” that was generated by the AI routines. Ex. 1020, ¶¶[0031]; [0033]; [0044]; [0050].

212. Mahaffy also explains that, if “there is a problem with the order” generated by the AI routines, “the operator 40 makes the required changes or additions” to the order generated by the AI routines, and “the CIT 12 is updated to indicate the changes and additions.” EX 1020, ¶[0047]. A POSA would understand that the operator would be unable to correct problems with the AI generated order unless the operator is provided with the AI generated order. For example, the operator must be provided with the AI generated order of “1 small french fries” in order to change that order to “1 large french fries,” or otherwise add other items to the order. In other words, the order is a record; in order to change or add to the record, one must have access to it.

213. It is therefore my opinion that the limitations of [4h] are disclosed by Mahaffy.

[4i] – performing a comparison of the audio stream of a customer order with the generated order by the auditor

214. As discussed in [4g] and [4h] above, Mahaffy discloses providing “the audio stream of a customer order,” and providing “the generated order” to the auditor.

215. Mahaffy further discloses “performing a comparison ... by the auditor” in the form of describing that, upon switching to the human-assisted workflow “when there is a problem with the order processing” that necessitates intervention, “the operator 40 is able to correct, verify or complete” the AI routine generated order using “the customer’s recorded verbal order.” EX 1020, ¶¶[0047]; [0050]; FIGS. 2A-2B.

216. A POSA would understand that, in order for the operator 40 to correct or verify the AI routine generated order, the operator 40 would naturally compare the AI routine generated order on the display 42 to the audio stream of the order, so as to identify what corrections need to be made or to verify that none are needed.

217. A POSA would understand that, if the AI routine generated order were not compared to the audio stream of the order, then the operator would not be able to verify that the AI routine generated order is correct, or identify any errors that require correction. For example, without comparing the AI generated order of “1 small french fries” to the customer audio stream of “1 large french fries,” the operator would not know that the AI order needs to be corrected.

218. Since Mahaffy describes that the operator is provided with both the AI generated order and the customer audio stream, and also describes that the operator is able to correct or verify the AI generated order, it naturally follows that such correction or verification involves the operator comparing the AI generated order and the customer audio stream.

219. It is therefore my opinion that the limitations of [4i] are disclosed by Mahaffy.

[4j] – switching to an auditor-assist mode in response to the comparison of the audio stream with the generated order

220. As discussed in [4i] above, Mahaffy discloses “the comparison of the audio stream with the generated order.”

221. Mahaffy further discloses “switching to an auditor-assist mode” by (after switching to the human-operated workflow, *see [1b]*) completing the order processing with the operator “without reversion to the AI routines.” EX 1020, ¶[0047], FIGS. 2A-2B. Mahaffy also discloses that the automated workflow (i.e., order processing via the AI routines) may continue to process the additional items for the order after the operator 40 corrects the order. EX 1020, ¶[0050], FIGS. 3A-3B. See also, Claim 3.

222. A POSA would understand that these two approaches to completing the order (i.e., with or without reversion to the AI routines) each stem from the operator intervening to correct the AI routine.

223. It would have therefore been obvious to a POSA to give the operator control of which approach to follow, once the initial correction to the AI generated order is made. In other words, it would have been obvious to a POSA to give the operator control of whether to either (a) revert to the automated workflow to complete the order or (b) to complete the order without reverting to the automated workflow. Accordingly, “switching to an auditor-assist mode” would occur when the operator decides that the order processing is to be completed without reverting back to the automated workflow, and implements that approach by continuing to complete the order according to the human-operated workflow rather than reverting to the automated workflow.

224. A POSA would recognize that modifying Mahaffy so that the operator controls whether to continue to process the order or to revert back to the AI routines merely combines known elements of Mahaffy to achieve a predictable result. The predictable result is that both of the approaches disclosed by Mahaffy are options for completing the order taking after operator intervention.

225. Moreover, doing so would be nothing more than an obvious design choice, since there are a limited number of possible options. That is, whether to

complete the order with the AI or the operator can be either decided before the operator intervention, during the operator intervention, or after the operator intervention. And it can be made automatically, by the operator, or by some other person. Thus, a POSA would have found it obvious to have the decision made by the operator during or after the operator intervention.

226. It is therefore my opinion that Mahaffy discloses or otherwise renders obvious to a POSA the limitations of [4j].

[4k] – updating the order processor based upon errors detected in the order by the auditor

227. Mahaffy discloses “updating the order processor based on errors,” because, “when there is a problem with the order processing” that necessitates intervention, “the operator 40 is able to correct, verify or complete” the order. EX1020, ¶¶[0047]; [0050]; FIGS. 2A-2B.

228. As discussed in [4i] above, Mahaffy discloses to a POSA that the corrections are based on a comparison of the order audio with the displayed order, which meets the limitation of “based on errors detected in the order by the auditor.”

229. This is precisely how the ‘706 Patent describes its updating of the order processor. If the auditor detects from its comparison of the AI generated order and the customer audio that the AI is not functioning properly, then the auditor can

intervene and update the order. EX 1001, 13:23-40, 4:21-25. There is no difference between this and what a POSA would understand from Mahaffy's teachings.

230. It is therefore my opinion that the limitations of [4k] are disclosed by Mahaffy.

[4l] – deactivating the order processor with the auditor

231. The limitations of [4l] track to [1h] and are satisfied by Mahaffy for the same reasons as explained above.

[4m] – transmitting the generated order to an on-site employee to communicate with the customer, to collect payment from the customer, and to provide notifications to release the ordered product or service to the customer

232. Mahaffy discloses “transmitting the generated order to an on-site employee ... to provide notifications to release the ordered product or service to the customer” by disclosing that “the order is sent at 128 to order fulfillment employees (kitchen staff and order assembly personnel) who prepare the order,” and that “the customer is provided at 130 with the customer's corresponding order.” EX 1020, ¶[0048].

233. Mahaffy does not appear to expressly describe that on-site employees “communicate with the customer” or “collect payment from the customer.”

234. However, Jasper describes an analogous order taking system in which customers at a drive-thru POS 70 are connected, via a VoIP connection, with a remote order taking agent 12a (e.g., at a call center). EX 1021, ¶¶[0047]; [0054]-[0055]; FIG. 1. The remote order taking agent 12a can be a human agent or an artificial intelligence based interactive voice response system. EX 1021, ¶[0045]. As the remote order taking agent 12a hears the order components, the order taking agent 12a inputs the order components into the remote agent terminal 42, causing the generation of order data identifying the ordered order components and the transmission of that order data to the store or system interface 80. EX 1021, ¶[0055].

235. Both Mahaffy and Jasper thus describe similar circumstances where VoIP connections are made between on-site customers and remote order taking agents.

236. Jasper additionally discloses “transmitting the generated order to an on-site employee to communicate with the customer” and “to collect payment from the customer” in the form of describing that the RPOS system 30 (i.e., the restaurant point-of-sale system) “may receive an order from the remote agent terminal 42 ... [and the] RPOS system 30 can transmit the order to at least one POS terminal, which an on-site employee then uses for fulfillment of the order and for collecting payment.” EX 1021, ¶[0057]. Jasper further discloses that the employee

communicates with the customer by “request[ing] the payment from the customer.”
EX 1021, ¶[0058].

237. A POSA would have found it obvious to modify Mahaffy in view of Jasper for at least the reasons discussed in Section IX.A. above.

238. In addition, a POSA would have found it obvious to modify Mahaffy to include the on-site employee functions of Jasper so that the operators of Mahaffy “need only specialize in taking orders and do not need to also function as a cashier,” as suggested by Jasper. EX 1021, ¶[0067]. A POSA would recognize that this is a simple combination of known elements with predictable results. The predictable result would be the efficient division of labor and specializations.

239. A POSA would have further found that “transmitting the generated order to an on-site employee to communicate with the customer” and “collect[ing] payment from the customer” is little more than utilizing the traditional drive-thru operational model, as described in Section IV above, after the order taking portion of the drive-thru process is complete. Thus, a POSA would recognize that the combination would have the additional predictable result of reverting to the traditional drive-thru operational model for communicating with the customer and collecting payment.

240. It is therefore my opinion that a POSA would have found the limitations of [4m] obvious based on the combined teachings of Mahaffy and Jasper.

F. Claim 5

wherein a business processor receives order information from the natural language processor (NLP) and performs the steps of:

associating the order items with one or more additional menu items or options associated with the order items; and

querying the customer about the one or more additional menu items or options associated with the order items, wherein the NLP updates the order based upon the customer response to the query

241. I understand that Claim 5 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

242. As discussed in [1e] above, Mahaffy discloses an NLP in the form of computer system 14 with AI routines are “able to recognize and parse the speech of the customer ... such that individual menu items ordered are properly added to the customer’s order list (the transaction) after each menu item is ordered.” EX 1020, ¶¶[0033]; [0044]; [0050].

243. Mahaffy further discloses a “business processor” in the form of “routines in the computer system which operate the virtual cashier [and] are adapted to follow additional techniques which are shown to increase restaurant sales.” EX 1020, ¶[0046]. Accordingly, the CIT 12 “may prompt the customer via computer-generated voice or displayed text to add other items to the menu list, and a complete

order may require multiple interactions between the customer and virtual cashier 38.” EX 1020, ¶[0046].

244. For example, Mahaffy discloses that “the virtual cashier preferably asks whether the customer would like french fries with an order which does not already include french fries, or whether for a nominal additional sum the customer would prefer to upsize the french fries and drink order, e.g., from medium to large.” EX 1020, ¶[0046].

245. Mahaffy thus discloses the claimed “associating the order items [(e.g., the order that does not already include french fries, or the medium drink)] with one or more additional menu items [(e.g., the offered french fries)] or options associated with the order items [(e.g., the upsize option)]” and also discloses “querying the customer about the one or more additional menu items or options [(e.g., the optional fries or the upsize option)] associated with the order items.”

246. Mahaffy further discloses “the NLP updates the order based upon the customer response to the query” because any “[a]dditional menu items orders are processed at 104 and the CIT is updated at 108 until the customer indicates at 110 that the order is complete.” EX 1020, ¶[0046].

247. Mahaffy also discloses that the sales routines “receive[] order information from the natural language processor” because, in order for the CIT to ask if the customer would like fries with an order that does not already include fries,

or if the customer would prefer to upsize a drink from medium to large, the sales routines of the computer system must be aware of the current order.

248. It is therefore my opinion that the limitations of Claim 5 are disclosed by Mahaffy.

G. Claim 6

*providing a text-to-speech processor to
convert queries from the business processor
into audio to be communicated to a customer*

*processing the text with the text-to-speech
processor to create an audio communication
of the query generated by the business
processor*

249. I understand that Claim 6 depends from Claim 5, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

250. Mahaffy discloses a “text-to-speech processor” by an “automated speech synthesis” on which “voice generation to interact with the customer may be based.” EX 1020, ¶[0044]. A POSA would understand that the automated speech synthesis discussed by Mahaffy refers to text-to-speech processing, because automated speech synthesis refers to the conversion of text into audible speech. See e.g., EX 1057.

251. As discussed in claim 5 above, Mahaffy discloses a “business processor” by “routines in the computer system which operate the virtual cashier

[and] are adapted to follow additional techniques which are shown to increase restaurant sales.” EX 1020, ¶[0046]. Mahaffy further discloses the claimed “convert queries from the business processor into audio to be communicated to the customer” and “create an audio communication of the query generated by the business processor” because the CIT 12 “may prompt the customer via computer-generated voice or displayed text to add other items to the menu list.” EX 1020, ¶[0046]. For example, “the virtual cashier preferably asks whether the customer would like french fries with an order which does not already include french fries, or whether for a nominal additional sum the customer would prefer to upsize the french fries and drink order, e.g., from medium to large.” EX 1020, ¶[0046].

252. It is therefore my opinion that the limitations of Claim 6 are disclosed by Mahaffy.

H. Claim 7

*reviewing an order by the auditor by
comparing the order generated by the
natural language processor to the
communication in the audio stream*

*updating the order based upon auditor
review*

253. I understand that Claim 7 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

254. As discussed in [1e] above, Mahaffy discloses “the order generated by the natural language processor” in the form of the customer’s order list (the transaction), which is generated by the computer system 14 with AI routines that are “able to recognize and parse the speech of the customer ... such that individual menu items ordered are properly added to the customer’s order list (the transaction) after each menu item is ordered.” EX1020, ¶¶[0033]; [0044]; [0050].

255. The remainder of the limitations track those in [4i] and [4k] above, and the limitations are satisfied by Mahaffy for the same reasons as explained above.

256. It is therefore my opinion that the limitations of Claim 7 are disclosed by Mahaffy.

I. Claim 8

*providing a text-to-speech processor to
convert queries from the business processor
into audio to be communicated to a customer*

*processing the text with the text-to-speech
processor to create an audio communication
of the query generated by the business
processor*

257. I understand that Claim 8 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

258. The limitations of Claim 8 track those of Claim 6, and are satisfied by Mahaffy for the same reasons as explained above.

J. Claim 9

alerting the auditor by the natural language processor (NLP) of any exceptions detected by the artificially intelligent order processing system, wherein the auditor can override the artificially intelligent order processing system based upon the detected exception

259. I understand that Claim 9 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

260. As discussed in [1e] above, Mahaffy discloses an NLP having exception-detection capabilities in the form of AI routines that are “able to recognize and parse the speech of the customer ... such that individual menu items ordered are properly added to the customer’s order list (the transaction) after each menu item is ordered,” and are also able to “recognize when the result of the speech recognition is correct and, when it is incorrect.” EX 1020, ¶¶[0033]; [0039]; [0044]; [0050].

261. Moreover, as discussed in [1g] above, Mahaffy further discloses “alerting the auditor ... of any exceptions detected by the artificially intelligent order processing system” by a VoIP connection that is created in response to “a problem with the order processing at 106,” and by which VoIP connection the operator is “able to correct, verify or complete” the customer’s order. EX1020, ¶¶[0047]; [0050].

262. A POSA would understand that the alert is by the AI routines (i.e., the NLP), because the VoIP connection (the alert) is initiated when the AI routines recognize that the result of the speech recognition is incorrect via a probabilistic model. EX 1020, ¶¶[0039]; [0047].

263. Mahaffy further discloses that “the auditor can override the artificially intelligent order processing system based upon the detected exception” because the operator is “able to correct, verify or complete” the customer’s order. EX 1020, ¶¶[0047]; [0050].

264. It is therefore my opinion that the limitations of Claim 9 are disclosed by Mahaffy.

K. Claim 10

wherein the auditor deactivates the order processor in response to a detected change in customer traffic

265. I understand that Claim 10 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

266. As discussed in [1b], [1g], and [4g] above, Mahaffy discloses “the auditor” as human operators 40 that work at the call center, and “the order processor” as a computer system having the AI routines that recognize and parse the speech of

the customer to add recognized menu items to the customer's order list. EX 1020, ¶¶[0031]; [0033]; [0044].

267. Jasper describes using the off-site aspects of its order taking system at specific times "to handle overflow orders, such as those order[s] which exceed the restaurant's or chain's own order taking capacity, e.g., lunch or dinner rush or during staffing shortages." EX 1021, ¶[0063].

268. A POSA would therefore understand Jasper to suggest that, when the restaurant's order taking capacity is not exceeded due to these circumstances, the system is deactivated in favor of the restaurant's order takers, namely, the on-site employees who enter orders into the RPOS 30 system. EX 1021, ¶¶[0056]; [0058]. Jasper thus discloses "deactivates ... in response to a detected change in customer traffic," in the form of deactivating its system (which is a substitute for on-site order taking) in favor of on-site order taking in response to the end of a lunch or dinner rush.

269. A POSA would have found it obvious to modify Mahaffy in view of Jasper for at least the reasons discussed in Section IX.A. above.

270. In addition, it would have been obvious to a POSA to modify Mahaffy to provide its operators with the ability to deactivate the AI routines in favor of on-site employee order taking in response to the end of a lunch or dinner rush, as suggested by Jasper. A POSA would be motivated to so modify Mahaffy in view of

Jasper in order to reduce the costs of paying operators to be on-call at times when on-site employees can handle the order-taking demand.

271. It is therefore my opinion that a POSA would have found the limitations of Claim 10 obvious based on the combined teachings of Mahaffy and Jasper.

L. Claim 11

wherein the auditor deactivates the order processor in response to a detected change in staffing at the site

272. I understand that Claim 11 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

273. As discussed in [1b], [4b], and [4g] above, Mahaffy discloses “the auditor” as human operators 40 that work at the call center, and “the order processor” as the computer system having the AI routines that recognize and parse the speech of the customer to add recognized menu items to the customer’s order list. EX 1020, ¶¶[0031]; [0033]; [0044].

274. Jasper describes using the off-site aspects of its order taking system at specific times “to handle overflow orders, such as those order[s] which exceed the restaurant’s or chain’s own order taking capacity, e.g., lunch or dinner rush or during staffing shortages.” EX 1021, ¶[0063].

275. A POSA would therefore understand Jasper to suggest that, when the restaurant's order taking capacity is not exceeded due to these circumstances, the system is deactivated in favor of the restaurant's order takers, namely, the on-site employees who enter orders into the RPOS 30 system. EX 1021, ¶¶[0056]; [0058]. Jasper thus discloses "deactivates ... in response to a detected change in staffing at the site," by deactivating its system (which is a substitute for on-site order taking) in favor of on-site order taking in response to the end of a staffing shortage.

276. A POSA would have found it obvious to modify Mahaffy in view of Jasper for at least the reasons discussed in Section IX.A. above.

277. In addition, it would have been obvious to a POSA to modify Mahaffy to provide its operators with the ability to deactivate the AI routines in favor of on-site employee order taking in response to the end of a lull or dinner rush, as suggested by Jasper. A POSA would be motivated to so modify Mahaffy in view of Jasper in order to reduce the costs of paying operators to be on-call at times when on-site employees can handle the order taking demand.

278. It is therefore my opinion that a POSA would have found the limitations of Claim 11 obvious based on the combined teachings of Mahaffy and Jasper.

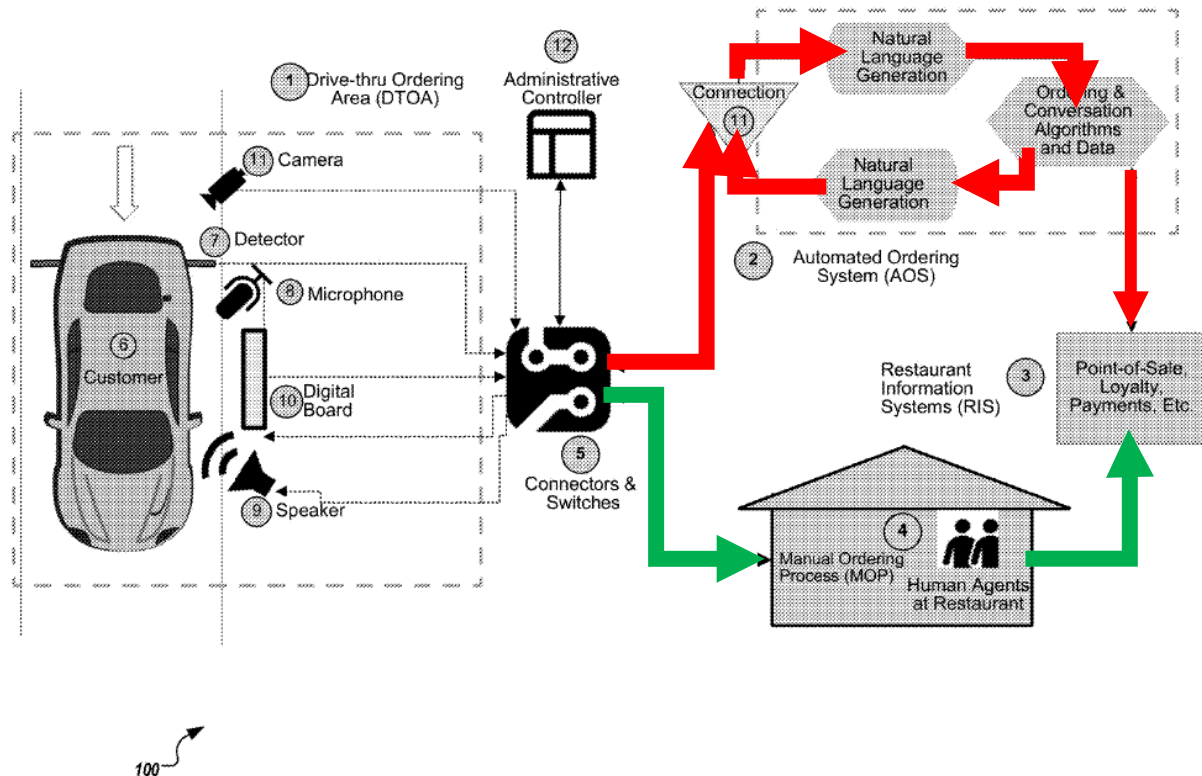
X. GROUND 2A – CLAIMS 1-4 ARE OBVIOUS OVER COLEMAN

A. Claim 1

[1-Pre] – A method comprising the steps of:

279. I understand that the “preamble” of a claim may or may not be considered a limitation of the claim. Regardless of whether the preamble is determined to be a limitation of this claim, it is my opinion that Coleman describes to a POSA the subject matter of the preamble.

280. Coleman generally relates to a system whose operation carries out a method for automated drive-thru ordering. FIG. 1, reproduced below with annotations, illustrates aspects of the system described by Coleman.



281. In an automated workflow (red arrows), an automated order system (“AOS”) 2 uses a natural language understanding (“NLU”) component to convert the spoken order into transcribed text and intent so as to processes the customer’s

order. EX 1022, ¶¶[0034]-[0035]; [0070]-[0071]. In a human-assisted workflow (green arrows bottom right of figure), the customer's spoken order is processed by a manual order process system ("MOP") 4, in which human agents take the order via interfacing with the customer using, e.g., headsets. EX 1022, ¶¶[0035]-[0036]; [0068]. The voice ordering process is rerouted from the AOS to the MOP if the automated workflow is not satisfactorily progressing to a completed order. EX 1022, ¶¶ [0021]-0024]; [0036]; [0071]-[0077].

282. I provide further details concerning the method introduced by the preamble in the sections that follow.

*[1a] – providing an audio stream of an order
of a customer positioned on a site*

283. Coleman discloses "an audio stream of an order of a customer" by "at least one microphone 8 is used to receive and transduce audible expressions from customers associated with the order interactions." EX 1022, ¶[0029]. The customer is "positioned on site" because the microphone 8 is part of a drive-thru ordering area ("DTOA") 1 of a restaurant. EX 1022, ¶¶[0027]-[0029].

284. Coleman further discloses that the order's audio stream is "provid[ed]" to a manual order process ("MOP") system 4, by describing that the MOP 4 "includes the ability for human agents at the restaurant site to interact with a customer 6 in the DTOA 1 by interfacing with the microphone 8, speaker 9 and detector 7 through wired or wireless communication equipment (e.g. headsets, etc.),

telephony communication, and/or computer software, which equipment or software may include the ability to record and play back audio to the speaker 9.” EX 1022, ¶[0036].

285. The audio stream is also “provid[ed]” to an automated ordering system (“AOS”) 2 that “has administrative control of the DTOA equipment,” and “can receive audio input from the microphone [during an order taking interaction] and can provide output to be visually displayed by the digital board 10 or audibly emitted by the speaker 9.” EX 1022, ¶¶[0069]-[0070].

286. A POSA would understand that microphones, speakers and other communication equipment operate to allow such interaction by transmitting audio streams of words spoken into the microphone. Accordingly, the order taking interaction of human agents and customers at the drive-thru, via microphones, speakers and other communication equipment, would be understood by a POSA to include providing an audio stream of the customer order to the MOP system 4 and to the AOS 2.

287. It is therefore my opinion that the limitations of [1a] are disclosed by Coleman.

*[1b] – providing an order processor having
a speech recognition module trained using
artificial intelligence programs*

288. Coleman discloses an “order processor” in the form of the AOS 2, discussed in [1-Pre], and the RIS 3 that “the restaurant uses to manage and execute transactions with customers.” EX 1022, ¶¶ [0035]; [0080]. EX 1022, ¶[0034]. In an automated workflow (red arrows of annotated FIG. 1 below), the AOS “manages the [customer] interactions after they have begun and processes the interactions ... while interpreting and responding to customer input.” EX 1022, ¶[0070].

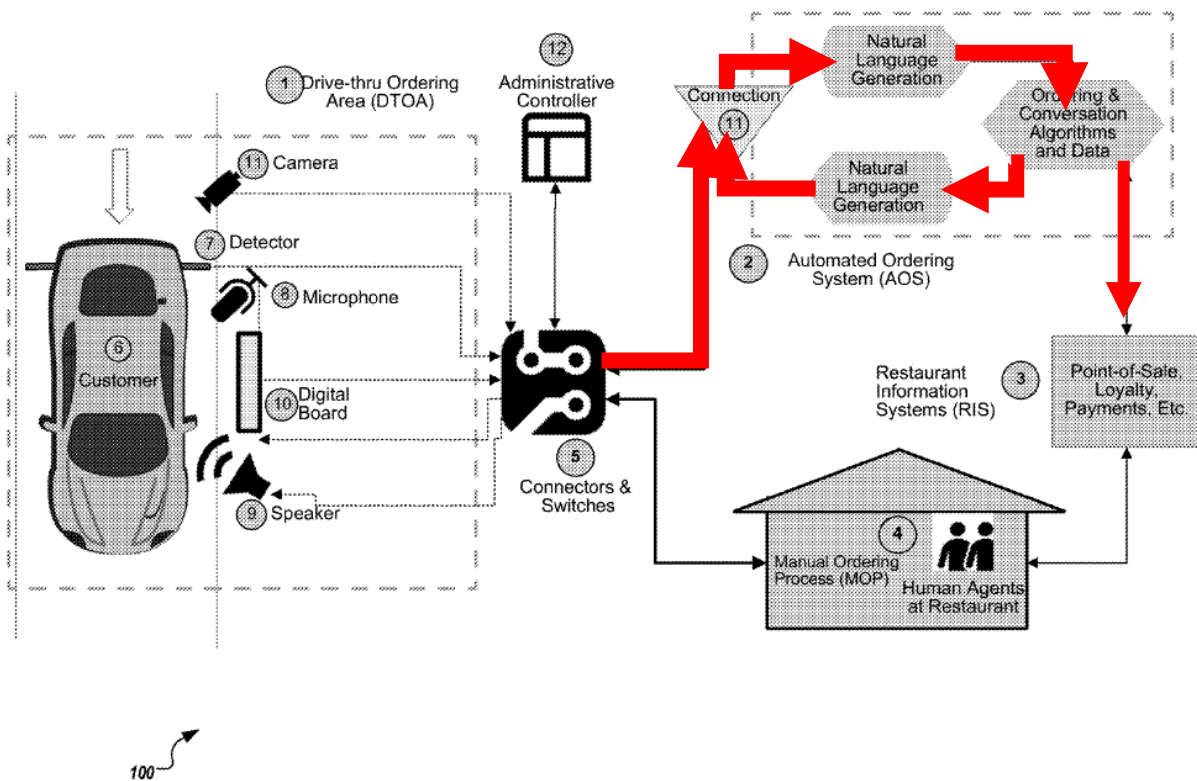


FIG. 1

EX 1022, FIG. 1. The AOS 2 transmits the order information to the RIS 3 so that the order can be fulfilled by the RIS 3 and on-site employees. EX 1022, ¶¶[0055]; [0080]. The customers thus “order in the same manner as they would place an order if speaking to a human worker.” EX 1022, ¶[0020].

289. Coleman further discloses a “speech recognition module” in the form of the NLU component “that converts human speech into transcribed text and intents.” EX 1022, ¶[0034]. As explained in Section IV.A. above, speech recognition or ASR is the conversion of human speech into transcribed text via computer. A POSA would therefore understand that the NLU component is a speech recognition module.

290. The NLU is also part of the AOS 2, which additionally includes “a set of ordering and conversation algorithms” that are “programmed to select actions to accurately process the customer’s order.” EX 1022, ¶¶ [0034]; [0047]. A POSA would further understand that processing an audio signal of the customer’s verbal order by the AOS 2 necessarily involves speech recognition.

291. A POSA would also understand that the NLU would be “trained using artificial intelligence.” Coleman discloses, or at least suggests, such training by describing that the set of ordering and conversation algorithms “may include machine learning models.” EX 1022, ¶ [0047].

292. As explained in Section IV.A. above, a POSA would understand that, at a basic level, most if not all speech-recognition modules at the time, including the NLU of Coleman, would be trained by providing corpus of sample audio that has been assigned text words or phrases to a neural network (i.e., a type of machine

learning model) to train the neural network's ability to pattern match against incoming audio.

293. Indeed, ASR neural networks at the time were trained, at a high level, by manually transcribing speech into words and phrases, and assigning those words and phrases to the ASR determined phoneme sequences of the speech, in order to build a corpus of known words and phrases.

294. The training also generally involved speaking to the ASR and correcting mistakes in the ASR's transcription of the speech in a training feedback loop. Where corrections are made, the corrected phoneme sequence to known word association is stored as a reference for the neural network to pattern match against (i.e., the training dataset is updated). Where corrections are not made, the phoneme sequence to known word association is reinforced.

295. This is also precisely how the '706 Patent describes its speech-to-text module 41 is trained. EX 1001, 9:40, et seq. Sample audio and its assigned human-transcribed text is provided to a neural network so as to build the pattern recognition program that is the speech-to-text module. EX 1001, 11:14-33. There is no difference between this and what a POSA would understand from Coleman's teachings.

296. A POSA would therefore understand that Coleman’s NLU is trained in accordance with the same basic ASR training process, and is thus “trained using artificial intelligence.”

297. Moreover, to the extent it is not inherent that Coleman’s NLU is “trained using artificial intelligence,” a POSA would nevertheless have found it obvious to train Coleman’s NLU using artificial intelligence in light of the general knowledge in the art. In particular, POSA would be readily familiar with the fact that most, if not all, ASR systems at the time were trained according to the general approach discussed above. Thus, it would have been obvious to a POSA to similarly train Coleman’s NLU in order to recognize words and phrases in users’ speech, as merely applying a known technique to improve a similar speech recognition system in the same way.

298. It is therefore my opinion that the limitations of *[1b]* are disclosed or otherwise rendered obvious by Coleman.

*[1c] – converting a word or words in the
audio stream to text using the speech
recognition module*

299. As discussed in *[1b]* above, Coleman discloses the “speech recognition module” in the form of the NLU.

300. Coleman further discloses that the NLU “converts human speech into transcribed text.” EX 1022, ¶[0034]. The human speech that the NLU converts into

text is the audio stream, discussed in [1a] above, that is received by the AOS 2 for processing. EX 1022, ¶¶[0069]-[0070].

301. In addition, a POSA would understand that this limitation merely recites the operation of the speech recognition module discussed in [1b] on the audio stream of the customer's order, in accordance with the basic ASR process. In particular, as explained in Section IV.A. above, a POSA would understand that most if not all speech-recognition routines at the time, including the NLU of Coleman, work by converting incoming an audio stream into a text-based phoneme sequence (the sequence of sound units, e.g., *th*, *p*, *d*, etc.) for pattern matching against known words or word sequences. For example, the audio "three SuperBurgers" may be converted to the phoneme sequence "th-r-ee--s-oo-p-u-r--b-u-r-g-u-r", which may be pattern-matched against known words to generate the text "three super burgers." Once identified, the known words or word sequences are pattern matched to known intents or meanings. For example, the text "three super burgers" may be pattern matched to the known intent of adding three SuperBurgers to the order.

302. It is therefore my opinion that the limitations of [1c] are disclosed by Coleman.

[1d] – processing the text communication with the speech recognition module to identify a word or words in the text of the converted audio stream according to a previous spoken word training

303. As discussed in [1b] and [1c] above, Coleman discloses the “speech recognition module” by its NLU, and the “text of the converted audio stream” by the NLU transcription of the audio stream.

304. Coleman discloses the remainder of this limitation in the form of the NLU converting the human speech into “transcribed text and intents.” EX 1022, ¶[0034]. In the context of the drive-thru ordering system of Coleman, a POSA would understand the “intents” of the customer’s verbal order include the intent to add identified menu items (i.e., words corresponding to menu items) to the customer’s order. Coleman indeed discloses that the system “interacts with customers using voice input from the customers ... that allow the customers to place and review orders conversationally ... in the same manner as they would place an order if speaking to a human restaurant worker.” EX 1022, ¶[0020]. Thus, the intent of the customer verbally placing an order is to add the spoken menu items to the order.

305. Coleman further suggests that the “intents” include the intent to add spoken menu items to the customer’s order by disclosing that its “ordering and conversion algorithms ... process the inputs from the NLU,” and “transmit[] the order information to the RIS 3 so that the order can be fulfilled.” EX 1022, ¶¶[0034]; [0047]; [0055]; [0080]. Again, a POSA would understand that the intent of the customer verbally placing an order is to add the spoken menu items to the order.

306. A POSA would also understand that the speech recognition includes neural networks that perform pattern matching to identify words “according to a previous spoken word training.” That is, as explained in Section IV.A. above, at the time of Coleman, a POSA would have expected an ASR neural network to be trained, at a high level, by manually transcribing speech into words and phrases, and assigning those words and phrases to the ASR determined phoneme sequences of the speech, in order to build a corpus of known words and phrases. The training also generally involved speaking to the ASR and correcting mistakes in the ASR’s transcription of the speech in a training feedback loop (i.e., the training dataset is updated). Where corrections are made, the corrected phoneme sequence to known word association is stored as a reference for the neural network to pattern match against. Where corrections are not made, the phoneme sequence to known word association is reinforced. A POSA would be readily familiar with the fact that most, if not all, ASR systems at the time, including the NLU of Coleman, would have been trained according to this general approach.

307. This is also precisely how the ‘706 Patent describes the training of its speech-to-text 41 module is trained. EX1001, 9:40, et seq. Sample audio and its assigned human-transcribed text is provided to a neural network so as “to build a pattern recognition program” that is the speech-to-text module. EX 1001, 11:14-33.

There is no difference between this and what a POSA would understand from Coleman's teachings.

308. To the extent it is not inherent that Coleman's NLU identifies words "according to previous spoken word training," a POSA would nevertheless have found it obvious to train Coleman's NLU using spoken word training, in light of the general knowledge in the art. In particular, POSA would be readily familiar with the fact that most, if not all, ASR systems at the time were trained according to the general approach discussed above. Thus, it would have been obvious to a POSA to similarly train Coleman's NLU in order to recognize words and phrases in users' speech according to previous spoken word training, as merely applying a known technique to improve a similar speech recognition system in the same way.

309. It is therefore my opinion that the limitations of [1c] are disclosed or otherwise rendered obvious by Coleman.

[1e] – providing a natural language processor having order assembly capabilities and exception detection capabilities, wherein the natural language processor receives recognized text from the order processor and creates or modifies an order based upon the recognized text

[1f] – generating an order with the natural language processor

310. Coleman discloses a "natural language processor" in the form of the "ordering and conversion algorithms that process the inputs from the NLU ... to

determine or select the next action to be taken by the AOS 2,” so as “to accurately process the customer’s order.” EX 1022, ¶¶ [0034]; [0047]. The selected “next action” includes transmitting order information to the RIS 3 so that the order can be fulfilled. EX 1022, ¶¶[0080]; [0047]; [0055].

311. Coleman also discloses the AOS can perform “modifications to the order or confirm[] certain aspects of the order.” EX 1022, ¶[0059]. Coleman thus further discloses “order assembly capabilities,” “creates or modifies an order,” and “generating an order,” by the ordering and conversion algorithms.

312. Moreover, as discussed in *[1d]* above, a POSA would understand that the ordering and conversion algorithms add recognized menu items from the customer’s speech to the customer’s order – and would therefore involve “receiv[ing] recognized text” corresponding to the menu items from the NLU.

313. Still further, as explained in Section IV.A. above, a POSA would understand that speech-recognition software, including the module of Coleman, works by converting incoming an audio stream into a text-based phoneme sequence (the sequence of sound units, e.g., *th*, *p*, *d*, etc.) for pattern matching against known words or word sequences. For example, the audio “three SuperBurgers” may be converted to the phoneme sequence “th-r-ee--s-oo-p-u-r--b-u-r-g-u-r”, which may be pattern-matched against known words to generate the text “three super burgers.”

314. Once identified, the known words or word sequences are pattern matched to known intents or meanings. For example, the text “three super burgers” may be pattern matched to the known intent of adding three SuperBurgers to the order. EX 1022, ¶[0040].

315. Thus, a POSA would understand that the ordering and conversion algorithms process the NLU’s transcript of the customer order so as to transmit order information to the RIS for fulfilment, which describes “the natural language processor receives recognized text from the order processor and creates or modifies and order based upon the recognized text.”

316. Coleman also discloses, in the context of an automatic transfer from the automated (“AOS”) 2 process to a manual (“MOP”) 4 process, detecting exceptions to using the AOS to process an order, in the form of detecting “suitable reasons” for the AOS 2 determining that the transfer from the AOS 2 process to the manual (“MOP”) 4 process should occur. Coleman describes that those “suitable reasons” include “whether the customer 6 is speaking or providing inputs in an unclear manner,” “whether the monitored customer behavior is particularly unusual or negative,” whether the customer speech is intelligible but “highly atypical,” whether connection has been lost or there is an “internal error,” or whether a human at the restaurant interrupted the connection. EX 1022, ¶¶[0071]-[0075]. Coleman thus discloses that the AOS has “exception detection capabilities.”

317. It is therefore my opinion that the limitations of [1e] and [1f] are disclosed or otherwise rendered obvious by Coleman.

[1g] – alerting an auditor of detected exceptions in the order, the auditor located off-site and connected to the order processor via a data connection

318. As discussed in [1e] above, Coleman discloses “detected exceptions” in the form of the suitable reasons for an automatic transfer of the ordering process by the AOS 2 to the MOP 4.

319. Coleman further discloses “alerting an auditor of detected exceptions in the order” in the form of disclosing that, “if an active ordering process with the AOS 2 is re-routed to the MOP 4, then [] human agents are alerted to the re-routed interaction and communicate with the customer 6 through the DTOA 1 equipment and can complete the order in the usual manual manner.” EX 1022, ¶[0077]. The human agents can be on-site or, as is relevant for this limitation, can be remotely connected to the RIS 3 via a data connection. EX 1022, ¶[0036].

320. A POSA would further understand that the act of alerting the human agents to the re-routed interaction alerts them “of detected exceptions in the order,” because the active ordering process would not have been re-routed to the human agents but for the detected exception.

321. Moreover, Coleman describes that alerting the human agent to the rerouted interaction includes transmitting to the human agent “contextual

information about the interactions” that allows the human agent to immediately assist in and take over the interaction. EX 1022, ¶¶[0036], [0077]. Because, as discussed in [1e] and [1f] above, rerouting the interaction is due to detected exceptions (i.e., determined “suitable reasons”), a POSA would understand that the contextual information, which would allow the human agent to immediately assist in and take over the interaction, identifies those suitable reasons so that the human agent may assist. Coleman thus further discloses alerting the human agents “of detected exceptions in the order.”

322. In addition, Coleman describes that the human agents are “able to listen in to the automated interactions with a customer” and “interrupt[] the connection to the AOS 2 using the administrative controller 12 or by manually manipulating or interacting with the connectors or switches 5.” EX 1022, ¶[0045]. More particularly, the human agents can “determine a new context” which justifies interrupting the AOS 2 “[i]f the automated process is not completing successfully, such as due to poor interactions or understanding of the customer.” EX 1022, ¶[0089].

323. A POSA would understand that these circumstances (e.g., that the automated process is not completing successfully due to poor understanding of the customer) correspond to the “suitable reasons” for transferring the ordering process from the AOS 2 to the MOP 4, which are discussed in [1e] above.

324. Coleman therefore also discloses “alerting an auditor of detected exceptions” in the form of the human agent listening in on the automated order taking process being alerted that the automated process of the AOS 2 is unsuccessful, and should be re-routed to the MOP 4, when the human agent hears the suitable reasons for the re-routing. The human agent is thus alerted to the “suitable reasons” for transitioning the ordering process from the AOS to the MOP in the course of listening in on the automated order taking to detect those “suitable reasons.”

325. To the extent Coleman does not explicitly describe the human agent listening in on the automated interaction for the “suitable reasons” to interrupt the automated process of the AOS 2 (so as to reroute the order taking to the MOP 4) a POSA would nevertheless recognize that the human agent is listening in for some reason to interrupt the automated process of the AOS 2. The only reasons Coleman discloses as prompting an interruption of the automated process of the AOS 2 are the “suitable reasons.” EX 1021, ¶¶[0071]-[0075]. While this disclosure is in the context of the AOS 2 interrupting its own automated process, a POSA would understand that these “suitable reasons” are suitable regardless of whether the AOS 2 or the human agent listening in on the automated process determines them. Coleman indeed notes that “another component” other than the AOS 2 can make the determination, with the human agent being the only other “component” of

Coleman's system that decides whether to interrupt the automated interaction of the AOS. EX 1021, ¶¶[0071]; [0076].

326. It is therefore my opinion that the limitations of [1g] are disclosed or otherwise rendered obvious by Coleman.

*[1h] – deactivating the order processor with
the auditor in response to the detected
exceptions in the order*

327. As discussed in [1e] above, Coleman discloses “detected exceptions” in the form of the suitable reasons for an automatic transfer of the ordering process by the AOS 2 to the MOP 4.

328. Coleman discloses the remainder of the limitation in the form of the automatic transfer of the ordering process by the AOS 2 to the MOP 4, namely, from the automatic workflow (red arrows) to a human-assisted workflow (green arrows):

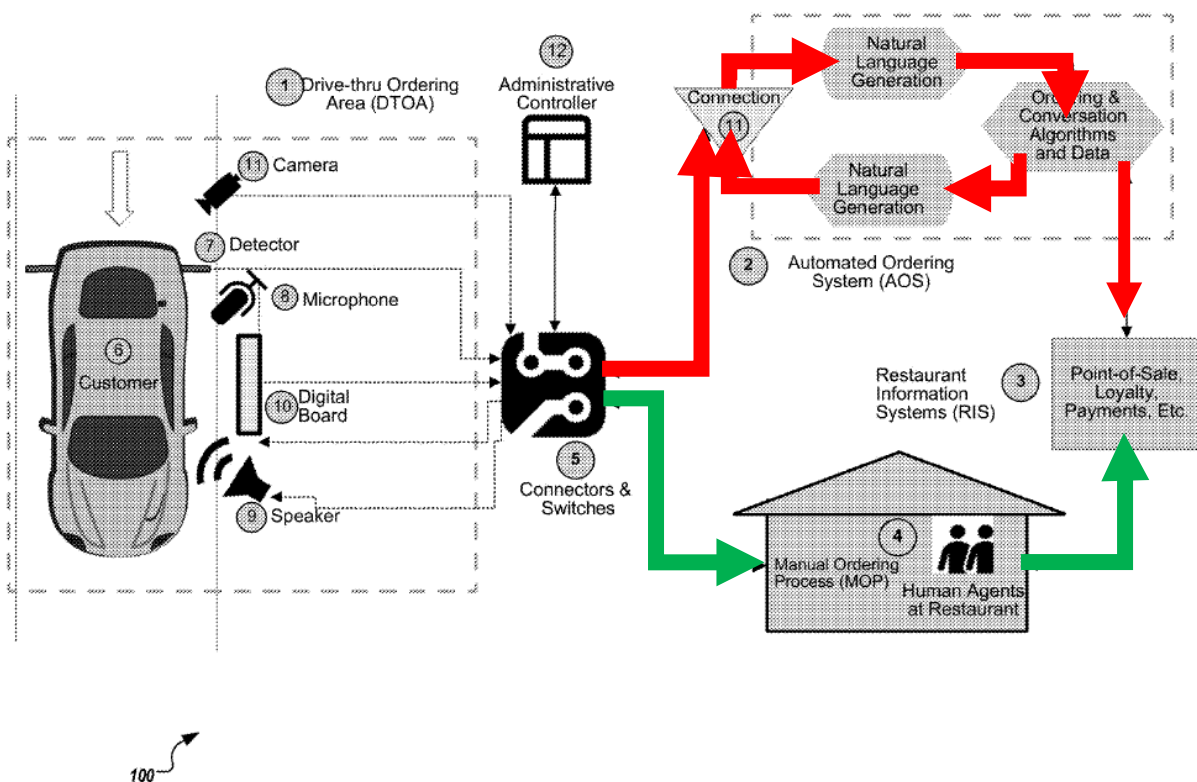


FIG. 1

EX 1022, FIG. 1.

329. More particularly, Coleman discloses that the switches 5 control the transfer (or re-routing) of the ordering process between the automatic workflow (red arrows) of the AOS2 and the human-assisted workflow (green arrows) of the MOP 4. EX 1022, ¶¶[0038]-[0046]. The switches also control whether the ordering process is initiated with the AOS 2 or the MOP 4. EX 1022, ¶¶[0023]; [0063]; [0067]. The switches 5 can be controlled by the human agents, via the administrative controller 12, to determine if orders “shall be taken (a) manually by human agents as part of the MOP 4 or (b) by the AOS 2.” EX 1022, ¶[0045]. A POSA would understand that actuating the switches 5 to reroute the ordering process from the

automatic workflow of the AOS 2 to the human-assisted workflow of the MOP 4 is an effective deactivation of the AOS 2. Coleman therefore discloses “deactivating the order processor with the auditor.”

330. Moreover, as also discussed in [1g] above, Coleman describes that the human agents are “able to listen in to the automated interactions with a customer” and “interrupt[] the connection to the AOS 2 using the administrative controller 12 or by manually manipulating or interacting with the connectors or switches 5.” EX 1022, ¶[0045]. More particularly, the human agents can “determine a new context” which justifies interrupting the AOS 2 “[i]f the automated process is not completing successfully, such as due to poor interactions or understanding of the customer.” EX 1022, ¶[0089].

331. A POSA would understand that these circumstances (e.g., that the automated process is not completing successfully due to poor understanding of the customer) correspond to the “detected exceptions” that are suitable reasons for transferring the ordering process from the AOS 2 to the MOP 4, which are discussed in [1e] above. Coleman therefore further discloses that the human agent deactivates the AOS 2 “in response to the detected exceptions.”

332. It is therefore my opinion that the limitations of [1h] are disclosed by Coleman.

*[1i] – prompting an on-site employee to
engage the customer and complete the order*

333. Coleman discloses this limitation in the form of the AOS 2 transmitting the order information to the RIS 3 so that the order can be fulfilled by the RIS 3 and the restaurant's human agents. EX 1022, ¶[0080]. More particularly, Coleman describes that its human agents handle drive-thru orders "in a manner that is consistent with the typical manual drive-thru ordering process throughout the restaurant industry." EX 1022, ¶[0036]. As discussed in Section IV above, a POSA would understand that the typical drive-thru ordering process involves prompting an on-site employee to put the ordered food in a bag and deliver it to the customer. A POSA would therefore understand that Coleman's human agents at the restaurant are prompted by the AOS 2 and/or RIS 3 to "engage the customer and complete the order" by assembling the order and delivering it to the customer.

334. A POSA would also understand such typical activity as inherent in Coleman's reference to the order being fulfilled by the restaurant's human agents in a manner that is consistent with the typical manual drive-thru ordering process. EX 1016 ("A dedicated drive-thru runner picks up the menu items, assembles them in a bag and brings the bag to a team member at the drive-thru window. ... [T]he person staffing the drive-thru window ... takes payment and passes the bag or other packaging to the customer.")

335. The '706 Patent indeed similarly describes that the "on-site employee 81 ... is still responsible for taking payment, and delivering food to a customer 25,"

in the course of which, the order is sent to the on-site POS system 61 that “generates a status indicator that payment has been received so that on-site employee 81 may fulfill the order by delivering the ordered food to the customer 25.” EX. 1001, 9:29-33; 13: 53-56. That the ‘706 Patent refers to the on-site employee as “still” responsible indicates that these responsibilities are conventional to the drive-thru process.

336. There is no difference between what the ‘706 Patent describes here and what a POSA would understand from Coleman’s teachings.

337. To the extent it is not inherent, a POSA would nevertheless have found it obvious to prompt an on-site employee to engage the customer and complete the order, in light of the general knowledge in the art discussed above. A POSA would understand that doing so is merely applying a known technique to improve a similar system in the same way, namely, providing a process for completing the order at the drive-thru location.

338. It is therefore my opinion that the limitations of [11] are disclosed or otherwise rendered obvious by Coleman.

B. Claim 2

wherein the order is processed in real time

339. I understand that Claim 2 depends from Claim 1, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

340. Coleman discloses this limitation in the form of a “drive-thru system implementation,” in which the “order is placed at the DTOA 1 and the customer drives to a window or other area to receive the order.” EX 1022, ¶¶[0025]-[0027]; FIG. 1.

341. More particularly, Coleman describes that “NLP can be used in machine or conversational interfaces to replace the need for another human to be interacting in real-time with a customer or user.” EX 1022, ¶[0003]. Coleman also describes that “[t]he system allows customers to order in the same manner as they would place an order if speaking to a human restaurant worker.” EX 1022, ¶¶[0020]; [0094]. A POSA would understand that processing the drive-thru orders, including those of Coleman, is done in real-time.

342. It is therefore my opinion that the limitations of Claim 2 are disclosed by Coleman.

C. Claim 3

wherein the alert to the auditor prompts the auditor to switch the order processor from an autonomous mode to an auditor-assist mode prior to deactivating the order processor

343. I understand that Claim 3 depends from Claim 1, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

344. As discussed in *[1h]* and *[1g]* above, Coleman discloses “deactivating the order processor” in the form of initiating the ordering interaction with the MOP 4 rather than the AOS 2, and “the alert to the auditor” by the human agents, who are listening in on the automated ordering interaction, hearing suitable reasons for re-routing the ordering interaction from the AOS 2 to the MOP 4.

345. Coleman further discloses an “autonomous mode” in the form of the AOS 2 “interact[ing] with the customer 6 by processing the voice input ... and generating response to the customer ... to complete an order.” EX 1022, ¶¶[0070]; [0034].

346. The “auditor-assist mode” is disclosed by Coleman in the form of the MOP 4 for human agents to “communicate with the customer 6 and manipulate the RIS 3 to transact the customer’s order.” EX 1022, ¶¶[0070]; [0036].

347. Coleman discloses the remainder of the limitation because determining to initiate the ordering interaction with the MOP 4 (and thus deactivate the AOS 2) can be based on “whether a prior attempt at an automated solution in a previous interaction was successful,” or “required a failover to [the] manual system.” EX 1022, ¶¶[0032]; [0067].

348. More particularly, a POSA would understand that the human agent, who is listening in on a prior AOS 2 interaction with the customer, may be alerted to the detected exceptions by hearing them, which may prompt the human agent to re-route the interaction from the AOS 2 to the MOP 4. The human agent may then, based on that prior failover to the MOP 4, determine that the next interaction is initiated with the MOP 4.

349. Coleman thus discloses that the auditor is prompted to re-route the previous ordering interaction “prior to” deactivating the AOS 2 in initiating the current ordering interaction.

350. It is therefore my opinion that the limitations of Claim 3 are disclosed by Coleman.

XI. GROUND 2B – CLAIMS 4-11 ARE OBVIOUS OVER COLEMAN IN VIEW OF WASHINGTON

A. Motivation to Combine

351. It is my opinion that a POSA considering Coleman would have looked to Washington to improve the ordering process of Coleman.

352. Coleman and Washington each relate to automating drive-thru interactions using an automated speech-to-order system. *E.g.*, EX 1022, ¶[0020]; EX 1023, ¶¶[0020]-[0022]. Each also recognizes that failing to accurately take the customer’s order is undesirable. *E.g.*, EX 1022, ¶[0021] (noting that the ability to handle errors and alleviate other issues in order taking is “highly desirable”); EX

1023, ¶[0020] (noting that “errors may be introduced into the order ... [which] errors cause customer dissatisfaction [and] decrease efficiency of the drive-thru system.”).

353. This problem of inaccurate order taking was also generally known in the field of automated order taking. EX 1022, ¶[0021] (“while NLP services continue to improve, on occasion, the customer’s behavior and/or the ordering environment may preclude the automated system from smoothly completing the ordering process.”); EX 1020, ¶¶[0039] (“important issue is the ability of the AI routines ... to recognize when the result of the speech recognition is correct and, when it is incorrect, to cause the system to task the customer for a clarification or cause intervention by a human operator”).

354. The solution proposed by Coleman minimizes potential order inaccuracies by, among other things, having a human agent “listen in or otherwise follow an ongoing interaction,” so that the human agent can “at any time, interrupt the automatic interaction to move the interaction to a manual process” if the automated process is not completing successfully, “such as due to poor interactions or understanding of the customer,” or for other “suitable reasons.” EX 1022, ¶¶[0089]; [0036]; [0071]-[0076].

355. Coleman also discloses that the human agent can be provided with “a listing of interactions performed so far” or other “contextual information” about the

interactions between the automated system and the customer and has the ability to “follow an ongoing automatic interaction.” EX 1022, ¶¶[0036], [0077], [0089].

356. The human agent can also be provided with “a summary of instructions received” from the automated system, in other words, a summary of the order taken by the automated system to that point, but Coleman is not explicit whether this happens before or after Coleman’s switchover to the manual process (the “manual order process” or “MOP 4”). EX 1022, ¶[0090].

357. Washington guards against order inaccuracies using a similar human-assist approach that gives the human agent more flexibility to correct the automated system. Like in Coleman, a human agent (an “operator 116”) of Washington can override the automated system. EX 1023, ¶[0070].

358. But Washington further discloses, whereas Coleman only suggests, that the operator can take the additional step of comparing the customer’s actual audio order with the autonomously generated order. EX 1023, ¶[0038] (The “operator 116 may review the menu items entered into the order input device 132 and then press a “Submit” button when satisfied that the menu items are correct (e.g., correspond to the actual menu items that the customer ordered)”). More particularly, the operator hears the order spoken into the microphone via headset (*id.*, ¶ [0020]), and can review the menu items entered into and displayed on the order input device 132 to

ensure the menu items match the order as given by the customer and/or the formatted transcript shown on the display 134 (*id.*, ¶¶ [0022]; [0036]; [0038]).

359. Washington also discloses, whereas Coleman only suggests, that if the operator determines there to be failures with the autonomously generated order, the operator can correct those failures. EX 1023, ¶¶[0023], [0038], [0039], [0069]-[0070].

360. A POSA would have found it obvious to modify Coleman, in line with what Coleman already suggests, to better guard against order inaccuracies by giving Coleman’s human agent the explicit ability to compare the automatically generated order with the customer’s spoken order prior to the switchover to Coleman’s auditor-assist mode, as suggested by Washington.

361. More particularly, Coleman already recognizes that a “suitable reason” for interrupting the autonomous order taking process can be that the speech-recognition is inaccurately interpreting the customer’s order. EX 1022, ¶[0080]. Washington teaches that a human operator auditing an automated order processing can identify such inaccurate interpretations and minimize inaccuracies in the order taking by overriding the automated order processing and manually correcting the order when a comparison of the entered menu items with the customer’s order reveal that they do not correspond. EX 1023, ¶[0038].

362. A POSA would therefore have modified Coleman to allow further human verification of the accuracy of a speech-recognition generated order by, as suggested by Washington, having Coleman's human agent review the AOS generated order for inaccuracies, as a "suitable reason" for interrupting the AOS order processing, while already listening in on the interaction. In other words, the human agent of Coleman would compare the autonomously generated order with the customer's audio order that the agent is listening in on for the purpose of identifying inaccuracies, as in Washington, and intervening where such inaccuracies are identified.

363. A POSA would understand that this combination would allow for more immediate intervention by Coleman's human agent when the AOS is not accurately interpreting the customer's order, than if the human agent were to only listen for audio-only indications that the AOS is inaccurately interpreting the customer's order (e.g., the customer verbally indicates the AOS order is incorrect). Consequently, customer dissatisfaction and inefficiencies would be reduced.

B. Claim 4

[4-Pre] – A method comprising the steps of:

[4a] – providing an audio stream of an order of a customer positioned on a site

[4b] – providing an order processor having a speech recognition module trained using artificial intelligence programs

[4c] – converting a word or words in the audio stream to text using the speech recognition module

[4d] – processing the text communication with the speech recognition module to identify a word or words in the text of the converted audio stream according to a previous spoken word training

[4e] – providing a natural language processor having order assembly capabilities and exception detection capabilities, wherein the natural language processor receives recognized text from the order processor and creates or modifies an order based upon the recognized text

[4f] – generating an order with the natural language processor

364. The majority of Claim 4 tracks *[1-Pre]* and *[1a]* to *[1f]*, and its limitations are satisfied by Coleman for the same reasons as explained above. The differences in Claim 4 are discussed below.

[4g] – providing an audio stream of a customer order to an auditor, the auditor located off-site and connected to the order processor via a data connection

365. Aside from the “providing” limitation, *[4g]* tracks *[1g]*, and its limitations are satisfied by Coleman for the same reasons as explained above.

366. Coleman further discloses “providing an audio stream of the customer order to an auditor” by disclosing that “even where the AOS 2 is triggered to start

an interaction, human agents may be listening in ... through headphones worn by one or more agents.” EX 1022, ¶[0036].

[4h] – providing the generated order to the auditor

367. Coleman discloses “providing the generated order to the auditor” by describing that the human agents may be “listening in ... through a listing of interactions performed so far.” Ex. 1022, ¶[0036].

368. A POSA would understand that, in the context of a drive-thru ordering system, the listing of interactions includes interactions for the customer adding menu items to the order. Indeed, Coleman discloses that the interactions include the “interpreting and responding to customer input” via which “the AOS 2 will continue to interact with the customer 6 by processing the voice input ... and generating responses ... to complete the order.” EX 1022, ¶[0070].

369. Moreover, as discussed in [4d] by way of [1d], such processing of the voice input includes the AOS 2 converting the voice input into “transcribed text and intents.” EX 1022, ¶[0034]. A POSA would understand that, in the context of a drive-thru ordering system, the intent of the customer voice input includes adding menu items to the customer’s order.

370. Coleman therefore discloses “providing the generated order” to its human agents at least in the form of the listing of interactions, which includes the transcribed text and intents, as well as the AOS generated responses.

371. It is therefore my opinion that the limitations of [4h] are disclosed by Coleman.

[4i] – performing a comparison of the audio stream of a customer order with the generated order by the auditor

[4j] – switching to an auditor-assist mode in response to the comparison of the audio stream with the generated order

372. As discussed in [4g] and [4h], Coleman discloses the “audio stream” and the “customer order.”

373. Coleman further discloses “switching to an auditor-assist mode” by re-routing the ordering process by the AOS 2 to the MOP 4, namely, from the automatic workflow (red arrows) to a human-assisted workflow (green arrows):

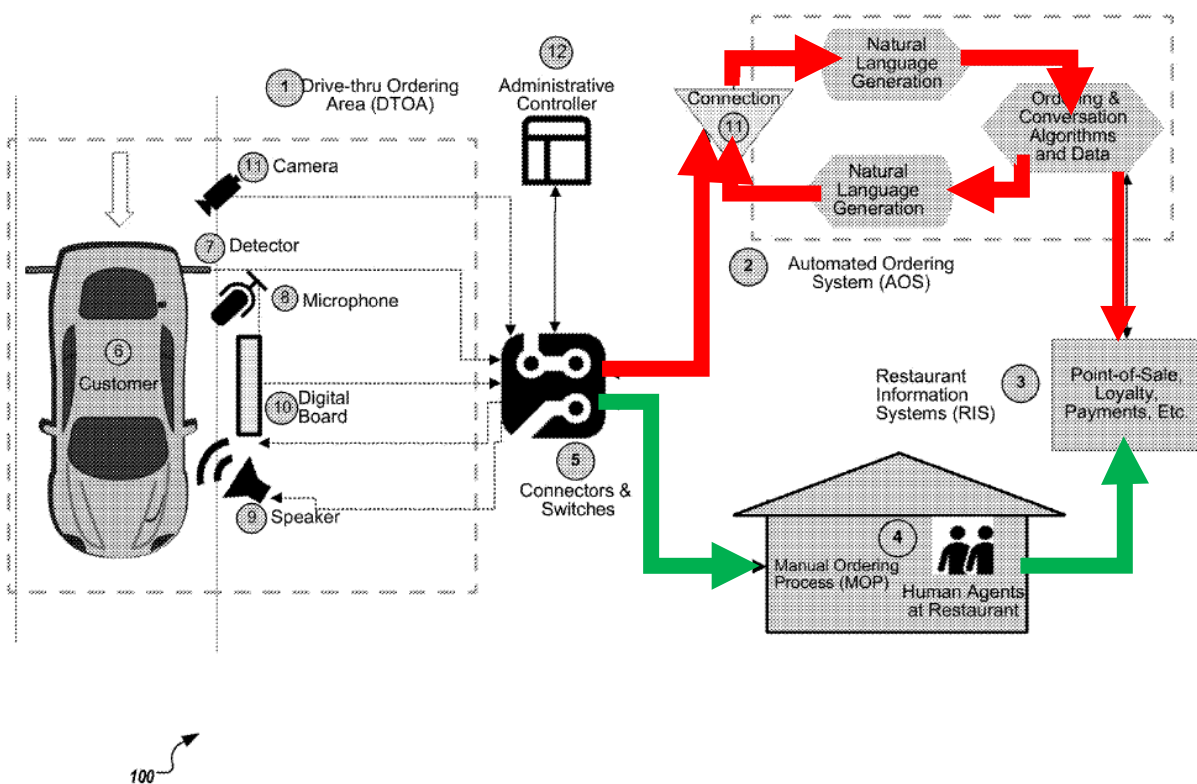


FIG. 1

374. More specifically, Coleman discloses that the switches 5 control the transfer (or re-routing) of the ordering process between the automatic workflow (red arrows) of the AOS2 and the human-assisted workflow (green arrows) of the MOP 4. EX 1022, ¶¶[0038]-[0046]. The switches 5 can be controlled by the human agents, via the administrative controller 12, to determine if orders “shall be taken (a) manually by human agents as part of the MOP 4 or (b) by the AOS 2.” EX 1022, ¶[0045].

375. Accordingly, the human agents are “able to listen in to the automated interactions with a customer” and “interrupt[] the connection to the AOS 2 using the administrative controller 12 or by manually manipulating or interacting with the

connectors or switches 5.” EX 1022, ¶[0045]. More particularly, the human agents can “determine a new context” which justifies interrupting the AOS 2 “[i]f the automated process is not completing successfully, such as due to poor interactions or understanding of the customer.” EX 1022, ¶[0089].

376. A POSA would understand that these circumstances (e.g., that the automated process is not completing successfully due to poor understanding of the customer) correspond to the “detected exceptions” that are suitable reasons for transferring the ordering process from the AOS 2 to the MOP 4, which are discussed in [1e] by way of [4e] above.

377. Coleman therefore discloses “switching to an auditor-assist mode,” in response to the human agents listening in and determining that “the automated process is not completing successfully.”

378. Coleman, however, does not expressly disclose “performing a comparison ... by the auditor” or that the switching is “in response to the comparison.”

379. Washington discloses “performing a comparison ... by the auditor” by describing a way of verifying the accuracy of a speech-to-text generated order with an “operator 116 [that] may review the menu items entered into the order input device 132 [by the speech-to-order system 106] and then press a “Submit” button when satisfied that the menu items are correct (e.g., correspond to the actual menu

items that the customer ordered).” EX 1023, ¶[0038]; ¶[0036] (“the order input device facilitates the operator 116 entering the order to be provided to the item ordering system”); ¶[0022] (“... the order is submitted into the item ordering system after the operator verifies (e.g., by submitting the order, etc.) that the menu items entered into the order input device match the menu items in the formatted transcript and/or the order as given by the customer”).

380. In the course of verifying the order accuracy, Washington discloses “providing an audio stream to an auditor” by “the customer speak[ing] their order into the microphone.” EX 1023, ¶[0036]. Washington also discloses that the operator has a headset via which to hear the actual order given by the customer into the microphone. *Id.*, ¶ [0020].

381. A POSA would understand that the audio stream of the order as given by the customer is provided to the operator so that the operator can verify the menu items entered into and displayed on the order input device match the order as given by the customer. The operator would indeed be unable to verify the accuracy of the AI generated order unless the operator is provided with the audio of the order as given by the customer.

382. Still further in the course of verifying the order accuracy, Washington also discloses “providing the generated order to the auditor” by “the point-of-sale terminal 110 [that] displays the formatted transcript highlighting the menu items

with the associated quantities and customizations.” EX 1022, ¶¶[0068]. More particularly, the speech-to-order system converts the customer speech into a formatted transcript and enters the menu items identified in the formatted transcript into the order input device 132. EX 1022, ¶[0038].

383. A POSA would thus understand that the operator of Washington verifies the accuracy of the AI generated order by comparing the menu items from formatted transcript to the audio of the order as given by the customer.

384. A POSA would have found it obvious to modify Coleman for at least the reasons discussed in Section XI.A. above.

385. In particular, a POSA would have found it obvious to modify Coleman to apply Washington’s way of verifying the accuracy of a speech-recognition generated order, by having Coleman’s human agent review the AOS generated order for inaccuracies, as a “suitable reason” for interrupting the AOS order processing, while already listening in on the interaction. The combination of Coleman and Washington meets the of [4i] and [4j] in the form of the human agent of Coleman listening in on the AOS interaction and comparing the AOS generated order with the customer audio to check for inaccuracies in the AOS generated order accuracy, as in Washington, that would warrant interruption of the automated order taking.

386. This modification is indeed already in line with what Coleman already suggests by disclosing that the human agents can “determine a new context” which

justifies interrupting the AOS 2 “[i]f the automated process is not completing successfully, such as due to poor interactions or understanding of the customer.” EX 1022, ¶[0089].

387. A POSA would understand that this combination better guards against order inaccuracies by giving Coleman’s human agent the explicit ability to compare the automatically generated order with the customer’s spoken order prior to the switchover to Coleman’s auditor-assist mode, as suggested by Washington.

388. A POSA would also understand that this combination would allow for more immediate intervention by Coleman’s human agent when the AOS is not accurately interpreting the customer’s order, than if the human agent were to only listen for audio-only indications that the AOS is inaccurately interpreting the customer’s order (e.g., the customer verbally indicates the AOS order is incorrect). Consequently, customer dissatisfaction and inefficiencies would be reduced.

389. It is therefore my opinion that a POSA would have found the limitations of [4i] and [4j] obvious based on the combined teachings of Coleman and Washington.

[4k] – updating the order processor based upon errors detected in the order by the auditor

390. Coleman does not expressly disclose this limitation.

391. Washington discloses this limitation in the form of the point-of-sale terminal 110 receiving an “input to override the menu items.” EX 1023, ¶[0070]. More specifically, in the operator assist mode, the point-of-sale terminal 110 displays the formatted transcript highlighting the menu items with the associated quantities and customizations. EX 1023, ¶¶[0038]; [0022]. The “operator 116 may review the menu items ... and then press a “Submit” button when satisfied that the menu items are correct (e.g., correspond to the actual menu items that the customer ordered).” *Id.* When the input to submit the order is not received, the point-of-sale terminal 110 determines whether it has received, via the order input device 132, an input to override the menu items input based on the formatted transcript. EX 1023, ¶[0070]. When the input to override the menu items is received, the point-of-sale terminal 110 receives input of the menu items with the associated quantities and customizations from the operator 116 via the order input device, and submits the order to the item ordering system 111. EX 1023, ¶[0070].

392. A POSA would have found it obvious to modify Coleman for at least the reasons discussed in Section XI.A. above.

393. In particular, a POSA would have found it obvious to modify Coleman to apply Washington’s way of verifying the accuracy of a speech-recognition generated order, by having Coleman’s human agent review the AOS generated order

for inaccuracies, as a “suitable reason” for interrupting the AOS order processing, while already listening in on the interaction.

394. A POSA would also understand that, in the course of interrupting the AOS order processing due to an order inaccuracy, the human agent would correct the order, as suggested by Washington.

395. This is precisely how the ‘706 Patent describes its updating of the order processor. If the auditor detects from its comparison of the AI generated order and the customer audio that the AI is not functioning properly, then the auditor can intervene and update the order. EX 1001, 13:23-40, 4:21-25. There is no difference between this and what a POSA would understand from Coleman’s and Washington’s teachings.

396. It is therefore my opinion that a POSA would have found the limitations of [4k] obvious based on the combined teachings of Coleman and Washington.

[4l] – deactivating the order processor with the auditor

397. The limitations of [4l] track to the same limitations of [1h] and are satisfied by Coleman for the same reasons as explained above.

[4m] – transmitting the generated order to an on-site employee to communicate with the customer, to collect payment from the customer, and to provide notifications to release the ordered product or service to the customer

398. Coleman discloses this limitation by disclosing that the customers are provided with “instructions on proceeding to pick up and/or pay for the ordered food and beverages,” while “the AOS 2 transmits the order information to the RIS 3 so that the order can be fulfilled by the RIS 3 and the restaurant’s human agents.” EX 1022, ¶[0080]. The human agents may be “at the restaurant site.” EX 1022, ¶[0036]. The RIS 3 is used to manage and execute transactions with customers, and can include a point-of-sale system that executes and records transactions, and includes a kitchen process management system. EX 1022, ¶[0035]. Coleman thus describes aspects of the traditional drive-thru operational model described in Section IV above.

399. A POSA would thus understand that the order is transmitted to the agent at the point-of-sale system “to communicate with the customer, to collect payment from the customer, and to provide notifications to release the ordered product ... to the customer.”

400. It is therefore my opinion that the limitations of [4m] are disclosed by Coleman.

C. Claim 5

wherein a business processor receives order information from the natural language processor (NLP) and performs the steps of:

associating the order items with one or more additional menu items or options associated with the order items; and

querying the customer about the one or more additional menu items or options associated with the order items, wherein the NLP updates the order based upon the customer response to the query

401. I understand that Claim 5 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

402. As discussed in [1e] above, Coleman discloses the “natural language processor” by the AOS 2. Coleman further discloses the “business processor” by the “ordering and conversation algorithms” programmed to “determine the optimal action [of the AOS 2] given the order status and inputs from the AOS 2.” EX 1022, ¶¶[0047]; [0070].

403. The ordering and conversion algorithms “may alter actions and the conversational response provided to a customer based on ... prior orders and interaction details associated with an identified customer ... to determine one or more recommendations associated with an order.” EX 1022, ¶[0058]. More specifically, “the items included in a current order may be used to identify one or more items to recommend” to the customer. EX 1022, ¶[0058].

404. Accordingly, examples of conversational responses of the ordering and conversion algorithms include: “suggesting one or more menu items for the current customer 6,” and “suggesting other modifications to the order.” EX 1022, ¶¶[0059];

[0079]. Coleman therefore further discloses “receives order information,” “associating the order items with one or more additional menu items or options associated with the order items,” and “querying the customer” about the additional menu items.

405. In the context of a drive-thru ordering system, a POSA would understand that the suggested items, if accepted by the customer, are added to the order. First, Coleman describes that its system “interacts with customers using voice input from the customers ... that allow the customers to place and review orders conversationally ... in the same manner as they would place an order if speaking to a human restaurant worker.” EX 1022, ¶[0020]. To accomplish this, the AOS’s “ordering and conversion algorithms ... process the inputs from the NLU,” and “transmit[] the order information to the RIS 3 so that the order can be fulfilled.” EX 1022, ¶¶[0034]; [0047]; [0055]; [0080]. The AOS 2 may “process many rounds of interactions with the customer 6 to complete the order.” EX 1022, ¶[0070]. Second, Coleman discloses that the drive-thru ordering area (“DOTA”) 1 includes a digital board where “updated items included in the order” may be presented to provide visual feedback to the customer. EX 1022, ¶[0031].

406. A POSA would thus understand that “the NLP updates the order based upon the customer response to the query” when the customer accepts the suggested order items.

407. It is therefore my opinion that the limitations of Claim 5 are disclosed by Coleman.

D. Claim 6

*providing a text-to-speech processor to
convert queries from the business processor
into audio to be communicated to a customer*

*processing the text with the text-to-speech
processor to create an audio communication
of the query generated by the business
processor*

408. I understand that Claim 6 depends from Claim 5, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

409. As discussed in claim 5 above, Coleman discloses the “queries from the business processor” and the “query generated by the business processor” by the ordering and conversation algorithms recommending or suggesting one or more menu items to the customer. EX 1022, ¶¶[0058]-[0059]; [0079]. Coleman further discloses the “text-to-speech processor” by a “Natural Language Generation (NLG) component that converts text to audible voice speech.” EX 1022, ¶[0034].

410. The remaining limitations are disclosed by Coleman because the potential actions of the AOS 2 include “generating a voice response through the NLG.” EX 1022, ¶¶[0048]-[0049]; [0070].

411. A POSA would understand that the NLG generated voice response includes recommending or suggesting menu items. First, Coleman describes that its system “interacts with customers using voice input from the customers ... that allow the customers to place and review orders conversationally ... in the same manner as they would place an order if speaking to a human restaurant worker.” EX 1022, ¶[0020]. Second, Coleman identifies an example “conversational response” of the AOS 2 as recommending or suggesting one or more menu items to the customer. EX 1022, ¶¶[0058]-[0059]; [0079].

412. A POSA would thus understand that the menu item recommendations are converted into audio responses to the customer.

413. It is therefore my opinion that the limitations of Claim 6 are disclosed by Coleman.

E. Claim 7

*reviewing an order by the auditor by
comparing the order generated by the
natural language processor to the
communication in the audio stream*

*updating the order based upon auditor
review*

414. I understand that Claim 7 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

415. As discussed in [1e] above, the combination of Coleman discloses “the order generated by the natural language processor.”

416. The rest of the limitations track those in [4i] and [4k] above, and its limitations are satisfied by the combination of Coleman and Washington for the same reasons as explained above. In particular, since the combination of Coleman and Washington disclose the narrower “by comparing the order ... to communication in the audio stream,” the combination necessarily discloses the broader “reviewing” the order.

417. It is therefore my opinion that the limitations of Claim 7 are disclosed in the combined teachings of Coleman and Washington.

F. Claim 8

*providing a text-to-speech processor to
convert queries from the business processor
into audio to be communicated to a customer*

*processing the text with the text-to-speech
processor to create an audio communication
of the query generated by the business
processor*

418. I understand that Claim 8 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

419. The limitations of claim 8 track those of claim 6, and are satisfied by Coleman for the same reasons as explained above.

G. Claim 9

alerting the auditor by the natural language processor (NLP) of any exceptions detected by the artificially intelligent order processing system, wherein the auditor can override the artificially intelligent order processing system based upon the detected exception

420. I understand that Claim 9 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

421. As discussed in [1e] above, Coleman discloses “exceptions detected” by detecting “suitable reasons” for the AOS 2 determining re-routing from the AOS 2 to the MOP 4 should occur, including “whether the customer 6 is speaking or providing inputs in an unclear manner,” “whether the monitored customer behavior is particularly unusual or negative,” whether connection has been lost or there is an “internal error,” or whether a human at the restaurant interrupted the connection. EX 1022, ¶¶[0071]-[0075].

422. Moreover, as discussed in [1g] above, Coleman further discloses “alerting the auditor ... of any exceptions detected” because, “if an active ordering process with the AOS 2 is re-routed to the MOP 4, then [] human agents are alerted to the re-routed interaction and communicate with the customer 6 through the DTOA 1 equipment and can complete the order in the usual manual manner.” EX 1022, ¶[0077].

423. As still further discussed in [1h] above, Coleman describes that the switches 5, which control the transfer (or re-routing) of the ordering process between the automatic workflow of the AOS 2 and the human-assisted workflow of the MOP 4, can be controlled by the human agents, via the administrative controller 12, to determine if orders “shall be taken (a) manually by human agents as part of the MOP 4 or (b) by the AOS 2.” EX 1022, ¶[0045]. The human agents are also “able to listen in to the automated interactions with a customer” and “interrupt[] the connection to the AOS 2 using the administrative controller 12 or by manually manipulating or interacting with the connectors or switches 5.” EX 1022, ¶[0045]. More particularly, the human agents can “determine a new context” which justifies interrupting the AOS 2 “[i]f the automated process is not completing successfully, such as due to poor interactions or understanding of the customer.” EX 1022, ¶[0089]. These circumstances correspond to the suitable reasons for transferring the ordering process from the AOS 2 to the MOP 4, which are discussed above in [1e].

424. A POSA would thus understand that the “auditor can override the artificially intelligent order processing system based upon the detected exception” because the human agent controlling switches 5 to re-route the ordering process from the AOS 2 to the MOP 4 upon listening in on the automated process and hearing suitable reasons to interrupt it.

425. It is therefore my opinion that the limitations of Claim 9 are disclosed by Coleman.

H. Claim 10

wherein the auditor deactivates the order processor in response to a detected change in customer traffic

426. I understand that Claim 10 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

427. As discussed in [1h] above, Coleman discloses that “the auditor deactivates the order processor,” by the human agent actuating the switches 5 to control whether the ordering process is initiated with the AOS 2 or the MOP 4. EX 1022, ¶¶[0023]; [0045]; [0063]; [0067].

428. Coleman discloses the remainder of this limitation because whether to use the AOS 2 or not can be determined based on the “volume of transactions with customers in the DTOA 1 and/or inside the restaurant within a recent time period,” “the relative volume of transactions ... as compared to the typical transaction volume,” “the current number of customers (or expected customers),” “the number of human agents currently available at the restaurant and a current number of transactions being performed inside of the restaurant,” and “an estimate or determination of the availability or responsiveness of the human agents operating

the MOP 4.” EX 1022, ¶[0067]. A POSA would therefore understand Coleman to suggest that, when the transaction volume does not exceed its available human agents, the AOS is deactivated in favor of the MOP, else the human agents would be overwhelmed.

429. It is therefore my opinion that the limitations of Claim 10 are disclosed by Coleman.

I. Claim 11

wherein the auditor deactivates the order processor in response to a detected change in staffing at the site

430. I understand that Claim 11 depends from Claim 4, which means that it contains all of the limitations from that claim. I have analyzed those limitations in view of the prior art above.

431. As discussed in [1h] above, Coleman discloses that “the auditor deactivates the order processor,” by the human agent actuating the switches 5 to control whether the ordering process is initiated with the AOS 2 or the MOP 4. EX 1022, ¶¶[0023]; [0045]; [0063]; [0067].

432. Coleman discloses the remainder of this limitation because whether to use the AOS 2 or not can be determined based on the “volume of transactions with customers in the DTOA 1 and/or inside the restaurant within a recent time period,” “the relative volume of transactions ... as compared to the typical transaction

volume,” “the current number of customers (or expected customers),” “the number of human agents currently available at the restaurant and a current number of transactions being performed inside of the restaurant,” and “an estimate or determination of the availability or responsiveness of the human agents operating the MOP 4.” EX 1022, ¶[0067]. A POSA would therefore understand Coleman to suggest that, when the transaction volume does not exceed its available human agents, the AOS is deactivated in favor of the MOP, else the human agents would be overwhelmed.

433. It is therefore my opinion that the limitations of Claim 11 are disclosed by Coleman.

XII. OBJECTIVE EVIDENCE OF NONOBVIOUSNESS

434. I understand that objective non-obviousness factors are to be considered in the obviousness analysis.

435. I am unaware of any basis for the Patent Owner to assert commercial success based on the technology patented in the ‘706 Patent. Moreover, to the extent that the Patent Owner would attempt to claim “commercial success,” it would appear not to stem from any novel features of the claimed invention since, as I explain above, each of the claimed concepts were known in the prior art.

436. I also understand that other objective factors of nonobviousness include “long-felt but unmet need,” “failed attempts by others,” or “unexpected results.” As

I explain above, the claimed concepts were so well-known in the prior art by the time of the purported invention that these factors do not appear to apply.

437. Accordingly, having considered all of the relevant factors for obviousness, it is my opinion that the challenged claims are unpatentable.

XIII. CONCLUSION

438. In my opinion, all the elements of the Challenged Claim limitations are disclosed, taught, or otherwise suggested by the prior art references discussed above, either alone or in combination, and that the claims are unpatentable in view of these prior art references.

439. I reserve the right to supplement my opinions to address any information obtained, or positions taken, based on any new information that comes to light throughout this or any other proceeding.

I declare under the penalty of perjury that the foregoing is true and accurate to the best of my ability.

Executed on: 12-20-2022



Boris Galitsky, Ph.D.

|