

**UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF TEXAS
MARSHALL DIVISION**

KAIFI LLC,

Plaintiff,

vs.

AMAZON.COM, INC.; AMAZON.COM
SERVICES LLC; and AMAZON WEB
SERVICES, INC.,

Defendants.

Case No. 2:24-cv-00542

**DEFENDANTS' AMENDED INVALIDITY CONTENTIONS AND ADDITIONAL
DISCLOSURES FOR U.S. PATENT NO. 8,930,196 PURSUANT TO LOCAL PATENT
RULES**

I. INTRODUCTION

Pursuant to the Court's Discovery Order (Dkt. No. 39), the Court's Docket Control Order (Dkt. No. 37), this District's Patent Local Rules, Plaintiff KAIFI LLC's ("KAIFI") Initial Infringement Contentions (served November 14, 2024), KAIFI's Amended Infringement Contentions (served January 10, 2024), and the Court's Order regarding Joint Motion for Leave to Amend Patent Local Rule Disclosures (Dkt. No. 59), Defendants Amazon.com, Inc., Amazon.com Services LLC, and Amazon Web Services, Inc. (collectively, "Defendants") submit the following Amended Invalidity Contentions and additional disclosures to KAIFI.

These Amended Invalidity Contentions address claims 1-20 of U.S. Patent No. 8,930,196, which are currently alleged by KAIFI to be infringed.

KAIFI's Amended Infringement Contentions are deemed to be its final contentions pursuant to Patent Local Rule 3-6. Should KAIFI supplement its Infringement Contentions with additional infringement theories with respect to the asserted claims (by Court order or otherwise), Defendants reserve the right to supplement their Invalidity Contentions to assert invalidity based on the additional infringement theories. Defendants also reserve the right to supplement their Invalidity Contentions in response to information learned in fact or expert discovery, including identification of additional prior art.

These Amended Invalidity Contentions are also submitted in compliance with and in satisfaction of the notice requirements of 35 U.S.C. § 282(c).

Defendants now serve these Amended Invalidity Contentions and Additional Disclosures.

II. PRELIMINARY MATTERS

Defendants' Amended Invalidity Contentions are based in whole or in part on its present understanding of the asserted claims and KAIFI's Infringement Contentions, including the priority

date of the '196 Patent as asserted by KAIFI in its Amended Infringement Contentions, which is May 28, 2008, and earlier invention date, which is alleged to be January 2007. That said, Defendants dispute that the asserted claims of the '196 Patent are entitled to the priority and/or invention dates that KAIFI has asserted in its Amended Infringement Contentions, and further dispute that the asserted claims are entitled to claim priority to the patent applications to which they purport to claim priority.

Defendants' Amended Invalidity Contentions are responsive at least to the same level of specificity as KAIFI's Initial and Amended Infringement Contentions. Defendants' Amended Invalidity Contentions may also take into account KAIFI's apparent claim constructions, to the extent KAIFI's constructions can be gleaned from KAIFI's Infringement Contentions. Such apparent constructions may be inconsistent with the constructions that Defendants ultimately will proffer in this case. By including prior art that would anticipate or render obvious the asserted claims of the '196 Patent based on KAIFI's disclosed and apparent claim constructions, or based on any other particular claim construction, Defendants are not adopting KAIFI's claim constructions, nor are Defendants admitting to the accuracy of any particular claim construction. Solely for the purpose of these Amended Invalidity Contentions, Defendants may, if necessary, apply alternative, and even inconsistent, claim construction positions. Defendants reserve all rights to further amend these Contentions after the Court issues its claim construction ruling, or if the Court permits KAIFI to amend its Infringement Contentions.

Defendants, however, do not concede that KAIFI's Initial and/or Amended Infringement Contentions provide the requisite level of specificity, and Defendants provide these Amended Invalidity Contentions without waiving any right to receive from KAIFI full and complete specific infringement contentions. Moreover, nothing herein admits in any way that any of the accused

devices or methods, or any of Defendants' other devices or methods, infringe any of the asserted claims.

B. Ongoing Investigation

Defendants' discovery and investigation in connection with this lawsuit is continuing. Thus, these Amended Invalidity Contentions are based on information obtained to date. Among other things, discovery is still underway, witnesses remain to be deposed, and the Court has not yet construed any terms of the patents-in-suit. Accordingly, Defendants' Amended Invalidity Contentions are subject to modification, amendment, or supplementation in accordance with the Court's Docket Control Order, the Local and Patent Rules of the Eastern District of Texas, and/or the Federal Rules of Civil Procedure as this action progresses and additional information is obtained.

As one example, Defendants identify below several systems that invalidate claims of the '196 Patent. Defendants have already or soon intend to issue subpoenas to the owners/makers/inventors of those systems, and expressly reserve the right to amend these contentions to add information obtained from these third parties.

III. DISCLOSURES PURSUANT TO P.R. 3-3

A. Disclosures Pursuant to P.R. 3-3(a)

Pursuant to P.R. 3-3(a), Defendants identify the following prior art to the '196 Patent. All of the prior art disclosed below relates to the field of speech recognition. As such, the below-disclosed pieces of prior art, collectively, are all relevant at least as background of the art to the '196 Patent, irrespective of the sub-section in which they are expressly cited. Defendants expressly reserve the right to at least use and rely on any of the prior art cited herein to establish or otherwise support Defendants' contentions as to what was known in the state of the art during the pertinent time frame for the '196 Patent.

1. Prior Art Patents and Published Patent Applications

The following prior art patents and published patent applications are relevant to the invalidity of the '196 Patent.

Reference	Number	Origin	Title	Inventor	Date of Issue or Publication
Aoyama '796	2003/0220796	US	Dialogue control system, dialogue control method and robotic device	Aoyama <i>et al.</i>	11/27/2003
Bellegarda '809	5,621,809	US	Computer program product for automatic recognition of a consistent message using multiple complimentary sources of information	Bellegarda <i>et al.</i>	03/26/1996
Bellegarda '774	5,502,774	US	Automatic recognition of a consistent message using multiple complimentary sources of information	Bellegarda <i>et al.</i>	03/26/1996
Braho '095	2007/0192095	US	Methods and systems for identifying errors in a speech recognition system	Braho <i>et al.</i>	08/16/2007
Braho '101	2007/0192101	US	Methods and systems for optimizing model adaptation for a speech recognition system	Braho <i>et al.</i>	08/16/2007
Catchpole '214	2006/030214	WO	A speech recognition circuit and method	Catchpole <i>et al.</i>	03/23/2006
Cohen '397	7,082,397	US	System for and method of creating and browsing a voice web	Cohen <i>et al.</i>	07/25/2006

Reference	Number	Origin	Title	Inventor	Date of Issue or Publication
Cohen '776	6,859,776	US	Method and apparatus for optimizing a spoken dialog between a person and a machine	Cohen <i>et al.</i>	02/22/2005
Cohen '576	6,560,576	US	Method and apparatus for providing active help to a user of a voice-enabled application	Cohen <i>et al.</i>	05/06/2003
De Souza '667	6,493,667	US	Enhanced likelihood computation using regression in a speech recognition system	De Souza <i>et al.</i>	12/10/2002
Gould '236	5,915,236	US	Word recognition system which alters code executed as a function of available computational resources	Gould <i>et al.</i>	06/22/1999
Hee-seok '562	100611562	KR	Computer security method using voice password	Jeong Hee-seok <i>et al.</i>	03/23/2005
Hee-seok '469	20060013469	KR	System and method of opening and closing a lock system using a voice password	Jeong Hee-seok <i>et al.</i>	02/10/2006
Hee-seok '242	100779242	KR	Speaker recognition methods of a speech recognition and speaker recognition integrated system	Jeong Hee-seok	11/26/2007
Hee-seok '951	20080026951	KR	Speech recognition methods for the robust distant-talking speech recognition system	Jeong Hee-seok	05/26/2008
Hee-seok '379	20050028379	KR	Computer security method using voice password	Jeong Hee-seok <i>et al.</i>	08/11/2006

Reference	Number	Origin	Title	Inventor	Date of Issue or Publication
Hogenhout '750	2409750	GB	Decoder for an automatic speech recognition system	Hogenhout <i>et al.</i>	06/07/2005
Huynh '801	2010/01798 01	US	Determining phrases related to other phrases	Huynh <i>et al.</i>	07/15/2010
Huynh '349	8,423,349	US	Filtering phrases for an identifier	Huynh <i>et al.</i>	04/16/2013
Jablokov '577	8,676,577	US	Use of metadata to post process speech recognition output	Jablokov <i>et al.</i>	03/18/2014
Jeon '261	2003/00972 61	US	Speech detection apparatus under noise environment and method thereof	Jeon <i>et al.</i>	05/22/2003
Jesensky '700	9,298,700	US	Determining similar phrases	Jesensky <i>et al.</i>	03/29/2016
Jesensky '644	8,706,644	US	Mining phrases for association with a user	Jesensky <i>et al.</i>	04/22/2014
Jesensky '643	8,706,643	US	Generating and suggesting phrases	Jesensky <i>et al.</i>	04/22/2014
Jesensky '770	9,569,770	US	Generating constructed phrases	Jesensky <i>et al.</i>	02/14/2017
Jiang '353	6,539,353	US	Confidence measures using sub-word-dependent weighting of sub-word confidence scores for robust speech recognition	Jiang <i>et al.</i>	03/25/2003
Junqua '990	5,677,990	US	System and method using N-best strategy for real time recognition of continuously spelled names	Junqua	10/14/1997

Reference	Number	Origin	Title	Inventor	Date of Issue or Publication
Kepuska '789	2003/0220789	US	Dynamic time warping of speech	Kepuska	11/27/2003
Kepuska '717	7,085,717	US	SCORING AND RE-SCORING DYNAMIC TIME WARPING OF SPEECH	Kepuska <i>et al.</i>	08/01/2006
Kim '837	2006/0122837	US	Voice interface system and speech recognition method	Kim <i>et al.</i>	06/08/2006
Kim '207	2006/0136207	US	Two stage utterance verification device and method thereof in speech recognition system	Kim <i>et al.</i>	06/22/2006
Konopka '188	2008/0059188	US	Natural Language Interface Control System	Konopka <i>et al.</i>	03/06/2008
Konopka '635	7,447,635	US	Natural Language Interface Control System	Konopka <i>et al.</i>	11/04/2008
Lahti '711	2005/0256711	US	Detection of end of utterance in speech recognition system	Lahti	11/17/2005
Lam '664	2006/0074664	US	System and method for utterance verification of chinese long and short keywords	Lam <i>et al.</i>	04/06/2006
Lee '574	6,151,574	US	Technique for adaptation of hidden markov models for speech recognition	Lee <i>et al.</i>	11/21/2000
Millman '413	2008/0140413	US	Synchronization of audio to reading	Millman <i>et al.</i>	06/12/2008

Reference	Number	Origin	Title	Inventor	Date of Issue or Publication
Mozer '392	7,418,392	US	System and method for controlling the operation of a device by voice commands	Mozer <i>et al.</i>	08/26/2008
Nadas '488	4,926,488	US	Normalization of speech by adaptive labelling	Nadas <i>et al.</i>	05/15/1990
Paden '454	8,301,454	US	Methods, apparatuses, and systems for providing timely user cues pertaining to speech recognition	Paden	10/30/2012
Qu '786	6,965,786	US	Annunciators for voice and data applications in wireless communication devices	Qu <i>et al.</i>	11/15/2005
Sang-hoon '493	KR20060064493	KR	Voice interface system and method	Sang-Hoon <i>et al.</i>	09/11/2006
Sherwood '768	6,163,768	US	Non-interactive enrollment in speech recognition	Sherwood <i>et al.</i>	12/19/2000
Sherwood '498	6,212,498	US	Enrollment in speech recognition	Sherwood <i>et al.</i>	04/03/2001
Smyth '453	0695453 B1	EP	Connected speech recognition	Smyth <i>et al.</i>	10/06/1999
Squires '043	6,092,043	US	Apparatuses and method for training and operating speech recognition systems	Squires <i>et al.</i>	07/18/2000
Stubley '045	6,092,045	US	Method and apparatus for speech recognition	Stubley <i>et al.</i>	07/18/2000
Takiguchi '832	2006/0122832	US	Signal enhancement and speech recognition	Takiguchi <i>et al.</i>	06/08/2006

Reference	Number	Origin	Title	Inventor	Date of Issue or Publication
Terrell '109	8,510,109	US	Continuous speech transcription performance indication	Terrell <i>et al.</i>	08/13/2013
Terrell '261	8,352,261	US	Use of intermediate speech transcription results in editing final speech transcription results	Terrell <i>et al.</i>	01/08/2013
Toyoda '348	6,452,348	US	Robot control device, robot control method and storage medium	Toyoda	09/17/2002
Tran '140	6,070,140	US	Speech recognizer	Bao Q. Tran	05/30/2000
Tunstall-Pedoe '308	7,013,308	US	Knowledge storage and retrieval system and method	William Tunstall-Pedoe	03/14/2006
Tunstall-Pedoe '656	2007/0055656	US	Knowledge repository	William Tunstall-Pedoe	03/08/2007
White '264	8,352,264	US	Corrective feedback loop for automated speech recognition	Marc White <i>et al.</i>	01/08/2013
White '653	6,785,653	US	Distributed voice web architecture and associated components and methods	James White <i>et al.</i>	08/31/2004
Zlokarnik '630	6,594,630	US	Voice-activated control for electrical device	Zlokarnik <i>et al.</i>	07/15/2003

Defendants also identify any patent or patent application publications that are practiced by or otherwise related to the systems discussed below, for example the patents cited in the Dragon NaturallySpeaking claim chart.

2. Non-Patent Prior Art Publications

The following non-patent prior art publications are relevant to the invalidity of the asserted claims of the '196 Patent:

Reference	Title	Author	Publication Date	Publisher
Abdulla	HMM-based techniques for speech segments extraction	W. Abdulla	2002	Scientific Programming
Bou-Ghazale	Hands-Free Voice Activation of Personal Communication Devices	Sahar E. Bou-Ghazale <i>et al.</i>	2000	IEEE
Bou-Ghazale II	A robust endpoint detection of speech for noisy environments with application to automatic speech recognition	S.E. Bou-Ghazale <i>et al.</i>	2002	IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.4
De Fouw	User-centred concept design of a speech controlled home system	De Fouw, J. <i>et al.</i>	1999	IPO
Fludd	Multi Remote Appliance Contoller	Keith M. Fludd <i>et al.</i>	2001	Proceedings of the IEEE 27 th Annual Northeast Bioengineering Conference
Gales	The Application of Hidden Markov Models in Speech Recognition	Mark Gales <i>et al.</i>	2007/2008	Foundations and Trends in Signal Processing
Holmes	Speech Synthesis and Recognition	John Holmes and Wendy Holmes	2001	Taylor & Francis
Jouvet	HYPOTHESIS DEPENDENT THRESHOLD SETTING	Jouvet <i>et al.</i>	1999	IEEE

Reference	Title	Author	Publication Date	Publisher
	FOR IMPROVED OUT-OF-VOCABULARY DATA REJECTION			
Juang	Automatic Speech Recognition – A Brief History of the Technology Development	Juang <i>et al.</i>	2005	
Kepuska I	Wake-Up-Word Speech Recognition Application for First Responder Communication Enhancement	Veton Këpuska <i>et al.</i>	2006	Proc. of SPIE
Kepuska II	LEADING AND TRAILING SILENCE IN WAKE-UP-WORD SPEECH RECOGNITION	Veton Këpuska <i>et al.</i>	May 2006	Proceedings of the 2006 IBMS Conference
Kepuska III	A novel Wake-Up-Word speech recognition system, Wake-Up-Word recognition task, technology and evaluation	Veton Këpuska <i>et al.</i>	December 2009	Nonlinear Analysis: Theory, Methods & Applications
Kuhne	Voice Activation Using Prosodic Features	M. Kuhne <i>et al.</i>	2004	INTERSPEECH 2004 – ICSLP 8th International Conference on Spoken Language Processing ICC
Kwak	Speaker identification and verification for intelligent service robots	K.C. Kwak	2007	Proceedings of the 2007 International Conference on Artificial Intelligence, ICAI

Reference	Title	Author	Publication Date	Publisher
Lee III	Automatic Speech Recognition: The Development of the SPHINX System	Kai-Fu Le	1989	The Springer International Series in Engineering and Computer Science
Lee I	Competing models-based text-prompted speaker independent verification algorithm	Heungkyu Lee <i>et al.</i>	2006	Speech communication
Lee II	Voice Code Verification System Supporting Multi-Modal Speech Interaction Using ASR and TTS	Heungkyu Lee <i>et al.</i>	2004	ICAT
Lee IV	Acoustic modeling for large vocabulary speech recognition	C.H. Lee <i>et al.</i>	1990	Computer and Speech Language
Mainardi	Controlling a prosthetic arm with a throat microphone	E. Mainardi <i>et al.</i>	2007	2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France
Manos	A study on out of vocabulary word modeling for a segment based keyword spotting system	Manos	1996	MIT
Paul	Speech Recognition Using Hidden Markov Models	D.B. Paul	1990	The Lincoln Laboratory Journal
Rabiner I	Recognition of Isolated Digits Using Hidden Markov Models With Continuous Mixture Densities	L.R. Rabiner <i>et al.</i>	July-August 1985	AT&T Technical Journal

Reference	Title	Author	Publication Date	Publisher
Rabiner II	A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition	Lawrence R. Rabiner	February 1989	Proceedings of the IEEE
Rabiner III	Fundamentals of Speech Recognition	Lawrence Rabiner; Biing-Hwang Juang	1993	Prentice-Hall International, Inc.
Ramachandran	Modern Methods of Speech Processing	R. P. Ramachandran et al.	1995	Kluwer Academic Publishers
Rudnick	The Design of Spoken Language Interfaces	Alexander I. Rudnick	March 1990	School of Computer Science, Carnegie Mellon University
Schmandt	Voice Communication with Computers	Chris Schmandt	1994	Van Nostrand Reinhold
Stiles	Testing and Improvement of the Triple Scoring Method for Applications of Wake up Word Technology	Andrew Stiles <i>et al.</i>	2007	Amalthea Reu Site
Wiggers	AUTOMATIC SPEECH RECOGNITION USING HIDDEN MARKOV MODELS	Ir. P. Wiggers <i>et al.</i>	September 2003	IN4012TU Real-time AI & Automatische Spraakherkennig
Young I	Token Passing: a Simple Conceptual Model for Connected Speech Recognition Systems	S.J. Young <i>et al.</i>	1989	Cambridge University
Young II	HMMS AND RELATED SPEECH RECOGNITION TECHNOLOGIES	Steve Young	2007	Springer Handbook on Speech Processing and Speech Communication

3. Items Offered for Sale or Publicly Used or Known

Numerous systems are prior art to the '196 Patent under at least 35 U.S.C. §§ 102(a), (b), (e), and/or (g), and anticipate and/or render obvious all asserted claims of the '196 Patent. These systems include physical implementations of prior art patents, patent applications, and publications identified herein that were known, made, used, sold, offered for sale, or imported into this country prior to the effective filing date of the '196 Patent, including those described by the prior art patents, patent applications, and publications listed in the previous section.

Defendants further identify the following products that were sold, offered for sale, publicly used, or known that incorporate the claimed elements of the '196 Patent.

No.	System or Product	Date of First Offer or Public Disclosure	Circumstances
1	HTK Toolkit	Earliest versions were released dating back to 1989. Charted Version 3.4 was released in December 2006.	The Hidden Markov Model Toolkit (HTK) is a portable toolkit for building and manipulating hidden Markov models. HTK was originally developed at the Machine Intelligence Laboratory (http://mi.eng.cam.ac.uk/ , formerly known as the Speech Vision and Robotics Group) of the Cambridge University Engineering Department (CUED) where it has been used to build CUED's large vocabulary speech recognition systems. In 1993 Entropic Research Laboratory Inc. acquired the rights to sell HTK and the development of HTK was fully transferred to Entropic in 1995 when the Entropic Cambridge Research Laboratory Ltd was established. HTK was sold by Entropic until 1999 when Microsoft bought Entropic. See https://htk.eng.cam.ac.uk/ .
2	Microsoft speech recognition	No later than 2005	Microsoft speech recognition is and was a commercially available speech recognition program developed and distributed by Microsoft Corporation. Microsoft speech recognition was

No.	System or Product	Date of First Offer or Public Disclosure	Circumstances
			distributed along with each Microsoft XP and Vista operating system as well as freely downloadable from the web. <i>See, e.g.,</i> https://web.archive.org/web/20050228093406/http://www.microsoft.com/speech/default.mspx ; https://web.archive.org/web/20070205145119/http://www.microsoft.com/speech/default.mspx .
3	IBM ViaVoice	Charted version was released in August 2002. Earlier versions may also invalidate.	IBM designed and developed several iterations of ViaVoice (hereinafter, “ViaVoice”) that allowed a computer user to talk to and instruct their computer regarding various tasks. IBM began advertising ViaVoice by no later than 1986. By August 2002, version 1, release 10, modification 0 of IBM ViaVoice for Windows Personal Edition US English (program number 5724-C96) was available for purchase.
4	Dragon NaturallySpeaking	Earliest versions were released dating back to 1997. Charted Version 9 was released in July 2006.	By July 1997, Dragon Systems had launched Dragon NaturallySpeaking, a continuous speech & voice recognition program for general-purpose use with a vocabulary of 23,000 words. <i>See</i> https://www.dragon-medical-transcription.com/history_speech_recognition.html .
5	“Wake-Up-Word” or “OnWord™” Spotting Technology by ThinkEngine Networks Inc.	2000s	In the early 2000s, ThinkEngine Networks invented, designed and developed unique solution to “Wake-Up-Word” or “OnWord™” Spotting Technology. <i>See</i> https://www.fit.edu/faculty-profiles/emeritus/kepuska-veton/ .
6	Yap Speech Cloud	By September 2007	Founded in 2006, Yap pioneered the world’s first high accuracy, fully automated cloud-based speech recognition platform. The Company offers a simple, scalable and cost effective approach to implementing advanced speech

No.	System or Product	Date of First Offer or Public Disclosure	Circumstances
			transcription services. By September 2007, Yap voice-to-text translation was available for cell phones. <i>See</i> https://techcrunch.com/2007/09/17/yap-voice-to-text-translation-on-your-cell/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xILmNvbS8&guc e_referrer_sig=AQAAALG7_P8uHQKT KoooC_P52nEh_nDvDCQ6N4orjeqNzv M2jbR-B2WbqamD_UVG2h2mFyNX2OeDnr-WjFtd5d7mOvj3pu3oUAJTQQHWpkzVvpWxc0FOz9pPB-Fhw21X-AMAL49S65IeglcNgKbHQNg0njsjddXE2vMKh4YZ6-NyuL2u .
7	True Knowledge Answer engine	2007	In 2007, True Knowledge released a commercial, open-domain question-answering platform.
8	Nomadic Radio	December 1995	Nomadic Radio, which was active December 1995 to December 2000, explored speech interfaces for communication purposes in a wearable computer. Nomadic radio was developed by Chris Schmandt and Nitin Sawhney.
9	Put That There	March 15, 1982	“Put That There” is a voice and gesture interactive system implemented at the Architecture Machine Group at MIT, and developed by Richard Bolt, Chris Schmandt, Eric A. Hulteen.
10	Ford/Microsoft Sync	2007	Ford released the Sync system, a speech recognition system for its cars and powered by Microsoft speech recognition technology, in 2007.
11	CMU Sphinx	1986 and on	The first version of CMU Sphinx was released in 1986. On information and belief, certain versions of Sphinx included wake word technology, at least by 1990.
12	ETRI Voice Navigation Engine	2006	ETRI made public its voice navigation system in 2006.

No.	System or Product	Date of First Offer or Public Disclosure	Circumstances
13	Open Sesame	By 2004	Biometric voice authentication software released by KoreaPowerVoice.
14	KAIST Albert Hubo	By 2006	Smart robot released by KAIST.
15	KoreaPowerVoice smart home systems	By 2008, likely earlier	
16	Clarion AutoPC and Joyride	1998	The Clarion AutoPC debuted at CES in January 1998. Microsoft technology powered the AutoPC.
17	Scansoft Automotive	2003	ScanSoft offered a suite of technologies and services to speech enable the devices and systems including automobiles.
18	The Multi Remote	2001	The Multi Remote is a project that automates the blinds of a window and a coffee maker through voice activation and infrared via remote control. An activation word, or preset call command, is required for the coffee maker.

Defendants continue to investigate the identities of the individuals or entities involved in the first sale, offer for sale, or public use of prior art products that incorporate the claimed elements of the '196 Patent, and Defendants will continue to produce additional documents relating to the invalidity of the asserted claims by those products as they become available (including those from third parties). Defendants also note that several companies not listed here may have prior art systems, such as the owners of the patents or patent applications listed above. Defendants reserve the right to provide additional invalidity contention information as additional information becomes available. Defendants also reserve the right to rely on any of the documents associated with these systems to the extent those documents qualify as printed publications.

Regarding the KoreaPowerVoice systems listed above, Defendants continue to investigate whether these systems constitute prior art under any subsection of 35 U.S.C. § 102,

and are actively seeking discovery from KAIFI on this issue. Based on KAIFI's limited production of KoreaPowerVoice documents, the Defendants can tell that KoreaPowerVoice had several products in use and on sale before the '196 Patent's earliest possible priority date. If those products were used and/or for sale in the United States, they may constitute prior art and could render the claims obvious. For example, KoreaPowerVoice appears to have sold a voice authentication product (Open Sesame) that when combined with one of KoreaPowerVoice's smart home systems (several of which appear to have been in use prior to the 196 Patent's earliest possible priority date), the combination would appear to render the claims obvious. Please note that detailed descriptions of the "Circumstances" were not included because KAIFI has marked such information Highly Confidential – Attorneys' Eyes Only (and, presumably, KAIFI is already aware of the circumstances).

4. Prior Inventions under 35 U.S.C. § 102(a)/ 102(g).

Defendants identify the following additional prior art systems under 35 U.S.C. § 102(a)/ 102(g).

No.	System or Product	Persons Involved/Circumstances
1	HTK Toolkit	The Hidden Markov Model Toolkit (HTK) is a portable toolkit for building and manipulating hidden Markov models. HTK was originally developed at the Machine Intelligence Laboratory (http://mi.eng.cam.ac.uk/ , formerly known as the Speech Vision and Robotics Group) of the Cambridge University Engineering Department (CUED) where it has been used to build CUED's large vocabulary speech recognition systems. In 1993 Entropic Research Laboratory Inc. acquired the rights to sell HTK and the development of HTK was fully transferred to Entropic in 1995 when the Entropic

No.	System or Product	Persons Involved/Circumstances
		Cambridge Research Laboratory Ltd was established. HTK was sold by Entropic until 1999 when Microsoft bought Entropic. See https://htk.eng.cam.ac.uk/ .
2	Microsoft speech recognition	Microsoft speech recognition is and was a commercially available speech recognition program developed and distributed by Microsoft Corporation. By no later than 2007, Microsoft speech recognition was distributed along with each Microsoft XP and Vista operating system as well as freely downloadable from the web. See, e.g., https://web.archive.org/web/20050228093406/http://www.microsoft.com/speech/default.aspx ; https://web.archive.org/web/20070205145119/http://www.microsoft.com/speech/default.aspx .
3	IBM ViaVoice	IBM designed and developed several iterations of ViaVoice (hereinafter, “ViaVoice”) that allowed a computer user to talk to and instruct their computer regarding various tasks. IBM began advertising ViaVoice by no later than 1986. By August 2002, version 1, release 10, modification 0 of IBM ViaVoice for Windows Personal Edition US English (program number 5724-C96) was available for purchase.
4	Dragon NaturallySpeaking	By July 1997, Dragon Systems had launched Dragon NaturallySpeaking, a continuous speech & voice recognition program for general-purpose use with a vocabulary of 23,000 words. See https://www.dragon-medical-transcription.com/history_speech_recognition.html .
5	“Wake-Up-Word” or “OnWord™” Spotting Technology by ThinkEngine Networks Inc.	In the early 2000s, ThinkEngine Networks invented, designed and developed unique solution to “Wake-Up-Word” or “OnWord™” Spotting Technology. See

No.	System or Product	Persons Involved/Circumstances
		https://www.fit.edu/faculty-profiles/emeritus/kepuska-veton/ .
6	Yap Speech Cloud	Founded in 2006, Yap pioneered the world's first high accuracy, fully automated cloud-based speech recognition platform. The Company offers a simple, scalable and cost effective approach to implementing advanced speech transcription services. By September 2007, Yap voice-to-text translation was available for cell phones. <i>See</i> https://techcrunch.com/2007/09/17/yap-voice-to-text-translation-on-your-cell/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2x1LmNvbS8&guc e_referrer_sig=AQAAALG7_P8uHQKT KoooC_P52nEh_nDvDCQ6N4orjeqNzv M2jbR-B2WbqamD_UVG2h2mFyNX2OeDnr-WjFtd5d7mOvj3pu3oUAJTQQHWpkzVvpWxc0FOz9pPB-Fhw21X-AMAL49S65IeglcNgKbHQNg0njsjddXE2vMKh4YZ6-NyuL2u .
7	True Knowledge Answer engine	In 2007, True Knowledge released a commercial, open-domain question-answering platform.
8	Nomadic Radio	Nomadic Radio, which was active December 1995 to December 2000, explored speech interfaces for communication purposes in a wearable computer. Nomadic radio was developed by Chris Schmandt and Nitin Sawhney.
9	Put That There	"Put That There" is a voice and gesture interactive system implemented at the Architecture Machine Group at MIT, and developed by Richard Bolt, Chris Schmandt, Eric A. Hulteen. The initial version was released in 1980 and the second version by 1982.
10	Ford/Microsoft Sync	Ford released the Sync system, a speech recognition system for its cars and

No.	System or Product	Persons Involved/Circumstances
		powered by Microsoft speech recognition technology, in 2007.
11	CMU Sphinx	The first version of open source speech recognition software CMU Sphinx was released in 1986. On information and belief, certain versions of Sphinx included wake word technology, at least by 1990.
12	ETRI Voice Navigation Engine	ETRI made public its voice navigation system in 2006.
13	Open Sesame	Biometric voice authentication software released by KoreaPowerVoice.
14	KAIST Albert Hubo	
15	KoreaPowerVoice smart home systems	
16	Clarion AutoPC and Joyride	The Clarion AutoPC debuted at CES in January 1998. Microsoft technology powered the AutoPC.
17	Scansoft Automotive	By 2003, ScanSoft offered a suite of technologies and services to speech enable the devices and systems including automobiles.

5. Inventorship

To the extent the idea of a wake word came from someone other than the named inventors, the '196 Patent is invalid for improper inventorship. For example, it appears that the idea of a wake word (the purported inventive feature of the '196 patent, *see* '196 Patent at 2:43-3:20), may have come from someone at M&Soft instead of the '196 patent's named inventors. *See, e.g.*, KAIFI0008732 (request for proposals from M&SOFT, Inc.). Defendants are actively seeking discovery from KAIFI on this issue.

B. Disclosures Pursuant to P.R. 3-3(b)

Pursuant to P.R. 3-3(b), the prior art at issue in these contentions anticipates and/or renders obvious the '196 Patent claims as follows:

Reference/Combination	Claims Anticipated	Claims Rendered Obvious
Tran '140	1-2, 4-9, 16, 17-20	1-20
Cohen '776	1, 8-9, 17-18	1-20
Konopka '188/'635	1-9, 16-20	1-20
Bou-Ghazale + Konopka '188/'635		1-20
Bou-Ghazale + Young I		1-20
HTK Toolkit		1-20
Microsoft Speech Recognition		1-20
IBM ViaVoice		1-20
Dragon NaturallySpeaking		1-20
Kepuska I + Kepuska '789		1-20
Qu '786		1-20
Millman '413		1-20
Sherwood '768		1-20
Mozer '392		1-20
Hogenhout '750		1-20
Lahti '711		1-20
Catchpole '214		1-20
Takiguchi '832		10-15
Gould '236		10-15
Jablokov '577		10-15
Lee I		2-8
Jouvet		2-7
Lam '664		2-7, 16, 18
Cohen '397	1, 8-9, 17-18	1-20
Bellegarda '809		2-4
Braho '095		2-4, 18
Braho '101		2-4, 18
De Souza '667		2, 4, 18
Gould '236	1, 9, 17-18	1-20
Hee-seok '242		2-5, 8, 18

Reference/Combination	Claims Anticipated	Claims Rendered Obvious
Hee-seok '951		2-6, 16
Huyhn '801		8, 16, 18
Huyhn '349		8, 16, 18
Jesensky '700		8, 16, 18
Jiang '353		2-5, 18
Jesensky '644		8, 16, 18
Kepuaka '789		1-20
Kepuska '717		1-20
Paden '454		2-4, 16, 18
Sherwood '498		1-20
Squires '043		1-20
Terrell '109		2-4, 16, 18
Terrell '261		2-4, 16, 18
White '264		2-4, 16, 18
Zlokarnik '630		8, 16, 18
Gales		2-4, 8
Kepuska II		1-20
Kepuska III		1-20
Lee II		2, 4-8
Rabiner II		2, 4
Stiles		1-20
Young II		2, 4-5
Nomadic Radio		1-20
Put That There		1-20
CMU Sphinx		1-20
Ford Sync		1-20
Auto PC		1-20
ETRI Speech Recognition System		1-20
Kim '837		1-20

Additional secondary references for the reference/combinations identified above are presented and discussed in the attached claim charts, which are fully incorporated herein.

The claims of the ‘196 Patent are entitled to the following priority date: August 22, 2012.

As noted in the chart above, the prior art renders all claims of the ‘196 Patent obvious. Defendants incorporate by reference all obviousness rationales and discussion set forth in the attached claim charts. The following additional obviousness rationales apply.

It would have been obvious to a person of skill in the art (“POSITA”) to combine any or all of the references for the systems or products identified above together, such that the combined teachings of the references render the asserted claims obvious. For example, it would have been obvious to combine one or more of the Dragon NaturallySpeaking references together. Similarly, it would have been obvious to combine one or more of the Microsoft Speech Recognition references together, etc. The references all concern the same products and subject matter and thus there is an implied suggestion to read the references together as a unified whole. Such combinations would be nothing more than taking a teaching from one reference and improving the teachings of a similar reference—and in these cases, a related reference—in the same way.

It would have been obvious to implement token passing (e.g., as taught in Young I, Young II, Hogenhout ‘750, Konopka ‘188, Konopka ‘635, Lahti ‘711, Catchpole ‘214) in any reference that uses hidden Markov models (HMMs) for speech recognition. As evidenced by the token passing prior art, token passing was a highly memory efficient solution to a well-known problem in HMM-based speech recognition systems; the storage of back-pointers that allows for finding the best word sequence hypothesis once all speech data has been processed. A POSITA would have been motivated to implement token passing to conserve processing power and resources, and to make the speech recognition system more memory efficient.

It would have been obvious to a POSITA to use a speech recognition network with models for silence, noise (such as extraneous noise or out of vocabulary words), and whatever word the network/system is searching for (such as an in-vocabulary or keyword) (e.g., the recognition networks taught by Konopka '188, Konopka '635, Bou Ghazale, Gould '236, Hee-seok '951) in any reference that uses HMMs. A POSITA would have been motivated to do so to improve recognition accuracy. The basic principle behind HMM-based speech recognition systems is that an utterance is compared with a number of speech models in order to find the model that best matches the utterance. By modeling both the sound (e.g., a particular word such as a wake word) the system is looking for as well as sounds the system is not looking for (e.g., silence, other words, and other noises like sneezing), the system is better able to recognize the speech it is looking for, and accuracy is improved.

Similarly, it would have been obvious to incorporate one or more anti-models (e.g., as taught by Konopka '188, Konopka '635, Lee I, Jouvett, and Lam '664) as well as likelihood/confidence measures of the same in any reference that uses an HMM model. A POSITA would have been motivated to do so to improve recognition accuracy. The basic principle behind HMM-based speech recognition systems is that an utterance is compared with a number of speech models in order to find the model that best matches the utterance. By modeling both the sound (e.g., a particular word such as a wake word) the system is looking for as well as sound the system is not looking for (e.g., anti-sound or garbage), the system is better able to recognize the speech it is looking for, and accuracy is improved. In the case of anti-models that can be created from HMMs, adding an anti-model does not even require development and training of another model. This is an additional reason a POSITA would be motivated to incorporate these types of anti-models into an existing speech recognition system.

It would have also been obvious to incorporate one or more garbage and/or silence models as well as likelihood/confidence measures of the same in any reference that uses an HMM model. As discussed above, by modeling both the sound (e.g., a particular word such as a wake word) the system is looking for as well as sounds the system is not looking for (e.g., garbage or silence), the system is better able to recognize the speech it is looking for. A POSITA would have been motivated to do so to improve recognition accuracy.

It would have been obvious to use any speech recognition system in the prior art (e.g., the speech recognition systems taught in Bellegarda '809, Braho '095, Braho '101, Catchpole '214, Cohen '397, Cohen '776, Cohen '576, De Souza '667, Gould '236, Hee-seok '562, Hee-seok '469, Hee-seok '242, Hee-seok '951, Hogenhout '750, Jablovskiy '577, Kepuska '789, Kepuska '717, Konopka '188, Konopka '635, Lahti '711, Lam '664, Millman '413, Mozer '392, Paden '454, Sherwood '768, Sherwood '498, Smyth '453, Squires '043, Terrell '109, Terrell '261, Tran '140, White '264, White '653, Zlokarnik, Yap system, True Knowledge System, HTK Toolkit system, Microsoft Speech Recognition, IBM ViaVoice, Dragon NaturallySpeaking, Bou-Ghazale, Gales, Jouvett, Aoyama '796, Jeon '261, Kim '837, Kim '207, Lee '574, Nadas '488, Sang-hoon '493, Stubble '045, Toyoda '348, Abdulla, Bou-Ghazale II, De Fouw, Fludd, Manos, Ramachandran, Schmandt, Clarion AutoPC, CMU Sphinx, ETRI Voice Navigation, Ford Sync, Put That There, and Nomadic Radio) for recognition/detection of a particular word such as a wake word, as these already existing systems are known to successfully recognize (a lot of) particular words. Whether recognizing a single word or many words (including the single word), they are both attempting to achieve the same goals with similar technology. And as discussed below and in the attached claim charts as well as in Defendants' eligibility contentions, the idea of recognizing a particular word

to wake up a device or a speech recognition system was well known in the art going back to at least the 1990s. *See, e.g.*, Ramachandran at 124; Rudnický at 14-15; Schmandt at 165.

It would have also been obvious to have substituted any one of the speech recognition systems in the prior art (e.g., the speech recognition systems taught in Bellegarda '809, Braho '095, Braho '101, Catchpole '214, Cohen '397, Cohen '776, Cohen '576, De Souza '667, Gould '236, Hee-seok '562, Hee-seok '469, Hee-seok '242, Hee-seok '951, Hogenhout '750, Jablokov '577, Kepuska '789, Kepuska '717, Konopka '188, Konopka '635, Lahti '711, Lam '664, Millman '413, Mozer '392, Paden '454, Sherwood '768, Sherwood '498, Smyth '453, Squires '043, Terrell '109, Terrell '261, Tran '140, White '264, White '653, Zlokarnik, Yap system, True Knowledge System, HTK Toolkit system, Microsoft Speech Recognition, IBM ViaVoice, Dragon NaturallySpeaking, Bou-Ghazale, Gales, Jouvét, Aoyama '796, Jeon '261, Kim '837, Kim '207, Lee '574, Nadas '488, Sang-hoon '493, Stubbley '045, Toyoda '348, Abdulla, Bou-Ghazale II, De Fouw, Fludd, Manos, Ramachandran, Schmandt, Clarion AutoPC, CMU Sphinx, ETRI Voice Navigation, Ford Sync, Put That There, and Nomadic Radio) for any other speech recognition system in the prior art. This substitution would have been obvious because of the similarity of the speech recognition systems to each other, and the fact that they were all attempting to achieve the same goals with similar technology.

It would have been obvious to implement a wake word (or wake words) in any speech recognition system in the prior art to save processing power by only having speech recognition active upon recognition of the wake word (e.g., as taught by Cohen '397, Cohen '776, Cohen '576, Hee-seok '562, Hee-seok '469, Kepuska '789, Konopka '188, Konopka '635, Millman '413, Mozer '392, Qu '786, Sherwood '768, Sherwood '498, Squires '043, Tran '140, HTK Toolkit system, Microsoft Speech Recognition system, IBM ViaVoice system, Dragon NaturallySpeaking

system, Bou-Ghazale, Kepuska I, Kepuska II, Kepuska III, Stiles, Ramachandran; Rudnick, Schmandt, Kim '837, Sang-hoon '93, De Fouw, Fludd, Manos, Clarion AutoPC, CMU Sphinx, Nomadic Radio, Put That There). It would also have been obvious to implement a wake word to offer an easy way to turn the system on and off, for example, for privacy or to minimize recognition errors when the system is not in active use. As Schmandt explains:

Explicit Indication of Attention

In many acoustic environments background noise is highly variable, which results in spurious insertion errors when short, loud sounds are picked up by the microphone. Insertion errors also occur if the user speaks with another person such as a telephone caller or a visitor. Such errors can be minimized by disabling recognition when it is not in active use.

One simple means of disabling recognition is the addition of a mechanical switch on the microphone cable. A hand-held device employing speech recognition may require a push-to-talk button in a noisy environment. If the user is at a computer, a mouse button may be used the same way. But any mechanical switch detracts from the benefit of being able to use recognition while one's hands are otherwise occupied.

Another alternative is to use voice commands such as "stop listening" and "pay attention." After the user says the first of these commands, recognition results are discarded until the second is spoken. Note that it is important to use phrases for this software switch that are unlikely to be spuriously recognized themselves. Another disadvantage of the stop listening approach is that the user is likely to forget that recognition is disabled until speaking several times with no apparent results.

Some versions of CMU's Sphinx recognizer used a key word to indicate attention and trigger recognition: "Computer, when is my meeting?" "Computer, cancel the meeting." Recognition control can also be implemented using the keyboard or mouse if the user works at a computer console. Finally, it is possible to use the direction in which the user speaks as the focus mechanism. The Media Lab's Conversational Desktop project [Schmandt *et al.* 1985] used such an approach; the user turned toward the screen when he or she wished to be heard, and an array of microphones detected this change of attention.

Schmandt at 165. And as Rudnick explains:

5 Interface Design

The proper design of a speech interface requires attention to a number of factors, including attention management, error correction, and interaction control.

Spoken communication does not have a direct physical focus in the way that keyboard entry does. In keyboard interaction, there is little ambiguity about the user's intent to communicate—the keyboard is mechanically activated by direct contact. There is no such explicit link in the case of speech. When a user speaks, the utterance may indeed be directed at the currently active application, but it could just as easily have been directed to another human in the environment or have been the result of thinking aloud. The problem is that the system does not have the concept of *attention*; it does not know when it should be paying attention to a potential input. There are a number of solutions to this problem.

The simplest is a “push to talk” system, where the user needs to perform some explicit mechanical act, such as depressing a button during the duration of the utterance or depressing a key just prior to the utterance. These solutions, while

dealing the problem at hand, are not satisfying, since they require the user to perform an incompatible or even disruptive act in order to signal an input to the system. We have experimented with the use of “attention” words that allow the user to activate the recognizer via a single keyword, such as *LISTEN*. If the recognizer is in the standby mode, it will not actively process speech until it encounters the attention word, subsequent to which it will transmit recognitions to the application. If we also provide a keyword for switching off attention, e.g. *STANDBY*, then the user has been given explicit control over the system's “attention” in a manner that is consonant with the modality. (We also have found it useful to have the system enter the standby mode if the quality of recognition falls below some level, on the assumption that the user has redirected his or her attention to another listener in the environment).

Rudnicky at 14-15. Additionally, Manos said the following in 1996:

Finally, word spotting can be used to initiate human-machine interaction. The user can turn on his computer and his large vocabulary continuous-speech recognition system by saying a particular word. Furthermore, people with handicaps will be able to control the opening of doors, switches, television sets, and many other household appliances by voice, using a word spotting system that only listens for specific commands and disregards all other acoustic input.

Manos at 11. Finally, as Ramachandran explained:

■ *Keyword and Phrase Detection:*

Detect occurrences of a particular keyword or phrase in audio data possibly containing the keyword or phrase. If possible, provide a score that characterizes the certainty that the keyword actually occurred. An application of this is in automatic sorting of audio data by its topic based on the keyword content [31, 39, 22], for example, selecting radio news stories that contain the words “health care” in them. Another application is in *open microphone* systems where the recognizer is always listening, waiting for the user to say a known utterance or a specific “trigger word” such as “Computer, turn on the lights.” In such an open-microphone system, extraneous speech not preceded by the word “Computer” must be ignored.

Ramachandran at 124.

Similarly, it would have been obvious to substitute a mechanical switch or push to talk feature with a wake word to avoid disruption and/or the use of one's hands when otherwise occupied (as explained by, for example, by the Schmandt, Rudnick, and Ramachandran passages above, and as also discussed in the CMU Sphinx chart).

For any confidence or likelihood values, it would have been obvious to keep those scores as logarithmic values (e.g., as taught by Konopka '188, Lee I, Lee II, Jouvett, Lam '664, and Tran '140, Jeon '261, Kim '837, Kim '207, Sang-Hoon '493). A POSITA would have been motivated to do so to save processing resources, as was well known.

It would have been obvious to use a confidence rate (e.g., as taught by Bellegarda '809, Braho '095, Braho '101, Catchpole '214, De Souza '667, Hee-seok '469, Hee-souk '242, Hee-souk '951, Hogenhout '750, Jablakov '577, Kepuska '789, Kepuska '717, Konopka '188, Konopka '635, Lahti '711, Lam '664, Millman '413, Paden '454, Sherwood '768, Sherwood '498, Terrell '109, Terrell '261, Tran '140, White '264, Bou-Ghazale, Gales, Jouvett, Lee I, Lee II, Rabiner II, Young II, the HTK Toolkit system, Jeon '261, Kim '837, Kim '207, Nadas '488, Sang-Hoon '493, Stubbley '045, Toyoda '348, and Abdulla) in any speech recognition system. A POSITA would have been motivated to do so to determine whether the speech recognition system actually

identified the speech, and to understand how confident the speech recognition system is in a recognition.

It would have been further obvious to have used a threshold confidence rate (e.g., as taught by Bellegarda '809, Braho '095, Braho '101, De Souza '667, Hee-souk '242, Hee-souk '951, Hogenhout '750, Jablov '577, Kepuska '789, Kepuska '717, Konopka '188, Konopka '635, Lahti '711, Lam '664, Millman '413, Paden '454, Sherwood '768, Sherwood '498, Terrell '109, Terrell '261, White '264, Bou-Ghazale, Gales, Juvet, Lee I, the HTK Toolkit system, Aoyama '796, Jeon '261, Kim '207, Toyoda '348) to determine when speech is recognized in any speech recognition system. A POSITA would have been motivated to do so to because there must be a point at which the speech recognition system is confident enough that it has recognized a word that it does in fact then recognize the word.

It would have been obvious to use a confidence rate (e.g., as taught by Bellegarda '809, Braho '095, Braho '101, Catchpole '214, De Souza '667, Hee-seok '469, Hee-souk '242, Hee-souk '951, Hogenhout '750, Jablov '577, Kepuska '789, Kepuska '717, Konopka '188, Konopka '635, Lahti '711, Lam '664, Millman '413, Paden '454, Sherwood '768, Sherwood '498, Terrell '109, Terrell '261, Tran '140, White '264, Bou-Ghazale, Gales, Juvet, Lee I, Lee II, Rabiner II, Young II, the HTK Toolkit system, Jeon '261, Kim '837, Kim '207, Nadas '488, Sang-Hoon '493, Stubbley '045, Toyoda '348, and Abdulla) to determine authenticity in any speech recognition system. A POSITA would have been motivated to do so to determine whether the speech recognition system actually identifies the speech. Speech recognition systems are looking to identify authentic speech instead of other sounds like noise.

It would have been obvious to have incorporated a rule check unit (e.g., as taught by Hee-seok '469, Hee-seok '242, Huyhn '801, Huhyn '349, Jesensky '700, Jesensky '644, Konopka '188,

Konopka '635, Sherwood '768, Sherwood '498, Tran '140, White '653, Zlokarnik '630, Gales, Kepuska I, Lee I, Lee II, Aoyama '796, Jeon '261, Kim '837, Kim '207, Lee '574, Sang-Hoon '493) in any speech recognition system. A POSITA would have been motivated to do so that the speech recognition system had a set of rules to follow for recognized speech and, thus, could act upon recognized speech successfully, thereby pleasing the user.

It would have been obvious to have incorporated a keyword spotting unit (e.g., as taught by Hee-seok '951, Huyhn '801, Huhyn '349, Jablov '577, Jesensky '700, Jesensky '644, Konopka '188, Konopka '635, Lam '664, Paden '454, Terrell '109, Terrell '261, Tran '140, White '264, Zlokarnik '630, Bou-Ghazale, Kepuska I, Kepuska III, Kim '837, Sang-Hoon '493) in any speech recognition system. A POSITA would have been motivated to do so that the speech recognition system could recognize and then implement a user's commands.

It would have been obvious to a POSITA to take any of the references that conduct speech recognition on a local device and move those processes in whole or in part to a remote location such as a cloud server depending on the system constraints, available components, privacy and other technological considerations. It likewise would have been obvious to take any of the references that conduct speech recognition in part or in whole at a remote location (remote of a device) and move those processes on the device, again depending on the system constraints, available components, privacy and other technological considerations. Such arrangements were well known in the art and well within the skill of a POSITA.

It would have been obvious to combine references with their corresponding implementations or source code (which Defendants expect to further develop during discovery).

For any references that implement the preset call command recognition unit and continuous speech recognition unit on the same structure and to the extent the claims require the preset call

command recognition unit and the continuous speech recognition unit to be separate structures, such references can be easily configured to have separate units or structures performing recognition of the attention word and recognition of user requests, particularly because a speech recognition system will need to be configured differently for different tasks. Depending on the requirements of the specific system, it would have been obvious to a POSITA to have separated the preset call command recognition unit and the continuous speech recognition unit, for example, to conduct continuous speech recognition on a remote server allowing for more processing power and to accommodate more complex requests. As another example, it would have been obvious to access to do so to access a different, perhaps larger and more easily updated, data set. Doing so would simply be combining prior art elements according to known methods to achieve predictable results, depending on the characteristics and goals of the deployed system.

For any references that teach a component other than the preset call command recognition unit (which recognizes input speech, such as a wake word) recognizing speech subsequent to the input speech, modifying such system to have the preset call command recognition unit recognize subsequent speech as claimed would be a simple design choice and/or obvious to try depending on the specific needs and constraints of the particular implementation of the system. Doing so would simplify any speech recognition system by having one component receive speech instead of two components that receive speech.

Additionally, to the extent the preambles are limiting, for any of the references that do not disclose *a method for recognizing a speech of a speaker by a device, an apparatus for recognizing a speech of a speaker, and/or a method for recognizing a speech of a speaker by a single device* (“preamble functionality”), it would have been obvious to use or combine them for that purpose. For example, to the extent any of Bou-Ghazale, Kepuska I, Kepuska II, Kepuska III, Stiles,

Microsoft Speech Recognition, IBM ViaVoice, Dragon Naturallyspeaking, the HTK Toolkit system, Cohen '397, Cohen '776, Gould '236, Hee-souk '469, Hee-Souk '562, Kepuska '789, Kepuska '717, Konopka '188, Konopka '635, Millman '413, Mozer '392, Sherwood '768, Sherwood '498, Tran '140, Kim '837, Nomadic Ratio, Put That There, Clarion AutoPC, CMU Sphinx, ETRI Voice Navigation, and Ford Sync (collectively, "List A") disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Bellegarda '809, Braho '095, Braho '101, Catchpole '214, Cohen '397, Cohen '776, Cohen '576, De Souza '667, Gould '236, Hee-seok '562, Hee-seok '469, Hee-seok '242, Hee-seok '951, Hogenhout '750, Jablov '577, Kepuska '789, Kepuska '717, Konopka '188, Konopka '635, Lahti '711, Lam '664, Millman '413, Mozer '392, Paden '454, Sherwood '768, Sherwood '498, Smyth '453, Squires '043, Terrell '109, Terrell '261, Tran '140, White '264, White '653, Zlokarnik, Yap system, True Knowledge System, HTK Toolkit system, Microsoft Speech Recognition, IBM ViaVoice, Dragon Naturallyspeaking, Bou-Ghazale, Gales, Jouvett, Aoyama '796, Jeon '261, Kim '837, Kim '207, Lee '574, Nadas '488, Sang-Hoon '493, Stubbley '045, Toyoda '348, Abdulla, Bou-Ghazale II, De Fouw, Fludd, Manos, Ramachandran, Rudnick, Schmandt, Clarion Auto PC, CMU Sphinx, ETRI Voice Navigation, Ford Sync, Nomadic Radio, Put That There (collectively, and for purposes of this paragraph only, "List B"). Such a combination would be nothing more than combining prior art elements (i.e., the claimed preamble functionality from any of the references in List B) according to known methods (i.e., adding preamble functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed preamble functionality). Such a combination would be mere use of a known technique (using the claimed preamble functionality from List B) to improve devices from List A. In addition, it would

have been obvious to do a simple substitution of any of the preamble functionality elements from any of the references in List B with any of the other preamble functionality elements in List B.

For any of the references that do not disclose *receiving, by a preset call command recognition unit, an input speech spoken by the speaker, wherein the preset call command recognition unit is configured to receive an input speech spoken by the speaker, receiving, by a preset call command recognition unit of the single device, an input speech spoken by the speaker, and/or wherein the preset call command recognition unit is configured to receive an input speech spoken by the speaker* (“receiving input speech functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Cohen ‘397, Cohen ‘776, Gould ‘236, Hee-souk ‘562, Hee-souk ‘469, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Millman ‘413, Mozer ‘392, Qu ‘786, Sherwood ‘768, Sherwood ‘498, Squires ‘043, Tran ‘140, Bou-Ghazale, Kepuska I, Kepuska II, Kepuska III, Stiles, Microsoft Speech Recognition, IBM ViaVoice, Dragon Naturallyspeaking, Kim ‘837, Sang-Hoon ‘493, De Fouw, Fludd, Manos, Ramachandran, Schmandt, Clarion AutoPC, CMU Sphinx, Nomadic Radio, and Put That There (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed receiving input speech functionality from any of the references in List B) according to known methods (i.e., adding receiving input speech functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed receiving input speech functionality). Such a combination would be mere use of a known technique (using the claimed receiving input speech functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the receiving input speech functionality elements

from any of the references in List B with any of the other receiving input speech functionality elements in List B.

For any of the references that do not disclose *receiving, by the preset call command recognition unit, a subsequent input speech spoken, by the speaker, subsequently to the input speech, receive a subsequent input speech spoken, by the speaker, subsequently to the input speech, receiving, by the preset call command recognition unit, a subsequent input speech spoken, by the speaker, subsequently to the input speech, and/or receive a subsequent input speech spoken, by the speaker, subsequently to the input speech* (“receiving subsequent input speech functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Cohen ‘397, Cohen ‘776, Gould ‘236, Hee-souk ‘562, Hee-souk ‘469, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Millman ‘413, Mozer ‘392, Qu ‘786, Sherwood ‘768, Sherwood ‘498, Squires ‘043, Tran ‘140, Bou-Ghazale, Kepuska I, Kepuska II, Kepuska III, Stiles, Microsoft Speech Recognition, IBM ViaVoice, and/or Dragon Naturallyspeaking, Kim ‘837, Sang-Hoon ‘493, De Fouw, Fludd, Manos, Ramachandran, Schmandt, Clarion AutoPC, CMU Sphinx, Nomadic Radio, and/or Put That There (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed receiving subsequent input speech functionality from any of the references in List B) according to known methods (i.e., adding receiving subsequent input speech functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed receiving subsequent input speech functionality). Such a combination would be mere use of a known technique (using the claimed receiving subsequent input speech functionality from List B) to improve devices from List A. In addition, it would have

been obvious to do a simple substitution of any of the receiving subsequent input speech functionality elements from any of the references in List B with any of the other receiving subsequent input speech functionality elements in List B.

For any of the references that do not disclose *comparing, by the preset call command recognition unit, the input speech with a preset call command, compare the input speech with a preset call command, and/or compare the input speech with a preset call command* (“comparing functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Cohen ‘397, Cohen ‘776, Gould ‘236, Hee-souk ‘562, Hee-souk ‘469, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Millman ‘413, Mozer ‘392, Sherwood ‘768, Sherwood ‘498, Squires ‘043, Tran ‘140, Bou-Ghazale, Kepuska I, Kepuska II, Kepuska III, Stiles, Microsoft Speech Recognition, IBM ViaVoice, Dragon Naturallyspeaking, Kim ‘837, Sang-Hoon ‘493, De Fouw, Fludd, Manos, Ramachandran, Schmandt, Clarion AutoPC, CMU Sphinx, Nomadic Radio, and/or Put That There (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed comparing functionality from any of the references in List B) according to known methods (i.e., adding comparing functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed comparing functionality). Such a combination would be mere use of a known technique (using the claimed comparing functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the comparing functionality elements from any of the references in List B with any of the other comparing functionality elements in List B.

For any of the references that do not disclose *determining, by the preset call command recognition unit, whether the input speech is recognized as the preset call command according to a result of the comparison, recognize the input speech as the preset call command according to a result of the comparison, and/or determining, by the preset call command recognition unit of the single device, whether the input speech is recognized as the preset call command according to a result of the comparison* (“result of the comparison functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Cohen ‘397, Cohen ‘776, Gould ‘236, Hee-souk ‘562, Hee-souk ‘469, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Millman ‘413, Mozer ‘392, Sherwood ‘768, Sherwood ‘498, Squires ‘043, Tran ‘140, Bou-Ghazale, Kepuska I, Kepuska II, Kepuska III, Stiles, Microsoft Speech Recognition, IBM ViaVoice, Dragon Naturallyspeaking, Kim ‘837, Sang-Hoon ‘493, De Fouw, Fludd, Manos, Ramachandran, Schmandt, Clarion AutoPC, CMU Sphinx, Nomadic Radio, and/or Put That There (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed result of the comparison functionality from any of the references in List B) according to known methods (i.e., adding result of the comparison functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed result of the comparison functionality). Such a combination would be mere use of a known technique (using the claimed result of the comparison functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the result of the comparison functionality elements from any of the references in List B with any of the other result of the comparison functionality elements in List B.

For any of the references that do not disclose *based on determining that the input speech is recognized as the preset call command, activating a continuous speech recognition unit configured to determine whether the subsequent input speech is recognized as an actual command, in order to reduce computational load on the device, only if the input speech is recognized as the preset call command, activating a continuous speech recognition unit of the single device that is configured to perform continuous speech recognition to determine whether the subsequent input speech is recognized as an actual command, in order to reduce overall computational load on the single device, and/or wherein the continuous speech recognition unit is configured to recognize the subsequent input speech as an actual command if the input speech is recognized as the preset call command by the preset call command recognition unit* (“activating functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Cohen ‘397, Cohen ‘776, Gould ‘236, Hee-souk ‘562, Hee-souk ‘469, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Millman ‘413, Mozer ‘392, Sherwood ‘768, Sherwood ‘498, Squires ‘043, Tran ‘140, Bou-Ghazale, Kepuska I, Kepuska II, Kepuska III, Stiles, Microsoft Speech Recognition, IBM ViaVoice, Dragon Naturallyspeaking, Kim ‘837, Sang-Hoon ‘493, De Fouw, Fludd, Manos, Ramachandran, Schmandt, Clarion AutoPC, CMU Sphinx, Nomadic Radio, and/or Put That There (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed activating functionality from any of the references in List B) according to known methods (i.e., adding activating functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed activating functionality). Such a combination would be mere use of a known technique (using the claimed activating functionality from List B) to improve

devices from List A. In addition, it would have been obvious to do a simple substitution of any of the activating functionality elements from any of the references in List B with any of the other activating functionality elements in List B.

For any of the references that do not disclose *wherein the preset call command recognition unit comprises a token passing unit configured to assume the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command* (“token passing functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Gould ‘236, Konopka ‘188, Konopka ‘635, Sherwood ‘768, Sherwood ‘498, and/or Tran ‘140 (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed token passing functionality from any of the references in List B) according to known methods (i.e., adding token passing functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed token passing functionality). Such a combination would be mere use of a known technique (using the claimed token passing functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the token passing functionality elements from any of the references in List B with any of the other token passing functionality elements in List B.

For any of the references that do not disclose *wherein determining whether the input speech is recognized as the preset call command further comprises computing, by a confidence detection unit, a confidence rate of the input speech, wherein the preset call command recognition unit further comprises ... a confidence detection unit configured to compute a confidence rate of the*

assumed input speech, and/or wherein determining whether the input speech is recognized as the preset call command further comprises computing, by a confidence detection unit, a confidence rate of the input speech (“confidence detection unit functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Bellegarda ‘809, Braho ‘095, Braho ‘101, Catchpole ‘214, De Souza ‘667, Hee-seok ‘469, Hee-souk ‘242, Hee-souk ‘951, Hogenhout ‘750, Jablovskiy ‘577, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Lahti ‘711, Lam ‘664, Millman ‘413, Paden ‘454, Sherwood ‘768, Sherwood ‘498, Terrell ‘109, Terrell ‘261, Tran ‘140, White ‘264, Bou-Ghazale, Gales, Jouvet, Lee I, Lee II, Rabiner II, Young II, the HTK Toolkit system, Aoyama ‘796, Jeon ‘261, Kim ‘837, Kim ‘207, Nadas ‘488, Sang-Hoon ‘493, Stubbley ‘045, Toyoda ‘348, and/or Abdulla (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed confidence detection unit functionality from any of the references in List B) according to known methods (i.e., adding confidence detection unit functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed confidence detection unit functionality). Such a combination would be mere use of a known technique (using the claimed confidence detection unit functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the confidence detection unit functionality elements from any of the references in List B with any of the other confidence detection unit functionality elements in List B.

For any of the references that do not disclose *wherein determining whether the input speech is recognized as the preset call command further comprises ... recognizing, by the confidence detection unit, the input speech as the preset call command based on the computed confidence*

rate, and/or wherein the preset call command recognition unit further comprises ... a confidence detection unit configured to ... recognize the assumed input speech as the preset call command based on the computed confidence rate (“recognizing based on confidence functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Bellegarda ‘809, Braho ‘095, Braho ‘101, Catchpole ‘214, De Souza ‘667, Hee-seok ‘469, Hee-souk ‘242, Hee-souk ‘951, Hogenhout ‘750, Jablov ‘577, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Lahti ‘711, Lam ‘664, Millman ‘413, Paden ‘454, Sherwood ‘768, Sherwood ‘498, Terrell ‘109, Terrell ‘261, Tran ‘140, White ‘264, Bou-Ghazale, Gales, Jouvett, Lee I, Lee II, Rabiner II, Young II, the HTK Toolkit system, Aoyama ‘796, Jeon ‘261, Kim ‘837, Kim ‘207, Nadas ‘488, Sang-Hoon ‘493, Stubbley ‘045, Toyoda ‘348, and/or Abdulla (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed recognizing based on confidence functionality from any of the references in List B) according to known methods (i.e., adding recognizing based on confidence functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed recognizing based on confidence functionality). Such a combination would be mere use of a known technique (using the claimed recognizing based on confidence functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the recognizing based on confidence functionality elements from any of the references in List B with any of the other recognizing based on confidence functionality elements in List B.

For any of the references that do not disclose *wherein recognizing the input speech comprises recognizing the input speech as the preset call command if the computed confidence*

rate is higher than a threshold value, and/or wherein the confidence detection unit is further configured to recognize the assumed input speech as the preset call command if the computed confidence rate is higher than a threshold value (“threshold functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Bellegarda ‘809, Braho ‘095, Braho ‘101, De Souza ‘667, Hee-souk ‘242, Hee-souk ‘951, Hogenhout ‘750, Jablov ‘577, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Lahti ‘711, Lam ‘664, Millman ‘413, Paden ‘454, Sherwood ‘768, Sherwood ‘498, Terrell ‘109, Terrell ‘261, White ‘264, Bou-Ghazale, Gales, Jouvett, Lee I, the HTK Toolkit system, Aoyama ‘796, Jeon ‘261, Kim ‘207, Toyoda ‘348, Abdulla (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed threshold functionality from any of the references in List B) according to known methods (i.e., adding threshold functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed threshold functionality). Such a combination would be mere use of a known technique (using the claimed threshold functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the threshold functionality elements from any of the references in List B with any of the other threshold functionality elements in List B.

For any of the references that do not disclose *wherein recognizing the input speech comprises determining whether or not the input speech is an authentic call command based on the computed confidence rate, and/or wherein the confidence detection unit is further configured to determine whether or not the assumed input speech is an authentic call command based on the computed confidence rate* (“determine authenticity functionality”) it would have been obvious to

use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Bellegarda '809, Braho '095, Braho '101, Catchpole '214, De Souza '667, Hee-seok '469, Hee-souk '242, Hee-souk '951, Hogenhout '750, Jablovsk '577, Kepuska '789, Kepuska '717, Konopka '188, Konopka '635, Lahti '711, Lam '664, Millman '413, Paden '454, Sherwood '768, Sherwood '498, Terrell '109, Terrell '261, Tran '140, White '264, Bou-Ghazale, Gales, Jouvett, Lee I, Lee II, Rabiner II, Young II, the HTK Toolkit system, Aoyama '796, Jeon '261, Kim '837, Kim '207, Nadas '488, Sang-Hoon '493, Stubbley '045, Toyoda '348, and/or Abdulla (collectively, and for purposes of this paragraph only, "List B"). Such a combination would be nothing more than combining prior art elements (i.e., the claimed determine authenticity functionality from any of the references in List B) according to known methods (i.e., adding determine authenticity functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed determine authenticity functionality). Such a combination would be mere use of a known technique (using the claimed determine authenticity functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the determine authenticity functionality elements from any of the references in List B with any of the other determine authenticity functionality elements in List B.

For any of the references that do not disclose *wherein recognizing the input speech comprises ... recognizing the input speech as the preset call command if it is determined that the input speech is the authentic call command*, and/or *recognize the assumed input speech as the preset call command if it is determined that the assumed input speech is the authentic call command* ("recognizing based on authenticity functionality") it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or

method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Bellegarda '809, Braho '095, Braho '101, Catchpole '214, De Souza '667, Hee-seok '469, Hee-souk '242, Hee-souk '951, Hogenhout '750, Jablov '577, Kepuska '789, Kepuska '717, Konopka '188, Konopka '635, Lahti '711, Lam '664, Millman '413, Paden '454, Sherwood '768, Sherwood '498, Terrell '109, Terrell '261, Tran '140, White '264, Bou-Ghazale, Gales, Jouvett, Lee I, Lee II, Rabiner II, Young II, the HTK Toolkit system, Aoyama '796, Jeon '261, Kim '837, Kim '207, Nadas '488, Sang-Hoon '493, Stubbley '045, Toyoda '348, and/or Abdulla (collectively, and for purposes of this paragraph only, "List B"). Such a combination would be nothing more than combining prior art elements (i.e., the claimed recognizing based on authenticity functionality from any of the references in List B) according to known methods (i.e., adding recognizing based on authenticity functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed recognizing based on authenticity functionality). Such a combination would be mere use of a known technique (using the claimed recognizing based on authenticity functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the recognizing based on authenticity functionality elements from any of the references in List B with any of the other recognizing based on authenticity functionality elements in List B.

For any of the references that do not disclose *wherein determining the input speech is recognized as the preset call command further comprises computing a log likelihood ratio of the input speech, wherein the token passing unit is further configured to compute a log likelihood ratio of the input speech and assume the input speech to be the preset call command based on the computed log likelihood ratio, and/or wherein determining whether the input speech is recognized as the preset call command further comprises computing a log likelihood ratio of the input speech*

(“computing log likelihood ratio functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Hee-Souk ‘469, Hee-Souk ‘242, Hee-Souk ‘951, Hogenhout ‘750, Jiang ‘353, Junqua ‘990, Konopka ‘188, Konopka ‘635, Lam ‘664, Sherwood ‘768, Sherwood ‘498, Tran ‘140, Bou-Ghazale, Jouvett, Lee I, Lee II, Young II, HTK Toolkit system, Jeon ‘261, Kim ‘837, Kim ‘207, Sang-Hoon ‘493 (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed computing log likelihood ratio functionality from any of the references in List B) according to known methods (i.e., adding computing log likelihood ratio functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed computing log likelihood ratio functionality). Such a combination would be mere use of a known technique (using the claimed computing log likelihood ratio functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the computing log likelihood ratio functionality elements from any of the references in List B with any of the other computing log likelihood ratio functionality elements in List B.

For any of the references that do not disclose *wherein determining the input speech is recognized as the preset call command further comprises ... assuming the input speech to be the preset call command based on the computed log likelihood ratio* (“assuming based on computed log likelihood ratio functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Hee-Souk ‘469, Hee-Souk ‘242, Hee-Souk ‘951, Hogenhout ‘750, Jiang ‘353, Konopka ‘188, Konopka ‘635,

Lam ‘664, Sherwood ‘768, Sherwood ‘498, Tran ‘140, Bou-Ghazale, Jouvét, Lee I, Lee II, Young II, Jeon ’261, and/or HTK Toolkit system (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed assuming based on computed log likelihood ratio functionality from any of the references in List B) according to known methods (i.e., adding assuming based on computed log likelihood ratio functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed assuming based on computed log likelihood ratio functionality). Such a combination would be mere use of a known technique (using the claimed assuming based on computed log likelihood ratio functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the assuming based on computed log likelihood ratio functionality elements from any of the references in List B with any of the other assuming based on computed log likelihood ratio functionality elements in List B.

For any of the references that do not disclose *wherein the computed log likelihood ratio is a natural logarithm of a probability on the input speech for a phoneme model of the preset call command, and the computed confidence rate is another log likelihood ratio* (“phoneme model functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Hee-souk ‘951, Konopka ‘188, Konopka ‘635, Lam ‘664, Tran ‘140, Jouvét, Lee I, Lee II, Jeon ’261, Kim ’837, Kim ’207, Sang-Hoon ‘493 (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed phoneme model functionality from any of the references in List B) according to known methods (i.e., adding phoneme model functionality as disclosed in the prior art) to achieve a predictable result (i.e., the

claimed phoneme model functionality). Such a combination would be mere use of a known technique (using the claimed phoneme model functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the phoneme model functionality elements from any of the references in List B with any of the other phoneme model functionality elements in List B.

For any of the references that do not disclose *wherein said another log likelihood ratio is computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command* (“subtracting functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Konopka ‘188, Konopka ‘635, Lam ‘664, Tran ‘140, Jouviet, Lee I, Jeon ‘261 and/or Lee II (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed subtracting functionality from any of the references in List B) according to known methods (i.e., adding subtracting functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed subtracting functionality). Such a combination would be mere use of a known technique (using the claimed subtracting functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the subtracting functionality elements from any of the references in List B with any of the other subtracting functionality elements in List B.

For any of the references that do not disclose *determining, using a rule check unit, whether or not the recognized subsequent input speech corresponds to a target sentence, and/or wherein the continuous speech recognition unit comprises ... a rule check unit configured to determine*

whether the recognized subsequent input speech corresponds to a target sentence (“rule check functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Hee-seok ‘469, Hee-seok ‘242, Huyhn ‘801, Huhyn ‘349, Jesensky ‘700, Jesensky ‘644, Konopka ‘188, Konopka ‘635, Sherwood ‘768, Sherwood ‘498, Tran ‘140, White ‘653, Zlokarnik ‘630, Gales, Kepuska I, Lee I, Lee II, Aoyama ‘796, Jeon ‘261, Kim ‘837, Kim ‘207, Lee ‘574, Sang-Hoon ‘493, (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed rule check functionality from any of the references in List B) according to known methods (i.e., adding rule check functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed rule check functionality). Such a combination would be mere use of a known technique (using the claimed rule check functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the rule check functionality elements from any of the references in List B with any of the other rule check functionality elements in List B.

For any of the references that do not disclose *a processor* (“processor functionality”) it would have been obvious to use or combine them for that purpose. any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Bellegarda ‘809, Braho ‘095, Braho ‘101, Catchpole ‘214, Cohen ‘397, Cohen ‘776, Cohen ‘576, De Souza ‘667, Gould ‘236, Hee-seok ‘469, Hee-seok ‘242, Hogenhout ‘750, Huhyn ‘801, Huhyn ‘349, Jesensky ‘700, Jiang ‘353, Jesensky ‘644, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Lahti ‘711, Lam ‘664, Millman ‘413, Mozer ‘392, Qu ‘786, Sherwood ‘768, Sherwood ‘498, Smyth ‘453, Squires ‘043, Terrell ‘109, Terrell ‘261, Tran ‘140,

White ‘264, Bou-Ghazale, Jeon ‘261, and/or Kepuska I (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed processor functionality from any of the references in List B) according to known methods (i.e., adding processor functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed processor functionality). Such a combination would be mere use of a known technique (using the claimed processor functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the processor functionality elements from any of the references in List B with any of the other processor functionality elements in List B.

For any of the references that do not disclose *a preset call command recognition unit using the processor* (“PCCRU using processor functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Cohen ‘397, Cohen ‘776, Gould ‘236, Hee-souk ‘562, Hee-souk ‘469, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Millman ‘413, Mozer ‘392, Qu ‘786, Sherwood ‘768, Sherwood ‘498, Squires ‘043, Tran ‘140, Bou-Ghazale, Kepuska I, Kepuska II, Kepuska III, Stiles, Microsoft Speech Recognition, IBM ViaVoice, and/or Dragon Naturallyspeaking (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the PCCRU using processor functionality from any of the references in List B) according to known methods (i.e., adding PCCRU using processor functionality as disclosed in the prior art) to achieve a predictable result (i.e., the PCCRU using processor functionality). Such a combination would be mere use of a known technique (using the claimed PCCRU using processor functionality from List B) to

improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the PCCRU using processor functionality elements from any of the references in List B with any of the other PCCRU using processor functionality elements in List B.

For any of the references that do not disclose *a continuous speech recognition unit using the processor* (“CSRU using processor functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Cohen ‘397, Cohen ‘776, Gould ‘236, Hee-souk ‘562, Hee-souk ‘469, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Millman ‘413, Mozer ‘392, Sherwood ‘768, Sherwood ‘498, Squires ‘043, Tran ‘140, Bou-Ghazale, Kepuska I, Kepuska II, Kepuska III, Stiles, Microsoft Speech Recognition, IBM ViaVoice, and/or Dragon Naturallyspeaking (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the CSRU using processor functionality from any of the references in List B) according to known methods (i.e., adding CSRU using processor functionality as disclosed in the prior art) to achieve a predictable result (i.e., the CSRU using processor functionality). Such a combination would be mere use of a known technique (using the claimed CSRU using processor functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the CSRU using processor functionality elements from any of the references in List B with any of the other CSRU using processor functionality elements in List B.

For any of the references that do not disclose *wherein the preset call command recognition unit further comprises an adaptive filter configured to receive the input speech and the subsequent input speech* (“filter functionality”) it would have been obvious to use or combine them for that

purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Rabiner III, Gould '236, Jablovkov '577, Nadas '488, and/or Takiguchi '832 (collectively, and for purposes of this paragraph only, "List B"). Such a combination would be nothing more than combining prior art elements (i.e., the filter functionality from any of the references in List B) according to known methods (i.e., adding filter functionality as disclosed in the prior art) to achieve a predictable result (i.e., the filter functionality). Such a combination would be mere use of a known technique (using the claimed filter functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the filter functionality elements from any of the references in List B with any of the other filter functionality elements in List B. Use of adaptive filters were well known in the art to handle adverse noise conditions. *See, e.g., Rabiner III at 310.*

For any of the references that do not disclose *wherein the continuous speech recognition unit comprises a keyword spotting unit configured to recognize the subsequent input speech as the actual command if the input speech is recognized as the preset call command* ("keyword spotting functionality") it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Hee-seok '951, Huyhn '801, Huhyn '349, Jablovkov '577, Jesensky '700, Jesensky '644, Konopka '188, Konopka '635, Lam '664, Paden '454, Terrell '109, Terrell '261, Tran '140, White '264, Zlokarnik '630, Bou-Ghazale, Kepuska I, Kim '837, and/or Kepuska III (collectively, and for purposes of this paragraph only, "List B"). Such a combination would be nothing more than combining prior art elements (i.e., the claimed keyword spotting functionality from any of the references in List B) according to known

methods (i.e., adding keyword spotting functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed keyword spotting functionality). Such a combination would be mere use of a known technique (using the claimed keyword spotting functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the keyword spotting functionality elements from any of the references in List B with any of the other keyword spotting functionality elements in List B.

For any of the references that do not disclose *wherein the single device is a portable device or an embedded system* (“portable or embedded functionality”) it would have been obvious to use or combine them for that purpose. For example, to the extent any of List A disclose a system or method that does not satisfy this claim limitation, then it would have been obvious to combine these references with Braho ‘095, Braho ‘101, Catchpole ‘214, Cohen ‘397, Cohen ‘776, De Souza ‘667, Gould ‘236, Hee-souk ‘242, Hee-souk ‘469, Hogenhout ‘750, Huyhn ‘801, Huyhn ‘359, Jablovkov ‘577, Jesensky ‘700, Jiang ‘353, Jesensky ‘644, Kepuska ‘789, Kepuska ‘717, Konopka ‘188, Konopka ‘635, Millman ‘413, Lahti ‘711, Lam ‘664, Millman ‘413, Mozer ‘392, Paden ‘454, Qu ‘786, Sherwood ‘768, Sherwood ‘498, Smyth ‘453, Squires ‘043, Terrell ‘109, Terrell ‘261, Tran ‘140, White ‘264, Zlokarnik ‘630, Bou-Ghazale, Kepuska I, HTK Toolkit system, Microsoft Speech Recognition, IBM ViaVoice, Kim ‘837, Kim ‘207, Sang-Hoon ‘493, Stubley ‘045, Toyoda ‘348, De Fouw, Fludd, and/or Dragon Naturallyspeaking (collectively, and for purposes of this paragraph only, “List B”). Such a combination would be nothing more than combining prior art elements (i.e., the claimed portable or embedded functionality from any of the references in List B) according to known methods (i.e., adding portable or embedded functionality as disclosed in the prior art) to achieve a predictable result (i.e., the claimed portable or embedded functionality). Such a combination would be mere use of a known technique (using the claimed portable or

embedded functionality from List B) to improve devices from List A. In addition, it would have been obvious to do a simple substitution of any of the portable or embedded functionality elements from any of the references in List B with any of the other portable or embedded functionality elements in List B.

C. Disclosures Pursuant to P.R. 3-3(c)

The attached claim charts are fully incorporated herein and identify where in each of the charted prior art references the limitations of the asserted claims of the '196 Patent are found. As discussed below, the following claim terms are governed by 35 U.S.C. § 112 ¶ 6: preset call command recognition unit; continuous speech recognition unit; token passing unit; confidence detection unit; rule check unit; and keyword spotting unit. The claim charts identify the structure(s), act(s), or material(s) in each item of prior art that performs the claimed function.

Defendants have endeavored to cite to the most relevant portions of the identified prior art. However, other portions of the identified prior art may additionally disclose, either expressly or inherently, one or more elements or limitations of the asserted claims. In an effort to focus the issues, Defendants identify only exemplary portions of the identified prior art. To the extent it is found a charted reference does not expressly disclose certain claim limitations, such limitations are inherent. Defendants reserve the right to rely on uncited portions of the identified prior art to establish the invalidity of the asserted claims. Moreover, Defendants reserve the right to rely on uncited portions of the identified prior art, rely on other references (irrespective of whether such references themselves qualify as prior art) to show the state of the art, and/or to rely on expert or fact testimony to provide context to, or aid in, understanding the cited portions of the identified prior art. Where Defendants cite to a particular drawing or figure in the attached claim charts, the citation encompasses the description of the drawing or figure, as well as any text associated with

the drawing or figure. Similarly, where Defendants cite to particular text concerning a drawing or figure, the citation encompasses that drawing or figure as well.

Relatedly, it is often the case that certain portions of patent disclosures build upon other disclosures – even if they are referred to as a separate or alternative embodiment. Thus, Defendants’ citations to structures or functions incorporate by reference all disclosures to related structures or functions, including any additional detail provided as to the operation or design of those structures or functions. Further, citations in each dependent claim include citations to each claim from which it directly or indirectly depends. And citations as to any recited limitation, step, or component in the claims apply wherever each such limitation, step, or component is repeatedly recited elsewhere in the claim.

D. Disclosures Pursuant to P.R. 3-3(d)

Pursuant to Patent Local Rule 3-3(d), Defendants hereby identify grounds of invalidity based on (1) lack of written description and enablement under 35 U.S.C. § 112, ¶ 1; and (2) indefiniteness under 35 U.S.C. § 112, ¶ 2. These Initial Invalidity Contentions shall not be construed as an admission that any claim construction advanced by Defendants in this case is in any way inconsistent, flawed, or erroneous. Nor should these contentions prevent Defendants from advancing claim construction and/or noninfringement positions in lieu of, or in addition to, invalidity positions. Further, these Initial Invalidity Contentions shall not be construed as an admission of or acquiescence to KAIFI’s purported constructions of the claim language or of other positions advanced by KAIFI during the course of this litigation. Indeed, Defendants’ Invalidity Contentions under 35 U.S.C. § 112 may depend on the Court’s claim construction, as well as KAIFI’s asserted claim scope.

Below, Defendants identify the issues under 35 U.S.C. § 112 of which they are presently aware. Pursuant to the applicable provisions of the Docket Control Order and Local Rules,

Defendants reserve the right to supplement or amend their contentions after the Court’s *Markman* ruling or if KAIFI amends or alters its infringement contentions in any way.

1. Written Description and Enablement

To satisfy the written description requirement of 35 U.S.C. § 112(a) or pre-AIA 35 U.S.C. § 112, ¶ 1, the specification of a patent must “clearly allow persons of ordinary skill in the art to recognize that the inventor invented what is claimed” and “had possession of the claimed subject matter as of the filing date.” *Ariad Pharm., Inc. v. Eli Lilly & Co.*, 598 F.3d 1336, 1351 (Fed. Cir. 2010) (en banc). In particular, an adequate written description must “demonstrate that the patentee possessed the full scope of the invention recited in [the] claim.” *LizardTech, Inc. v. Earth Resource Mapping, Inc.*, 424 F.3d 1336, 1345 (Fed. Cir. 2005). To satisfy the enablement requirement, “the specification of a patent must teach those skilled in the art how to make and use the full scope of the claimed invention without undue experimentation.” *Genentech, Inc. v. Novo Nordisk A/S*, 108 F.3d 1361, 1365 (Fed. Cir. 1997) (quotations omitted); *see also Amgen v. Sanofi*, 143 S.Ct. 1243, 1244 (2023) (“If a patent claims an entire class of processes, machines, manufactures, or compositions of matter, the patent’s specification must enable a person skilled in the art to make and use the entire class. In other words, the specification must enable the full scope of the invention as defined by its claims. The more one claims, the more one must enable.”).

The specification of the ‘196 Patent fails to meet the requirements of 35 U.S.C. § 112 ¶ 1 because it fails to disclose, enable or teach—or fails to disclose, enable or teach the full scope of—at least the following claim phrases:

“preset call command recognition unit” (Claims 1, 9-10, 17).¹ The ‘196 Patent does not provide adequate written description or enable the “preset call command recognition unit” recited in claims 1, 9-10, and 17. As an initial matter, neither the word “preset” nor the phrase “call command recognition unit” appear in the ‘196 Patent outside of the claims. Moreover, contrary to the “preset” language of the claims, the ‘196 Patent only describes a *configurable*, i.e., *non-preset*, call command. As shown in the examples below, the ‘196 Patent consistently describes the first step of its invention as *configuring* a recognition network to determine what call command will be recognized:

- ‘196 Patent at 3:5-20: “According to the construction of the present invention . . . a call command such as ‘Naraeya’ or ‘Nepiya’ is selected.”
- *Id.* at 4:48-61: Describing first “select[ing] a call command such as ‘Neraeya’ or ‘Nepiya’ and configur[ing] a Minimum recognition network” to recognize the selected call command.
- *Id.* at 6:29-35: “The call command/continuous speech recognition system configures a recognition network for call commands (for example, ‘Naraeya’, ‘Nepiya’, etc.).”
- *Id.* at 7:5-8: “First, the call command recognition network is configured (Step 1). The real-time recognition network for a recorded input speech frame is decoded and is then stored in tokens (Step 2).”
- *Id.* at 5:44-48: “The keyword searching unit 423 includes a token passing unit for configuring a speech frame, which includes a silence interval accompanied by noise and call commands from the speaker, in the form of a minimum number of tokens[.]”

At no point does the ‘196 Patent describe any call command or any call command recognition unit that is “preset” (nor could it). Due to the nature of speech recognition, a network to look for one word will inherently be different than the network to look for another word. For example, a

¹ In the headings and conclusions, Defendants identify the claim(s) that contain the particular term(s) being discussed. It should be understood that all claims that depend from the identified invalid claim(s) are also invalid.

network for each word will require different configurations and training. None of this is described in or enabled by the '196 Patent specification, which describes speech recognition more of a goal than something that is explained or enabled.

Where a patent claims one way of doing something and the specification describes another way of doing something, courts regularly find the claim invalid for lack of written description. Additionally, the complete absence of the terms “preset” and “call command recognition unit” in the specification further demonstrates the '196 Patent does not provide adequate written description for the claimed “preset call command recognition unit.”

Moreover, the '196 Patent would not enable one skilled in the art to make or use the claimed “preset call command recognition unit.” Notably, “preset call command recognition unit” is not a term known in the art and, again, the '196 Patent provides no guidance as to how to make or use a preset call command recognition unit.

Additionally, KAIFI's initial infringement contentions as to this claim limitation appear to accuse (1) “a combination of hardware, processing, and software functions both on devices and on the cloud server;” (2) Amazon's wake word engine, and/or (3) the AVS client as being the claimed preset call command recognition unit. Initial Infringement Contentions, Appendix B at 11-23, 32, 39, 120-147, 157, 167, 175, 258-269, 278. The '196 Patent certainly does not adequately describe or enable this broad a scope of the term “preset call command recognition unit.” Instead, and as discussed above, the '196 Patent's description/enablement of this term is (at best) limited and inconsistent with the claim language. The '196 Patent additionally does not adequately describe or enable this broad a scope because it does not disclose or explain any of its elements as being on hardware, software, or a cloud server (none of these are even disclosed in the '196 Patent).

Thus, claims 1, 9-10, and 17 (and all claims that depend therefrom) are invalid for lack of proper written description and enablement.

“preset call command” (Claims 1-7, 9-17, 19-20). As discussed above for “preset call command recognition unit,” the ‘196 Patent only describes a *configurable*, i.e., *non-preset*, call command. Thus, “preset call command” as in claims 1-7, 9-17, 19-20 (and all claims that depend therefrom) also lacks adequate written description and enablement.

“continuous speech recognition unit” (Claims 1, 9, and 17). Like “preset call command recognition unit” discussed above, “continuous speech recognition unit” is not a term of art and the specification of the ‘196 Patent does not disclose or explain what a “continuous speech recognition unit” is. In fact, this term does not appear in the ‘196 Patent outside of the claims. Instead, without explaining what any of these terms mean or providing examples, the ‘196 Patent discloses a “continuous speech recognition system,” a “continuous speech recognition network,” a “continuous speech recognition engine,” and a “continuous speech recognition technique.” The ‘196 Patent does not explain the differences between these four terms, nor does it attempt to explain the difference between the claimed “continuous speech recognition unit” and these four terms. For these reasons, the ‘196 Patent does not show the inventors were in possession of the claimed “continuous speech recognition unit” and does not identify to a person skilled in the art how to make or used the claimed “continuous speech recognition unit.” Thus, claims 1, 9, and 17 (and all claims that depend therefrom) lack proper enablement and written description.

Moreover, the claim language appears directed to any and all continuous speech recognition units, with no limitations or exceptions. Even if the ‘196 Patent enables a certain type of continuous speech recognition unit (which Defendants dispute), it certainly doesn’t enable any and all continuous speech recognition units.

Additionally, KAIFI's Initial Infringement Contentions appear to accuse "a combination of hardware, processing, and software functions both on devices and on the cloud server" as being the "continuous speech recognition unit." Initial Infringement Contentions, Appendix B at 135, 183.² The '196 Patent does not adequately describe or enable this broad a scope at least because it does not disclose any components as being hardware, software, or a cloud server. Claims 1, 9, and 17 (and all claims that depend therefrom) are thus invalid for lack of proper written description and enablement for this reason as well.

"token passing unit" (Claims 1, 9, and 17). Claims 1, 9, and 17 each require a "token passing unit configured to assume the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command." The '196 Patent discloses two specific token passing units: one that is part of the continuous speech recognition network ('196 Patent at 4:35-37; 5:9-10; 5:59-61) and another that is part of the keyword searching unit 423 of the call command recognition network 420 (*id.* at 5:7-9; 5:44-48). The '196 Patent does not disclose a token passing unit that is part of a preset call command recognition unit, as claimed. *See* '196 Patent at cl. 1 ("wherein the preset call command recognition unit comprises a token passing unit"), cl. 9 ("the preset call command recognition unit comprises a token passing unit"), cl. 17 ("wherein the preset call command recognition unit comprises a token passing unit"). The claimed "token passing unit" lacks adequate written description and enablement for this reason alone.

² Although KAIFI does not use the "a combination of hardware, processing, and software functions both on devices and on the cloud server" language in its discussion of claims 1 and 17, KAIFI appears to be applying a similar, broad scope of the term. *See* Initial Infringement Contentions, Appendix B at 47-60, 291-304.

Even if KAIFI argues the claimed “token passing unit” is the keyword searching unit 423’s token passing unit, the ‘196 Patent still does not properly describe or enable the claim language. Specifically, the ‘196 Patent’s disclosure/enablement of keyword searching unit 423’s token passing unit is very specific and limited to the following: “[t]he keyword searching unit 423 includes a token passing unit for configuring a speech frame, which includes a silence interval accompanied by noise and call commands from the speaker, in the form of a minimum number of tokens, and then searching for call commands” ‘196 Patent at 5:44-52. This disclosure has nothing to do with, and therefore does not describe or enable, “configured to assume the input speech to be the preset call command according to [a] result of [a] comparison using a minimum recognition network,” as claimed. Thus, claims 1, 9, and 17 additionally lack written description and enablement because the ‘196 Parent (1) does not show the named inventors were in possession of a token passing unit that is part of the preset call command recognition unit and is “configured to assume the input speech to be the preset call command according to [a] result of [a] comparison using a minimum recognition network,” (2) it does not identify to a POSITA how to make or use the claimed token passing unit of the preset call command recognition unit.

Moreover, KAIFI’s Infringement Contentions appear to accuse one or more of the following as being the claimed token passing unit:

- A user speech stream;
- An indicator notifying that a device has detected the wake word and is now processing a request;
- Detecting only the sound waves of a chosen wake word;
- An ASR model;
- Using ASR;
- ASR that is trained for silence and noise;
- ASR that adds a silence label at start and end of the label sequence and in between words;

- ASR using end pointers;
- A shared memory ring buffer;
- Wake word indexes;
- A voice request;
- Wake word and speech recognition;
- Performing “true accepts” of the wake word;
- An ASR decoder;
- Performance conditions that include noise; and
- Keyword spotting algorithms.

Initial Infringement Contentions, Appendix B at 61-78, 194-212, 232, 305-322. Amazon disagrees that “token passing unit” can be interpreted to cover any one of the extremely broad categories and items identified by KAIFI above. If it can, “token passing unit” is invalid for lack of written description and enablement. As discussed above, the ‘196 Patent discloses nothing about a token passing unit that is part of a preset call command recognition unit, as claimed. To the extent the token passing unit that is part of the ‘196 Patent’s call command recognition network is considered to be the claimed token passing unit, the ‘196 Patent’s disclosure is very specific and limited to the following: “[t]he keyword searching unit 423 includes a token passing unit for configuring a speech frame, which includes a silence interval accompanied by noise and call commands from the speaker, in the form of a minimum number of tokens, and then searching for call commands” ‘196 Patent at 5:44-52. This limited disclosure does not support or describe or enable the incredibly broad scope that KAIFI is now trying to give this limitation. Thus, to the extent that KAIFI’s infringement contentions are credited, claims 1, 9, and 17 (and all claims that depend therefrom) are invalid for written description and enablement for this reason as well.

“minimum recognition network composed of a silence interval accompanied by noise and the preset call command” (Claims 1, 9, and 17). The ‘196 Patent never defines minimum

recognition network or silence interval. Claims 1, 9 and 17 require a “minimum recognition network” that is “composed of” (closed language, so composed of *only*) “a silence interval accompanied by noise and the [singular] preset call command.” However, other parts of the patent describe a minimum recognition network that includes *multiple* call commands. See ‘196 Patent at Abstract; 1:13-16. “Minimum recognition network” is not a term of art, nor is it shown in any of the drawings. The ‘196 Patent does not describe or enable how a minimum recognition network can be made or used including only a silence interval, noise, and a preset call command.

KAIFI’s infringement contentions appear to accuse ASR training for silence and noise (which is, itself, a confusing and unclear position), and ASR adding a silence label at the start and end of the label sequence and in between words as being the claimed “minimum recognition network.” Initial Infringement Contentions, Appendix B at 75, 209, 319. To the extent KAIFI is successful in construing this limitation to mean something other than a silence interval accompanied by noise and a singular preset call command, the ‘196 Patent fails to adequately disclose or enable such functionality. Moreover, KAIFI is judicially estopped from arguing otherwise since during prosecution, to overcome a prior art rejection, the applicant argued that “minimum recognition network composed of a silence interval accompanied by noise and the preset call command” “is a one word sequence of a one word command followed by silence and then a command[.]” Following the addition of this language to all of the independent claims, the ‘196 Patent was allowed.

Moreover, “silence interval” is not defined, explained, or enabled by the specification.

Claims 1, 9, and 17 (and all claims that depend therefrom) are invalid for this reason as well.

“actual command” (Claims 1, 9, 16, 17). Although the goal of determining an “actual command” (whatever that may mean) is disclosed in passing in the ‘196 Patent, how to achieve this goal is not. There is nothing in the ‘196 Patent to show the inventors had possession of this claim limitation or that a person skilled in the art could make or use it. Thus, claims 1, 9, 16, and 17 (and all claims that depend therefrom) are invalid.

“processor” / “preset call command recognition unit using the processor” / “continuous speech recognition unit using the processor” (Claim 9). The ‘196 Patent does not adequately describe or enable the claimed “processor” and/or use of the claimed processor by the preset call command recognition unit and the continuous speech recognition unit, as required by claim 9.

The word “processor” does not appear anywhere in the ‘196 Patent except for in claim 9. *See, generally,* ‘196 Patent. “[P]rocessor” was not included in the originally filed claim 9. Instead, “processor” and the claimed use thereof by the preset call command recognition unit and continuous speech recognition unit was added to claim 9 during prosecution in response to a rejection by the examiner under 35 U.S.C. § 101. *See* ‘196 Patent Prosecution History, November 27, 2013 Reply to Non-Final Office Action. As discussed below, the claimed processor and its use by the preset call command recognition unit and continuous speech recognition unit is new matter that is not adequately supported by the originally filed specification, and claim 9 is invalid for adequate written description.

First, again, there is no processor disclosed in the as-filed ‘196 Patent. *Second*, a processor is not inherent in the as-filed ‘196 Patent. The as-filed ‘196 Patent does not even delineate between hardware and software and other elements.³ *Third*, as discussed above, the ‘196 Patent does not describe or explain what the “preset call command recognition unit” or the “continuous speech

³ Neither of the words “hardware” or “software” appear in the ‘196 Patent.

recognition unit” are, and the drawings provide little to no guidance as these elements are shown as simple black boxes.

Accordingly, the specification does not include any written description support for a processor or how a “call command recognition unit” or “continuous speech recognition unit” would “us[e] the processor.” The addition of a processor to claim 9 changes the scope of the claim beyond what is supported by the as-filed ‘196 Patent and therefore is impermissible new matter. At a minimum, the ‘196 Patent certainly does not adequately support or enable the “preset call command recognition unit” and the “continuous speech recognition unit” using the *same* processor, as claim 9 requires.

As discussed below for the device limitations, KAIFI’s infringement contentions appear to accuse devices of the Defendants, devices of non-Defendants, and cloud-based services *together* as meeting these claim limitations. To the extent KAIFI contends multiple processors (e.g., one in the cloud and one on an Amazon device, or one on a non-Amazon device and one in the cloud) meet the processor limitation of claim 9, the ‘196 Patent does not describe or enable this broad of a scope. As discussed above, the ‘196 Patent does not disclose a processor, let alone multiple processors in different places and devices that can somehow be read to meet this claim limitation.

In sum, claim 9 (and all claims that depend therefrom) is invalid under § 112 ¶ 1.

“comparing” and “determining” steps (Claims 1 and 17); “compare” and “recognize” steps (Claim 9). Claims 1 and 17 of the ‘196 Patent recite comparing and determining steps related to identifying input speech as the call command. Claim 9 requires a preset call command recognition unit configured to perform compare and recognize steps. Specifically, the independent claims recite:

Claim 1: “comparing, by the preset call command recognition unit, the input speech with a preset call command; determining, by the preset call command recognition unit,

whether the input speech is recognized as the preset call command according to a result of the comparison[.]”

Claim 9: “wherein the preset call command recognition unit is configured to ... compare the input speech with a preset call command; and recognize the input speech as the preset call command according to a result of the comparison[.]”

Claim 17: “comparing, by the preset call command recognition unit, the input speech with a preset call command; determining, by the preset call command recognition unit of the single device, whether the input speech is recognized as the preset call command according to a result of the comparison[.]”

As discussed below, the ‘196 Patent does not provide adequate written description support or enablement for these claim limitations.

First of all, the 196 patent does not explain or describe the compare, determine, and recognize steps in the specification, nor does it delineate between the three of them.

The ‘196 Patent suggests that the comparing, determining, and recognizing steps somehow involve a confidence rate and/or confidence detection. *See* ‘196 Patent at 2:43-48; 3:6-20; 3:45-50; 3:54-63; 4:5-11; 4:15-31; 4:48-61; 4:65-5:6; 5:19-22; 5:44-52; 5:65-67; 6:4-9; 6:14-18; 6:24-35; 7:1-3; 7:18-24; 7:32-39; 7:52-8:8; 8:16-41; Fig. 7. The ‘196 Patent teaches “the confidence rate is determined by the following equation: $LLR_k(0, \lambda_k) = \log p(0|\lambda_k) - \log p(0|\lambda_\xi)$ where LLR is log likelihood ratio, λ_x is a phoneme model, and λ_k is an anti-phoneme model.” *Id.* at 4:15-20; 8:3-9. The ‘196 Patent does not, however, (1) explain what the phoneme or anti-phoneme models in the confidence rate equation could be, (2) provide an example of a phoneme or anti-phoneme model, (3) provide an example of a calculated confidence rate, (4) explain how the calculated confidence rate can be used to perform the comparing and determining steps of claims 1 and 17 or how the preset call command recognition unit is configured to perform the compare and recognize steps of claim 9; or (5) differentiate between the claimed comparing, recognizing, and determining.

Generating and training speech recognition models are complex and difficult problems. The ‘196 Patent certainly doesn’t enable all phoneme models, nor does it enable all anti-phoneme

models (again, no models are disclosed). Particularly without any of (1)-(5) above, it is clear that the inventors did not have possession of the “comparing” and “determining” steps of claims 1 and 17 or the “compare” and “recognize” steps of claim 9, and that they did not enable such an invention for others. Claims 1, 9, and 17 (and all claims that depend therefrom) are invalid for this additional reason.

“confidence rate” (Claims 2, 10, 19); “computed confidence rate” (Claims 2-4, 6, 10-12, 14, 19). Each of claims 2-4, 6, 10-12, 14, 19 recite a “confidence rate” or “computed confidence rate[.]” As discussed above, the ‘196 Patent specification says “the confidence rate is determined by the following equation: $LLR_k(0, \lambda_k) = \log p(0|\lambda_k) - \log p(0|\lambda_x)$ where LLR is log likelihood ratio, λ_x is a phoneme model, and λ_k is an anti-phoneme model.” ‘196 Patent at 4:15-20; 8:3-9. But nowhere in the ‘196 Patent does the applicant describe what the phoneme or anti-phoneme models are, how to generate the phoneme or anti-phoneme models, provide any limits on what the phoneme and anti-phoneme models can be, or provide a working example. Thus, the ‘196 Patent’s confidence rate appears to apply to any phoneme and anti-phoneme models. The ‘196 patent lacks written description and enablement because it does not show (1) that the named inventors were in possession of a confidence rate (or computed confidence rate) generated by any combination of phoneme and anti-phoneme model and (2) because it does not identify to a POSITA how to make or use confidence rates (or computed confidence rates) with any combination of phoneme and anti-phoneme model without undue experimentation.

KAIFI’s infringement contentions further demonstrate that the claimed “confidence rate” and “computed confidence rate” does not comply with § 112, 1st paragraph. For example, KAIFI appears to equate “Natural Language Understanding” as the “phoneme model” and “Hidden Markov Models” as the “anti-phoneme model[.]” *See* Initial Infringement Contentions, Appendix

B at 94-96. Neither of these (incredibly broad) techniques are described or explained in the ‘196 Patent. Thus, to the extent KAIFI’s infringement contentions are credited, “confidence rate” lacks written description and enablement because (1) the ‘196 Patent does not show the named inventors were in possession of either natural language understanding or hidden Markov models and (2) it does not identify to a POSITA how to make or use natural language understanding and hidden Markov models to calculate a confidence rate without undue experimentation.

Thus, claims 2-4, 6, 10-12, 14, 19 (and all claims that depend therefrom) are invalid.

“phoneme model” and “anti-phoneme model” (Claims 6 and 7); “log likelihood ratio” (Claims 5-7, 13-15, 20). As discussed above, and even though neither of these techniques are disclosed in the ‘196 Patent, KAIFI’s infringement contentions equate “Natural Language Understanding” as the “phoneme model” and “Hidden Markov Models” as the “anti-phoneme model[.]” See Initial Infringement Contentions, Appendix B at 94-96. Natural language understanding and hidden Markov models are, broadly speaking, complex techniques that can be used to recognize speech, but they are not specific models that can be plugged into the ‘196 Patent’s confidence rate equation. See ‘196 Patent at 4:15-20; 8:3-9. The claimed “phoneme model” and “anti-phoneme model” lack written description and enablement because the ‘196 Patent (1) does not show the named inventors were in possession of any phoneme or anti-phoneme models, and (2) does not identify to a POSITA how to make or use a phoneme and anti-phoneme models together to calculate a confidence rate without undue experimentation.

To the extent KAIFI’s infringement contentions are credited, the claimed “phoneme model” and “anti-phoneme model” also lack written description and enablement because the ‘196 Patent (1) does not show the named inventors were in possession of either natural language understanding or hidden Markov models and (2) does not identify to a POSITA how to make or

use natural language understanding and hidden Markov models together to calculate a confidence rate without undue experimentation.

Since the claimed “log likelihood ratio” is calculated using the phoneme and anti-phoneme models, it too lacks written description and enablement.

Claims 5-7, 13-15, 20 (and all claims that depend therefrom) are invalid.

“confidence detection unit” (Claims 2, 10-12, 19). The ‘196 Patent fails to adequately describe or enable the “confidence detection unit” of claims 2, 10-12, and 19. The ‘196 Patent specification does not disclose what the confidence detection unit is, nor does it provide any examples. The drawings provide no further information, as the confidence detection unit is shown as a black box at 433. The only boundaries the ‘196 Patent appears to put on the confidence detection unit are functional:

- “a confidence detection unit for detecting the confidence of each recognized call command in relation to whether the recognized call command is an authentic or false command.” ‘196 Patent at 5:49-52;
- “The confidence detection unit 433 is a component for determining whether a recognized keyword is authentic or false by detecting the confidence of the keyword.” *Id.* at 5:65-67.

The ‘196 Patent does not disclose or enable anything that is capable of performing the functions above and, thus, claims 2, 10-12, and 19 do not comply with § 112, first paragraph.

KAIFI’s infringement contentions as to this limitation appear to accuse one or more of the following as being the confidence detection unit:

- Only listening to requests after the wake word is detected;
- Keyword spotting technology;
- Far field speech recognition;
- Deep learning technology;
- A suite of learning techniques, including supervised, semi-supervised, and unsupervised learning;
- A machine learning model; and

- Amazon’s ASR System.

See Initial Infringement Contentions, Appendix B at 79-84, 218-231, 327-333. Amazon disagrees that “confidence detection unit” can be interpreted to cover any one of the extremely broad categories and items identified by KAIFI above. If it can, “confidence detection unit” is further invalid for lack of written description and enablement. As discussed above, the ‘196 Patent does not describe any of these categories or items in connection with a confidence detection unit (or in connection with anything else). Thus, to the extent KAIFI’s infringement contentions are credited, claims 2, 10-12, and 19 (and all claims that depend therefrom) are invalid for written description and enablement for this reason as well.

“rule check unit” (Claims 8, 16). The ‘196 Patent provides even less guidance for “rule check unit” than “confidence detection unit.” The only description of a “rule check unit” in the entire patent is: “[t]he rule check unit 444 is a post-processing component for determining whether the results of the recognition correspond to a recognition target sentence.” ‘196 Patent at 6:1-3. This extremely limited disclosure does not adequately disclose or enable the claimed rule check unit. Crucially, there is no evidence of what the rule check unit is or how it determines recognition of a target sentence. For these reasons, claims 8 and 16 do not comply with § 112, first paragraph.

KAIFI’s infringement contentions as to this limitation appear to accuse one or more of the following as being the rule check unit:

- Amazon’s ASR model;
- A speech recognizer;
- A Speechlet/Skill that looks at intent and determines if it can handle a request;
- Cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies on the recognized actual subsequent input speech command.

See Initial Infringement Contentions, Appendix B at 97-103; 244-245. Amazon disagrees that “rule check unit” can be interpreted to cover any one of the extremely broad categories and items

identified by KAIFI above. If it can, “rule check unit” is further invalid for lack of written description and enablement. As discussed above, the ‘196 Patent does not describe any of these categories or items in connection with a rule check unit (or in connection with anything else). Thus, to the extent KAIFI’s infringement contentions are credited, claims 8 and 16 (and all claims that depend therefrom) are invalid for written description and enablement for this reason as well.

“authentic call command” (Claims 4, 12). Claims 4 and 12 of the ‘196 Patent recite “determining whether or not the input speech is an authentic call command based on the computed confidence rate; and recognizing the input speech as the preset call command if it is determined that the input speech is the authentic call command[]” and “wherein the confidence detection unit is further configured to determine whether or not the assumed input speech is an authentic call command based on the computed confidence rate and recognize the assumed input speech as the preset call command if it is determined that the assumed input speech is the authentic call command[,]” respectively. The ‘196 Patent does not adequately describe or enable the “authentic call command” recited in claims 4 and 12. Instead, the ‘196 Patent’s description/enablement of “authentic call command” is limited to the following:

- “Accordingly, whether the found call command is an authentic call command intended by the speaker or a false call command is determined, thus enabling the recognition of continuous speech to be performed on commands spoken subsequent to the call command.” ‘196 Patent at 5:23-27;
- “The confidence detection unit 433 is a component for determining whether a recognized keyword is authentic or false by detecting the confidence of the keyword.” ‘196 Patent at 5:65-67;
- “The keyword searching unit 423 includes a token passing unit for configuring a speech frame, which includes a silence interval accompanied by noise and call commands from the speaker, in the form of a minimum number of tokens, and then searching for call commands, and a confidence detection unit for detecting the confidence of each recognized call command in relation to whether the recognized call command is an authentic or false command.” ‘196 Patent at 5:40-52.

Thus, although the ‘196 Patent connects an “authentic call command” to a “confidence detection unit” and a “computed confidence rate,” the ‘196 Patent does not provide any guidance or examples of (1) what a confidence detection unit is, (2) what a computed confidence rate is (as discussed above), (4) what “authentic” means in the context of speech, (5) what is authentic speech and what is not authentic speech, or (5) critically, how an “authentic call command” would be determined from a “confidence detection unit” or a “computed confidence rate[.]” In other words, the ‘196 Patent describes the goal of determining an authentic call command, but does not provide any guidance as to how that is accomplished. Claims 4 and 12 (and all claims that depend therefrom) lack written description and enablement for this reason as well.

“device” (claim 1); “apparatus” (claim 9); “single device” (claims 17-19). KAIFI’s infringement contentions appear to accuse devices of the Defendants, devices of non-Defendants, and cloud-based services *together* as meeting these claim limitations. To the extent KAIFI’s construction is credited, the ‘196 Patent does not adequately describe or enable this broad a scope for the terms “device,” “apparatus,” and “single device.” Instead, the ‘196 Patent states “the advantages of continuous speech recognition can be applied even to application fields such as portable devices or embedded systems, which have low computational processing ability and a small storage capacity.” ‘196 Patent at 6:10-13. There is nothing in the ‘196 Patent that describes or enables the use of cloud services, or even connecting the devices of the ‘196 Patent to the cloud or to other devices. For example, no internet, cellular, or other connection is disclosed or explained. For this reason as well, and to the extent KAIFI’s arguments are credited, claims 1, 9, and 17-19 (and all claims that depend therefrom) lack written description and enablement.

“assuming the input speech to be the preset call command based on the computed log likelihood ratio” (claim 5); “wherein the computed log likelihood ratio is a natural logarithm

of a probability on the input speech for a phoneme model of the preset call command, and the computed confidence rate is another log likelihood ratio” (claim 6); “wherein said another log likelihood ratio is computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command” (claim 7). Claim 5 requires “computing a log likelihood ratio of the input speech” and “assuming the input speech to be the preset call command based on the computed log likelihood ratio.” Claim 6 defines the computed log likelihood ratio of claim 5 as “a natural logarithm of a probability on the input speech for a phoneme model of the preset call command.” Claim 6 also introduces another log likelihood ratio by saying “the computed confidence rate is another log likelihood ratio.” As discussed below, “the computed confidence rate” is indefinite for lacking proper antecedent basis. Claim 7 then further defines the another log likelihood ratio as “computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command.”

The ‘196 Patent’s disclosures related to log likelihood ratios are specific and limited. The ‘196 Patent discloses one log likelihood ratio in equation 1, and a second log likelihood ratio in equation 2:

Equation 1: $LLR_k(0, \lambda_k) = \log p(0 | \lambda_k)$ (1) where LLP is the log likelihood ratio, λ_k is a k-th model parameter, and $p(0 | \lambda_k)$ is the probability of the λ_k model of the input speech frame. *See* ‘196 Patent at 7:8-15.

Equation 2 (also called “the confidence rate”): $LLR_k(0, \lambda_k) = \log p(0 | \lambda_k) - \log p(0 | \lambda_{\bar{k}})$ (2) where LLR is log likelihood ratio, λ_k is a phoneme model, and $\lambda_{\bar{k}}$ is an anti-phoneme model. *See id.* at 4:15-20; 8:3-15.

With respect to Equation 1, the ‘196 Patent says “[t]he real-time recognition network for a recorded input speech frame is decoded and is then stored in tokens (Step 2). In this case, the individually stored bit values are sorted according to the log likelihood ratio given by the following Equation [1.]” *Id.* at 7:8-10. With respect to Equation 2, the ‘196 Patent says:

With respect to input speech (call command) assumed to be a call command ‘Naraeya’ or ‘Nepiya’ according to the recognized auto segmentation information, a ratio of the probability $p(0|\lambda k)$ of phoneme models for the call command recognition network to the probability $p(0|\lambda k)$ of anti-phoneme models, other than the assumed phoneme models, is computed, and thus the confidence rate of the call command is computed. This is represented by the following Equation (2)[.]

Id. at 7:63-8:4. Thus, the ‘196 Patent connects assuming input speech to be the preset call command with Equation 2, not Equation 1.

As claimed, Equation 2 must correspond to the “another log likelihood ratio” because Equation 2 is “computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command,” as recited in claim 7. In that case, the claimed computed log likelihood ratio must correspond with Equation 1, but there is no written description or enablement for “assuming the input speech to be the preset call command based on” Equation 1, only Equation 2.

If, on the other hand, you assume the claimed computed log likelihood ratio corresponds with Equation 2, then there is no written description or enablement for “computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command,” as recited in claim 7.

For these reasons, as well as the reasons discussed above for phoneme and anti-phoneme models, claims 5-7 (and all claims that depend therefrom) lack written description and enablement.

“wherein the token passing unit is further configured to compute a log likelihood ratio of the input speech and assume the input speech to be the preset call command based on the computed log likelihood ratio” (claim 13); “wherein the computed log likelihood ratio is a natural logarithm of a probability on the input speech for a phoneme model of the preset call command, and the computed confidence rate is another log likelihood ratio” (claim 14); “wherein said another log likelihood ratio is computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command” (claim 15). Claims 13-15 (and all claims that depend therefrom) are invalid for lack of written description and enablement for reasons similar to those discussed above for claims 5-7. Additionally, the ‘196 Patent provides no written description or enablement for the token passing unit computing a log likelihood ratio.

“in order to reduce the computational load on the device” (claim 1); “in order to reduce overall computational load on the single device” (claim 17). As discussed below in collection with indefiniteness below, the ‘196 Patent does not explain what “reduce” is being compared to, or how much reduction is taking place. Certainly not all reduction in computational load is enabled by the ‘196 Patent, and claims 1 and 17 (and all claims that depend therefrom) lack adequate written description and enablement for the full scope of these terms.

2. Indefiniteness

“A patent is invalid for indefiniteness if its claims, read in light of the specification delineating the patent, and the prosecution history, fail to inform, with reasonable certainty, those skilled in the art about the scope of the invention.” *Nautilus, Inc. v. Biosig Instruments, Inc.*, 134

S. Ct. 2120, 2124 (2014). The patent itself must “describe the exact scope of an invention.” *Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 373 (1996). “The definiteness requirement must take into account the inherent limitations of language,” but at the same time, “a patent must be precise enough to afford clear notice of what is claimed.” *Nautilus, Inc.*, 134 S.Ct. at 2128-29.

The following claim terms are indefinite because they fail to inform a person of ordinary skill the scope of the invention with reasonable certainty:

“preset call command recognition unit;” “continuous speech recognition unit;” “token passing unit;” “confidence detection unit;” “rule check unit;” and “keyword spotting unit” (Claims 1-2, 8-13, 16-17, 19). The ‘196 Patent claims include the following six limitations ending with “unit”: “preset call command recognition unit;” “continuous speech recognition unit;” “token passing unit;” “confidence detection unit;” “rule check unit;” and “keyword spotting unit” (hereinafter, collectively, the “Unit Limitations”). The Unit Limitations are means-plus-function limitations under 35 U.S.C. § 112, sixth paragraph. As discussed below, although the Unit Limitations do not include the words “means” or “for,” the presumption that § 112, sixth paragraph does not apply is overcome because the Unit Limitations fail to recite sufficiently definite structure or else recite function without reciting sufficient structure for performing that function.

Although not employing the word “means,” the Unit Limitations use “unit,” which is a similar nonce word that can operate as a substitute for “means” in the context of § 112, sixth paragraph. The Federal Circuit, the Eastern District of Texas, and other federal district courts have all found “unit” to be a non-structural generic placeholder that can replace “means for” to overcome the presumption that § 112, sixth paragraph, does not apply.

Starting with the claims, each Unit Limitation modifies the “unit” nonce word with functional adjectival qualifiers (e.g., “token passing” and “confidence detection”). The rest of the

surrounding claim language only provides functional descriptions of the Unit Limitations. Additionally, none of the Unit Limitations has a known meaning in the art, let alone a known structural meaning.

Turning now to the specification, none of the Unit Limitations are clearly defined in the specification (and are certainly not clearly defined in terms of structure). Instead, the specification describes the Unit Limitations largely in terms of their functions. It isn't even clear if the Unit Limitations are hardware, software, or something else. Tellingly, the '196 Patent does not even provide examples of any of the Unit Limitations. Additionally, the '196 Patent uses different nonce terms "unit," "network," and "component" interchangeably for some Unit Limitations. This is further evidence that the Unit Limitations are subject to § 112, sixth paragraph.

For at least the reasons discussed above, the Unit Limitations are subject to § 112, sixth paragraph. A means-plus-function claim is indefinite if the specification fails to disclose adequate corresponding structure to perform the claimed function. Because the Unit Limitations are subject to § 112, sixth paragraph, the '196 Patent must describe adequate corresponding structure for the Unit Limitations to comply with § 112, second paragraph. It does not.

A structure disclosed in the specification is corresponding structure only if the specification or prosecution history clearly links or associates that structure to the function recited in the claim. Simply put, the '196 Patent fails to clearly link or associate *any* structure with any of the functions recited in the claim.

At best, the only structure disclosed in the entire '196 Patent is a generic computer. However, because the '196 Patent fails to clearly link or associate this generic computer with any of the functions recited in the claim, it is not corresponding structure for the purpose of § 112, sixth paragraph. The same can be said for KAIFI's apparent construction of "preset call command

recognition unit” and “continuous speech recognition unit” as “a combination of hardware, processing, and software functions both on devices and on the cloud server[.]” The ‘196 Patent does not clearly link or associate “a combination of hardware, processing, and software functions both on devices and on the cloud server” with any of the Unit Limitations. Notably, as mentioned above, the ‘196 Patent does not mention hardware or software. The ‘196 Patent also does not disclose a cloud server, nor does it disclose anything that could connect a device to the cloud, to a server, or to another device (e.g., a cellular network or internet connection).

Even if a generic computer and/or processor or KAIFI’s “combination of hardware, processing, and software functions both on devices and on the cloud server” were considered corresponding structure for the Unit Limitations—neither should be—algorithms sufficient to perform *every* claimed function would still need to be disclosed in the ‘196 Patent for the claims to comply with § 112, second paragraph. But such algorithms are not disclosed. By way of example, the absence of an algorithm sufficient to perform any one of the following claimed functions renders the claims indefinite because the functions appear in each of independent claims 1, 9, and 17:

- Preset call command recognition;
- Receiving an input speech spoken by the speaker;
- Receiving a subsequent input speech spoken, by the speaker, subsequently to the input speech;
- Comparing the input speech with a preset call command;
- Continuous speech recognition;
- Token passing; and
- Assuming the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command.

As discussed above, preset call commands are not disclosed in the ‘196 Patent and, thus, neither is an algorithm sufficient to perform preset call command recognition. As another example, the ‘196 Patent does not disclose an algorithm sufficient to perform continuous speech recognition.

Simply put, the ‘196 Patent does not describe how speech recognition is performed at all. As yet another example, the ‘196 Patent does not disclose an algorithm sufficient to perform token passing. The ‘196 Patent does not describe passing of tokens at all. Similar arguments can be made for every single claimed function of the Unit Limitations.

For at least these reasons, the claims of the ‘196 Patent fail to disclose adequate corresponding structure to perform the claimed functions. Put another way, a person of skill in the art would be unable to recognize sufficient (or any) structure in the ‘196 Patent specification and associate it with the corresponding functions of the claims. As such, the Unit Limitations render the claims invalid as indefinite.

The Unit Limitations are indefinite for additional reasons (regardless of whether they are means plus function limitations). “Preset call command recognition unit;” “continuous speech recognition unit;” “token passing unit;” “confidence detection unit;” “rule check unit;” and “keyword spotting unit” are inherently ambiguous terms whose meets and bounds are not apparent to the public. A person of skill in the art, considering the ‘196 Patent, would not know what they could and could not do to avoid these terms in the claims. None of them have an ordinary meaning in the art, and none of them is defined in the specification. Additionally, the ‘196 Patent provides no examples of the Unit Limitations, and they are only shown in the drawings as black boxes. In this way, the scope of the claims cannot be ascertained and the Unit Limitations render claims 1-2, 8-13, 16-17, 19 (and all claims that depend therefrom) indefinite.

“preset call command recognition unit” (Claims 1-7, 9-17, 19-20). This term renders the claims indefinite as detailed above for the Unit Limitations. This term additionally renders the claims indefinite because, as discussed above for lack of enablement and written description, the ‘196 Patent does not describe or explain the term “preset call command recognition unit.”

Moreover, “preset call command recognition unit” is not a term known in the art. It is unclear what would fall within the scope of the claimed “preset call command recognition unit” and what would fall outside. For these reasons as well, claims 1-7, 9-17, 19-20 (and all claims that depend therefrom) are indefinite.

“preset call command” (Claims 1-7, 9-17, 19-20). As discussed above for enablement and written description, the ‘196 Patent fails to disclose or teach preset call commands. It is unclear what would fall within the scope of the claimed “preset call command” and what would fall outside, and, thus, the claim term and claims 1-7, 9-17, 19-20 (and all claims that depend therefrom) are indefinite.

“continuous speech recognition unit” (Claims 1, 9, 16, 17). This term renders the claims indefinite as detailed above for the Unit Limitations. This term additionally renders the claims indefinite because, as discussed above for lack of enablement and written description, the ‘196 Patent does not describe or explain the term “continuous speech recognition unit” (and, instead, discloses a “continuous speech recognition system,” a “continuous speech recognition network,” a “continuous speech recognition engine,” and a “continuous speech recognition technique,” without explaining these terms or differentiating between them). Moreover, “continuous speech recognition unit” is not a term known in the art. It is unclear what would fall within the scope of the claimed “continuous speech recognition unit” and what would fall outside. For these reasons as well, the claim term and claims 1, 9, 16, 17 (and all claims that depend therefrom) are indefinite.

“token passing unit” (Claims 1, 9, and 17).

This limitation renders the claims indefinite as detailed above for the Unit Limitations. It is also indefinite for additional, independent reasons. First, the term “token” in this limitation renders the claims indefinite. “Token” does not have a specific meaning in the art and is, instead,

used to describe numerous different things. For example, a token can be a word, a phoneme, a portion of speech smaller than a phoneme, a state of a speech recognition model such as a hidden Markov model, etc. Nowhere in the '196 Patent is "token" clearly defined, and the term is used in a circular way that makes it impossible to discern its scope in the claims. For example, the '196 Patent says "each of the predetermined tokens includes the speech frame and a silence interval accompanied by noise," and then says "in this case, the speech frame is configured such that a beam width thereof is limited to 20 or 30 tokens." '196 Patent at 4:3-4; 4:12-13. It is impossible for a token to include the speech frame at the same time a "beam width" (which is not defined) of the speech frame "is limited to 20 or 30 tokens." Claims that require something impossible are indefinite.

Second, this limitation is indefinite because of "token passing[.]" The '196 Patent does not explain or describe the term "token passing." There is a network protocol known in the art called token passing that involves passing of tokens between devices in a network. However, this functionality is not described in the '196 Patent. In the context of speech recognition this functionality has been applied to hidden Markov models. But hidden Markov models are not disclosed in the 196 Patent. If "token passing" means something other than the ordinary and customary meaning in the art, the '196 Patent would need to explain what "token passing" means in the context of the '196 Patent, and it does not.

Further confusing this claim language, as discussed above for lack of enablement/written description, KAIFI's infringement contentions appear to take the position that one or more of any number of broad categories and/or items is the claimed "token passing." If a person of skill in the art were to consider KAIFI's arguments, it would be even more unclear what the scope of this

limitation is. The claim term and claims 1, 9, and 17 (and all claims that depend therefrom) are thus indefinite.

“minimum recognition network composed of a silence interval accompanied by noise and the preset call command” (Claims 1, 9, and 17). As discussed above, the ‘196 Patent does not adequately describe this claim limitation. In particular, it does not distinguish it from the minimum recognition network disclosed in the specification that includes multiple call commands. It also does not explain how a minimum recognition network can be used with a preset call command instead of a configurable call command. Additionally, minimum is a term of degree, the scope or extent of which is not defined by the ‘196 Patent. For these reasons, the claim limitation and claims 1, 9, and 17 (and all claims that depend therefrom) are indefinite.

“actual command” (Claims 1, 9, 16-17). As discussed above for enablement and written description, the ‘196 Patent does not describe or explain what an “actual command” is, or how to determine what is an “actual command.” Moreover, “actual command” is not a term known in the art to have a specific meaning. As such, the term and claims 1, 9, 16-17 (and all claims that depend therefrom) are indefinite.

“in order to reduce the computational load on the device” (claim 1); “in order to reduce overall computational load on the single device” (claim 17). “Reduce” and “reduce overall” are terms of degree, the scope or extent of which is not defined by the ‘196 Patent, nor would a person skilled in the art reading the ‘196 Patent be able to ascertain it. As such, these terms and claims 1 and 17 (and all claims that depend therefrom) are indefinite.

“comparing” and “determining” steps (Claims 1 and 17); “compare” and “recognize” steps (Claim 9). As discussed above for lack of enablement/written description, the ‘196 Patent

does not describe, explain, or differentiate between these steps. As such, the terms and claims 1, 9, and 17 (and all claims that depend therefrom) are indefinite.

“confidence rate” (Claims 2, 10, 19); “computed confidence rate” (Claims 2-4, 6, 10-12, 14, 19). These terms are indefinite because, as discussed above for lack of written description/enablement, the ‘196 Patent does not explain how to calculate these rates and they do not have an accepted meaning in the art. Accordingly, the terms and claims 2-4, 6, 10-12, 14, and 19 (and all claims that depend therefrom) are indefinite.

“[T]he computed confidence rate” limitation of claim 6 is additionally indefinite for lack of an antecedent basis. A claim is indefinite if a term does not have proper antecedent basis where such basis is not otherwise present by implication or the meaning is not reasonably ascertainable. Claim 6 depends on claim 5, which itself depends on claim 1. However, neither claim 5 nor claim 1 mention any “computed confidence rate” upon which claim 6’s “computed confidence rate” could depend. Thus, claim 6’s computed confidence rate does not have any explicit antecedent basis.

Further, an antecedent basis for “the computed confidence rate” is neither implied nor reasonably ascertainable from claim 1 or claim 5. Claim 1 does not describe calculating or imply calculating any rate representing the confidence of any determination. To the contrary, claim 1 only describes making a binary determination of “whether the input speech is recognized as the preset call command.” ‘196 Patent at cl. 1. Such a true-or-false determination does not imply the computation of any rate. Claim 5 discloses “computing a log likelihood ratio of the input speech.” *Id.* at cl. 5. However, claim 6 references “the computed log likelihood ratio” and “the computed confidence rate” as different elements. *Id.* at cl. 6. Because the computed confidence ratio and the computed log likelihood ratio are different elements, claim 5’s “log likelihood ratio”

necessarily cannot be “the computed confidence rate” claimed in claim 6. Thus, there is no explicit or implied antecedent basis for “the computed confidence rate” in claim 1, 5, or 6, and the claim term is indefinite.

“the recognized subsequent input speech” (Claims 8, 16). “[T]he recognized subsequent input speech” limitations of claims 8 and 13 are indefinite for lack of antecedent basis. Claim 8 depends on independent claim 1. Although claim 1 recites “activating a continuous speech recognition unit configured to determine whether the subsequent input speech is recognized as an actual command,” claim 1 does not actually require recognizing a subsequent input speech. Similarly, claim 16 depends from independent claim 9. Independent claim 9 recites “wherein the continuous speech recognition unit is configured to recognize the subsequent input speech as an actual command if the input speech is recognized as the preset call command by the preset call command recognition unit[,]” but it does not require recognizing the subsequent speech. Claim 16 recites “a keyword spotting unit configured to recognize the subsequent input speech as the actual command if the input speech is recognized as the preset call command[,]” and, thus, also does not require recognizing the subsequent speech. Thus, there is no explicit or implied antecedent basis for “the recognized subsequent input speech” in claims 8 and 16 and the claims (and all claims that depend therefrom) are indefinite.

“a preset call command recognition unit using the processor / a continuous speech recognition unit using the processor” (Claim 9). The ‘196 Patent claim terms that recite “using the processor” impermissibly mix statutory classes by simultaneously claiming an apparatus and the method of using the apparatus, and thus are indefinite.

Claims that recite both an apparatus and a method of using that apparatus are indefinite under section 112, paragraph 2. Here, claim 9 of the ‘196 Patent attempts to claim both an

apparatus for recognizing speech, and a method of using the apparatus. The ‘196 Patent claims an “apparatus for recognizing [] speech of a speaker” comprising “a processor, a preset call command recognition unit” and “a continuous speech recognition unit.” ‘196 Patent, cl. 9. The ‘196 Patent then claims a method for using the apparatus by claiming a method for using the processor. Particularly, the ‘196 Patent claims both “recognition unit[s] using the processor” to “recognize [] input speech as the preset call command” and “recognize the subsequent input speech as an actual command.” *Id.* Because claim 9 claims both an apparatus and a method of using the apparatus, such a claim (and all claims that depend therefrom) is not sufficiently precise to provide competitors with an accurate determination of the metes and bounds of protection involved, and is indefinite under § 112, second paragraph.

“wherein the preset call command recognition unit comprises a token passing unit configured to assume the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command (Claims 1, 9, 17). This clause is indefinite for the several reasons discussed above including the recognition unit and token passing unit limitations. This clause is additionally indefinite because it is not clear how a preset call command recognition unit comprising a token passing unit could use “the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command.”

First, and as discussed above, “minimum recognition network” is not a phrase known in the art. Nor is it ever explicitly defined in the ‘196 Patent. Second, “composed of” is closed language. The ‘196 Patent does not describe how a network of only a silence interval accompanied by noise and the preset call command can be used for a comparison. To the contrary, the ‘196

Patent teaches a recognition network that includes tokens. To the extent “minimum recognition network” means a minimum number of tokens, it is not clear what token means, as discussed above.

Claims 1, 9, and 17 (and all claims that depend therefrom) are indefinite for this reason as well.

“phoneme model” and “anti-phoneme model” (Claims 6 and 7); “log likelihood ratio” (Claims 5-7, 13-15, 20). As discussed above, the ‘196 Patent does not describe or explain “phoneme model” or “anti-phoneme model.” Accordingly, the terms are indefinite. Since it uses the phoneme and anti-phoneme models, “log likelihood ratio” is indefinite for similar reasons. Claims 5-7, 13-15, and 20 (and all claims that depend therefrom) are indefinite.

“assuming the input speech to be the preset call command based on the computed log likelihood ratio” (claim 5); “wherein the computed log likelihood ratio is a natural logarithm of a probability on the input speech for a phoneme model of the preset call command, and the computed confidence rate is another log likelihood ratio” (claim 6); “wherein said another log likelihood ratio is computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command” (claim 7); “wherein the token passing unit is further configured to compute a log likelihood ratio of the input speech and assume the input speech to be the preset call command based on the computed log likelihood ratio” (claim 13); “wherein the computed log likelihood ratio is a natural logarithm of a probability on the input speech for a phoneme model of the preset call command, and the computed confidence rate is another log likelihood ratio” (claim 14); “wherein said another log likelihood ratio is computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the

preset call command” (claim 15). As discussed above for § 112, first paragraph, the ‘196 Patent does not describe or explain the claimed “computed log likelihood ratio” or “another log likelihood ratio.” Accordingly, these clauses are indefinite, because they do not reasonably inform a person of ordinary skill in the art what would and would not be within their scope. For this reason as well, claims 5-7 and 13-15 (and all claims that depend therefrom) are indefinite.

“authentic call command” (Claims 4, 12). As discussed above, the ‘196 Patent does not describe or explain “authentic call command.” Accordingly, the term is indefinite because it does not convey to a person of ordinary skill in the art what does and does not fall within its scope. For this reason as well, claims 4 and 12 (and all claims that depend therefrom) are indefinite.

I. DISCLOSURES PURSUANT TO P.R. 3-4

Defendants’ previous Patent Local Rule 3-4(a) document production can be found at AMZ_KAIFI_00003726 - AMZ_KAIFI_00048949. Defendants previous Local Rule 3-4(b) document production can be found at AMZ_KAIFI_00048950 - AMZ_KAIFI_00068331. Defendants hereby produce additional Local Rule 3-4(b) documents as AMZ_KAIFI_00080289 - AMZ_KAIFI_00087938. Certain of these documents have been designated under the terms of the Protective Order. Amazon.com, Inc.’s Petitions for Inter Partes Review of the Asserted Patents (including all relevant exhibits, some of which are cited above) have been produced at AMZ_KAIFI_00069399 - AMZ_KAIFI_00078431. These Petitions (and all of their exhibits) are fully incorporated herein.

Dated: March 14, 2025

Respectfully submitted,

HUESTON HENNIGAN LLP

By: /s/ Chrsitina V. Rayburn
Christina V. Rayburn (Lead Counsel)
Neil G. Anderson
Thomas B. King
Karen L. Younkens
Christine Woodin
HUESTON HENNIGAN LLP
620 Newport Center Dr., Ste. 1300
Newport Beach, California 92660
Telephone: 949-287-5940
Facsimile: 888-866-4825
Crayburn@hueston.com
Nanderson@hueston.com
Tking@hueston.com
Kyounkens@hueston.com
Cwoodin@hueston.com
*Attorneys for Defendants Amazon.com,
Inc., Amazon.com Services LLC, and
Amazon Web Services, Inc.*

Deron R. Dacus
The Dacus Law Firm PC
Ddacus@dacusfirm.com
821 ESE Loop 323
Tyler, Texas 75701
Telephone: 903-705-1117
Facsimile: 903-581-2543

CERTIFICATE OF SERVICE

I hereby certify that a true and correct copy of the above and foregoing document has been served on all counsel of record via email on March 14, 2025.

/s/ Chrsitina V. Rayburn

Christina V. Rayburn