

Eligibility Contentions Appendix A – U.S. Patent No. 8,930,196

The following charts contain Defendants’ Subject-Matter Eligibility Contentions demonstrating that claims 1-20 of U.S. Patent No. 8,930,196 (the “’196 Patent”) are patent-ineligible under 35 U.S.C. § 101 pursuant to the Court’s Standing Order Regarding Subject Matter Eligibility Contentions. Because the Court has yet to issue a claim construction in this case and—to the extent that KAIFI contends that there are factual disputes that relate to eligibility of the asserted claims of the ’196 Patent—fact and expert discovery is ongoing, these contentions are preliminary only and Defendants reserve the right to supplement or modify these contentions in accordance with the Court’s Local Rules and the Docket Control Order in this case. Additionally, and in further consideration of the preliminary stage of the case, Defendants note that the pinpoint citations referenced in these charts are not exhaustive, and Defendants reserve the right to rely on additional legal case authority. Further, the citations to intrinsic and extrinsic evidence are exemplary only, and Defendants reserve the right to rely on other portions of intrinsic and extrinsic evidence not expressly cited herein, including but not limited to evidence from the prior art and evidence from the specification, claims and prosecution history of the asserted patents.

Further, these charts incorporate KAIFI’s apparent interpretation of the elements within the asserted claims, as reflected in KAIFI’s infringement contentions to date. Defendants do not concede that such interpretation is correct, and reserves its rights to supplement these contentions accordingly.

Defendants reserve their right to supplement these contentions based on further discovery, including any supplemental infringement contentions or any interrogatory response provided by KAIFI purporting to rebut these ineligibility contentions. Defendants further reserve their right to supplement or amend these contentions based on any claim constructions provided by the Court.

Pursuant to the Court’s Standing Order Regarding Subject Matter Eligibility Contentions, Defendants provide the following chart identifying each exception to eligibility (e.g., abstract idea, law of nature, and natural phenomenon) to which each Challenged Claim is directed and the factual and legal basis therefor. Defendants contend that Claim 1 of the ’196 Patent is representative of the Challenged Claims 1-20, but Defendants separately sets forth herein the exception to eligibility for each.

Challenged Claims	Exception to Eligibility, and Basis Therefor
1. A method for recognizing a speech of a speaker by a device, the method comprising: receiving, by a preset call command recognition unit, an input speech spoken by the speaker;	All of the Challenged Claims are directed to the abstract idea of recognizing a particular word and then listening for speech after the spoken keyword. This is what humans do every day. For example, when our name is spoken (e.g., “Andrew”), we hear it, and we then listen for what is spoken next, such as an instruction or a request (e.g., “Andrew, will you please take out the trash?”). The challenged claims are also directed to the abstract idea of recognizing the particular word using (a) what we know the word sounds like, (b) what we know other words and other noise sound like, and (c) what we know silence sounds like. Again, this is what humans do in normal, everyday conversation. For example, we know when “Andrew” is spoken because it comports with what we know “Andrew” sounds like and because “Andrew” doesn’t sound like what we know

Challenged Claims	Exception to Eligibility, and Basis Therefor
<p>receiving, by the preset call command recognition unit, a subsequent input speech spoken, by the speaker, subsequently to the input speech;</p> <p>comparing, by the preset call command recognition unit, the input speech with a preset call command;</p> <p>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; and</p> <p>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; and</p> <p>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</p>	<p>other words like “Megan” and “John” sound like, or what silence sounds like, or what extraneous noise like laughing sounds like.</p> <p>35 U.S.C. § 101 lists four broad categories of patent-eligible subject matter, but is subject to “an important implicit exception” that “[l]aws of nature, natural phenomena, and abstract ideas,” which form the “basic tools of scientific and technological work,” are not eligible for patent protection. <i>Alice Corp. Pty. Ltd. v. CLS Bank Int’l</i>, 573 U.S. 208, 216 (2014) (internal quotation marks omitted). “The ‘abstract ideas’ category embodies ‘the longstanding rule that an idea of itself is not patentable.’” <i>Id.</i> at 209 (quoting <i>Gottschalk v. Benson</i>, 409 U.S. 63, 67 (1972) (quoting <i>Rubber-Tip Pencil Co. v. Howard</i>, 87 U.S. 498 (1874))).</p> <p><i>Alice</i> set forth a two-step inquiry. “First, we determine whether the claims at issue are directed to one of those patent-ineligible concepts.” <i>Id.</i> at 217. In making this inquiry, the claims “must be considered as a whole.” <i>Alice</i>, 573 U.S. at 218 n.3 (quoting <i>Diamond v. Diehr</i>, 450 U.S. 175, 188 (1981)). And courts consider “what the patent asserts to be the focus of the claimed advance over the prior art.” <i>Solutran, Inc. v. Elavon, Inc.</i>, 931 F.3d 1161, 1168 (Fed. Cir. 2019) (cleaned up). “Instead of a definition [for abstract idea] ... courts ... examine earlier cases in which a similar or parallel descriptive nature can be seen—what prior cases were about, and which way they were decided.” <i>Amdocs (Isr.) Ltd v. Openet Telecom, Inc.</i>, 841 F.3d 1288, 1294 (Fed. Cir. 2016).</p> <p>If the claims recite an abstract idea, the second step requires a “search for an ‘inventive concept’— an element or combination of elements that is ‘sufficient to ensure that the patent in practice amounts to significantly more than a patent upon the [ineligible concept] itself.’” <i>Id.</i> at 218 (quoting <i>Mayo Collaborative Servs. v. Prometheus Labs., Inc.</i>, 566 U.S. 66, 73 (2016)). “A claim that recites an abstract idea must include ‘additional features’ to ensure ‘that the claim is more than a drafting effort designed to monopolize the abstract idea.’” <i>Id.</i> at 221 (quoting <i>Mayo</i>, 566 U.S. at 78) (alterations omitted). The recited “additional features must be more than well-understood, routine, conventional activity.” <i>Ultramercial, Inc. v. Hulu, LLC</i>, 772 F.3d 1335, 1339 (Fed. Cir. 2013) (internal quotations omitted).</p> <p>Recognizing a word based on what it, other noises, and silence sounds like, and then listening for speech after recognizing the word, as claimed, is an abstract idea.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
<p>minimum recognition network composed of a silence interval accompanied by noise and the preset call command.</p>	<p>First, the claim can be performed by the human mind alone and is, therefore, directed to an abstract idea. <i>See, e.g., CyberSource Corp. v. Retail Decisions, Inc.</i>, 654 F.3d 1366, 1373 (Fed. Cir. 2011) (a “method that can be performed by human thought alone” is an abstract idea); <i>Alice</i>, 573 U.S. at 208 (“method[s] of organizing human activity” are abstract ideas); <i>PersonalWeb Techs. LLC v. Google LLC</i>, 8 F.4th 1310, 1316 (Fed. Cir. 2021) (mental processes are “a telltale sign of abstraction”). As discussed above, humans recognize a keyword (such as their name) and once recognized, listen for whatever is said next. The human mind does so based on what it knows the word sounds like, what other words and noises sound like, as well as what silence sounds like.</p> <p>“In the context of software-based inventions, <i>Alice/Mayo</i> step one often turns on whether the claims focus on the specific asserted improvement in computer capabilities or, instead, on a process that qualifies as an abstract idea for which computers are invoked merely as a tool.” <i>Trinity Info Media, LLC v. Covalent, Inc.</i>, 72 F.4th 1355, 1362–63 (Fed. Cir. 2023). There are no asserted improvements in computer capabilities disclosed in the ’196 Patent. Instead, the ’196 Patent is clear that its alleged advancement over the prior art is using call commands or keywords to wake up some sort of generic device or computer (that is not described in the patent). <i>See</i> ’196 Patent at 2:43-3:20. However, as discussed above, using a keyword or a call command to draw someone’s attention to what happens next (such as a request or command) is a human activity that can be performed in the human mind. Such ideas are not patent eligible. <i>See, e.g., Affinity Labs of Texas, LLC v. DIRECTV, LLC</i>, 838 F.3d 1253, 1257 (Fed. Cir. 2016) (if the “focus of the claimed advance over the prior art” shows that “the claim’s character as a whole is directed to” steps that “can be performed in the human mind, or by a human using a pen and paper,” the claim is directed to an abstract idea) (cleaned up).</p> <p>The Challenged Claims merely describe desired functions without providing any technical details, further suggesting the challenged claims are directed to an abstract idea. <i>See, e.g., Affinity Labs of Texas, LLC v. Amazon.com, Inc.</i>, 838 F.3d 1266 (Fed. Cir. 2016) (finding claims that did “no more than describe a desired function or outcome, without providing any limiting detail that confines the claim to a particular solution to an identified problem[.]” abstract); <i>IPA Technologies, Inc. v. Amazon.com, Inc.</i>, 307 F. Supp. 3d 356 (D. Del. 2018) (finding claims that recited “the basic steps that would be required” to achieve a particular aim abstract); <i>Univ. of Fla. Rsch. Found., Inc. v. Gen. Elec. Co.</i>, 916 F.3d 1363, 1368 (Fed. Cir. 2019) (finding claims abstract where they “fail[ed] to provide any technical details for the tangible components, ... instead predominately describ[ing] the system and methods in purely functional terms.”) Like the claims, the ’196 Patent specification uses primarily functional terminology to describe the alleged invention. <i>See, e.g., Pebble Tide LLC v. Arlo Techs.</i>, No. CV 18-1767-LPS, 2020 WL 509183, at *1 (D. Del. Jan. 31, 2020) (“This</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>conclusion at Step One is supported by the fact that the representative claim lacks limiting technical details. Neither of the claims, nor for that matter the specification, explain[s] how the claimed invention’s components perform their recited functions. Rather, they describe those components in purely functional terms.”). As one example, the ’196 Patent does not disclose how it performs the claimed speech recognition functions. As another example, the ’196 Patent does not disclose how it performs the claimed token passing function.</p> <p>Courts considering similar claims have routinely found them abstract (and ineligible). <i>See, e.g., Content Extraction & Transmission LLC v. Wells Fargo Bank, N.A.</i>, 776 F.3d 1343 (Fed. Cir. 2014) (“concept of data collection, recognition, and storage” is abstract idea); <i>Enco Systems, Inc. v. DaVincia, LLC</i>, 447 F.Supp.3d 916 (E.D. Mo. 2020), <i>aff’d Enco Sys., Inc. v. DaVincia, LLC</i>, 845 F. App’x 953 (Fed. Cir. 2021) (finding claims directed to employing speech-to-text technology abstract); <i>Parus Holdings, Inc. v. Sallie Mae Bank</i>, 137 F. Supp. 3d 660, 672 (D. Del. 2015), <i>aff’d</i>, 677 F. App’x 682 (Fed. Cir. 2017) (finding a “computer and telecommunications network for receiving, sending and managing information from a subscriber to the network and from the network to a subscriber” including via speech recognition abstract); <i>MModal Servs. Ltd. v. Nuance Commc’ns, Inc.</i>, No. 1:18-CV-00901, 2019 WL 13059774, at *1 (N.D. Ga. May 13, 2019) (finding claims directed to a “free-form narration of a patient encounter provided by a clinician,’ provid[ing] ‘to a user the one or more options’ corresponding to particular semantic concepts, and ‘updating a textual representation of the free-form narration’ based on the correlations” abstract); <i>In re TLI Communications LLC Patent Litigation</i>, 87 F.Supp.3d 773 (E.D. Va. 2015) (finding “taking, organizing, classifying, and storing photographs” is an abstract idea where claims could be performed by human thought alone, and further finding that limiting this idea to speech does not make the concept patent eligible); <i>Quad City Pat., LLC v. Zoosk, Inc.</i>, 498 F. Supp. 3d 1178, 1185 (N.D. Cal. 2020) (“First, detecting participant ‘speech, language, emotion, social intelligence, character and characteristics’ is routinely performed by humans when conducting face-to-face transactions. More importantly, the claims do not recite any novel computer implementation of such analysis. Instead, the claim ‘describe[s] a desired function or outcome’ of recognizing participant characteristics without limiting that function to any technical solution. <i>Affinity Labs of Tex.</i>, 838 F.3d at 1269. Finally, the limitation is not directed to improving computer functionality. Quad City does not identify any technical problem that is resolved by identifying participant characteristics. Instead, the benefits of recognizing participant characteristics by a computer appear to be the same as those of recognizing participant characteristics using human beings—namely, better marketing.”); <i>W. View Rsch., LLC v. Audi AG</i>, 685 F. App’x 923 (Fed. Cir. 2017) (“receiving or collecting data queries, analyzing the data query, retrieving and processing the information constituting a response to the initial data query, and generating a visual or audio response to the initial data query,” including involving speech recognition, is an abstract idea).</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>To the extent this abstract idea is considered more than one such ideas, the combination of these abstract ideas does not render Claim 1 any less abstract. <i>See Sensormatic Elecs., LLC v. Wyze Labs, Inc.</i>, 484 F. Supp. 3d 161, 165 (D. Del. 2020), <i>aff'd</i>, No. 2020-2320, 2021 WL 2944838 (Fed. Cir. July 14, 2021) (“Courts often invalidate patents that are directed to a combination of abstract ideas.”) (citing cases); <i>Broadband iTV, Inc. v. Amazon.com, Inc.</i>, No. 6:20-CV-00921-ADA, 2022 WL 4703425, at *14 (“Combining two abstract ideas . . . is not inventive.”); <i>RecogniCorp. LLC v. Nintendo Co.</i>, 855 F.3d 1322, 1327 (Fed. Cir. 2017) (“Adding one abstract idea . . . to another abstract idea . . . does not render the claim non-abstract.”); <i>Consumer 2.0, Inc. v. Tenant Turner, Inc.</i>, 343 F. Supp. 3d 581, 590-91 (E.D. Va. 2018), <i>aff'd</i> 796 F. App’x 752 (Fed. Cir. 2020) (“[T]he mere combination of these abstract processes is insufficient to bring the patent-at-issue out of the realm of the abstract.”) (citing cases).</p> <p>Claim 1 of the ’196 Patent does not contain an inventive concept such that it does more than claim an ineligible abstract idea. Instead, it recites only broad and result-based aspirational and functional language such as “comparing . . . the input speech with a preset call command,” “determining . . . whether the input speech is recognized as the preset call command according to a result of the comparison,” and “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.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. <i>See, e.g., Two-Way Media Ltd. v. Comcast Cable Commc’ns, LLC</i>, 874 F.3d 1329, 1337 (Fed. Cir. 2017) (finding the use of results-based functional language failed to “sufficiently describe how to achieve these results in a non-abstract way”).</p> <p>At best, Claim 1 seeks to apply the abstract idea(s) using a generic electronic “device” or computer structure. But “[i]nstructing one to ‘apply’ an abstract idea and reciting no more than generic computer elements performing generic computer tasks does not make an abstract idea patent-eligible.” <i>Intellectual Ventures I LLC v. Capital One Bank (USA)</i>, 792 F.3d 1363, 1368 (Fed. Cir. 2015) (citing <i>Alice</i>, 134 S. Ct. at 2359-60). The ’196 Patent does not describe the claimed device or any of the claimed Unit Limitations in a meaningful way.¹ Accordingly, there is nothing in the specification to suggest that the claimed device or any of the Unit</p>

¹ 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” (collectively, the “Unit Limitations”). As discussed in Defendants’ Invalidity Contentions Cover

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Limitations are more than generic, off the rack structures well known in the art. Put another way, if any of them <i>were</i> something more than generic, off the rack structures well known in the art, the '196 Patent would need to explain that, and it does not. Again, the '196 Patent is clear that its alleged advancement over the prior art is using call commands or keywords to wake up a speech recognition device. There is no allegation of advancement or improvements to the claimed device/single device or whatever the Unit Limitations in the '196 Patent are. This is insufficient to create an inventive concept. <i>See, e.g., Enco</i> (“nowhere does the 191 Patent suggest its purpose is to specifically improve speech-to-text processing software, and nowhere does it explain how this is to be accomplished—across myriad software iterations—at a technical level. Simply put, there are references to things like ‘autoflush counters’ and computer-determined communication relays between speech-to-text processors and encoders, but nothing at all to explain how those functional operations are implemented at an inventive level. This is insufficient to create an inventive concept that transforms the patent into something more than a patent upon the ineligible concept itself.”).</p> <p>Furthermore, the '196 Patent admits that (1) speech recognition systems for recognizing single words were known, (2) speech recognition systems for recognizing many words were known, and (3) that it was known to use recognition networks to recognize speech:</p> <p style="padding-left: 40px;">Speech recognition systems using the above-described various access methods can be classified into isolated word speech recognition systems that recognize isolated words according to the type of vocalization, and continuous speech recognition systems that recognize speech made by continuously speaking words.</p> <p style="padding-left: 40px;">Of these speech recognition systems, an isolated word speech recognition system is a scheme for performing recognition in the sequence of Voice Activity Detection (VAD) feature extraction→pattern comparison→recognition, and is a technology which is suitable for small computational loads or small-scale speech recognition, but is unsuitable for commercialization because unregistered words are rejected only by confidence detection with the result that errors frequently occur in confidence detection depending on noisy environments, and because a special event such as a recording button for speech recognition is required.</p>

Pleading, the Unit Limitations are means-plus-function limitations under 35 U.S.C. § 112, sixth paragraph, and render the claims indefinite under 35 U.S.C. § 112, second paragraph, for failing to disclose adequate corresponding structure. None of the Unit Limitations are clearly discussed or defined in the specification (and are certainly not clearly defined in terms of structure). Instead, the specification describes the Unit Limitations and terms similar to the Unit Limitations (without explaining the differences) largely in terms of their functions. Moreover, the '196 Patent does not provide examples of any of the Unit Limitations, and the '196 Patent may use different nonce terms “unit,” “network,” and “component” interchangeably and unclearly.

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>FIG. 1 is a diagram briefly showing the construction of a conventional continuous speech recognition network. The continuous speech recognition network employs a scheme that performs recognition by performing post-processing on a recognized word sequence using a language model. This scheme is mainly used in large-capacity speech recognition systems of ten thousand or more words.</p> <p>A continuous speech recognition system is a scheme for performing recognition in the sequence of feature extraction→pattern comparison (acoustic model comparison)→language model comparison→post-processing→recognition, is suitable for large-scale speech recognition systems in high-specification server level Personal Computers (PCs), and also enables a real-time large-scale speech recognition system to be implemented with the development of storage media and computational processing abilities. However, there are disadvantages because, in order to install such a continuous speech recognition system in terminals which have gradually become lightweight, still as of yet insufficient storage capacity and complicated floating computation cause problems, and a large computational load and storage capacity for a large amount of data are required.</p> <p>'196 Patent at 1:44-2:16; <i>see also</i> Fig. 1, 6:4-9.</p> <p>The specification also makes clear that the devices disclosed in the '196 Patent are generic, off-the-rack structures well-known in the art:</p> <p style="padding-left: 40px;">Therefore, 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.</p> <p>'196 Patent at 6:10-13. There is nothing new or novel alleged about these devices.</p> <p>Moreover, this claim would preempt the broad subject matter of using a device to recognize a particular word based on what it and other sounds and silence sound like, and then listening for speech after the keyword. The breadth of this claimed subject matter further suggests that the claims are directed to an abstract idea and are patent-ineligible subject matter. <i>See, e.g., Bancorp Servs., LLC v. Sun Life Assur. Co. of Canada (U.S.)</i>, 687</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>F.3d 1266, 1280 (Fed. Cir. 2012) (finding that claimed abstract idea would improperly preempt broad subject matter).</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions, which are fully incorporated herein.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>2. The method of claim 1, wherein determining whether the input speech is recognized as the preset call command further comprises:</p> <p>computing, by a confidence detection unit, a confidence rate of the input speech; and</p> <p>recognizing, by the confidence detection unit, the input speech as the preset call command based on the computed confidence rate.</p>	<p>Like claim 1, this claim is directed to the abstract idea of recognizing a particular word based on what we know it and other sounds and silence sound like, and then listening for speech after recognizing the particular word. <i>See supra</i> claim 1. Claim 2 adds the additional abstract idea of recognizing the particular word, or keyword, based on how confident we are it has been spoken.</p> <p>Like claim 1, the subject matter of claim 2 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves determining how confident we are that a spoken word is one we know. Also like claim 1, determining confidence is not an improvement in computer capabilities, and claim 2 merely describes desired functions without providing any technical details.</p> <p>Claim 2 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “computing ... a confidence rate of the input speech” and “recognizing ... the input speech as the preset call command based on the computed confidence rate.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. For example, the ’196 Patent does not explain how a confidence score is computed (particularly because, as explained in Defendants Invalidity Contentions, the ’196 Patent does not disclose the speech recognition model(s) used), nor does it explain how it is used to recognize input speech.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>At best, claim 2 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>3. The method of claim 2, wherein recognizing the input speech comprises:</p> <p>recognizing the input speech as the preset call command if the computed confidence rate is higher than a threshold value.</p>	<p>Like claim 2, this claim is directed to the abstract idea(s) of (1) recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword, and (2) recognizing the keyword based on how confident we are it has been spoken. <i>See supra</i> claims 1, 2. Claim 3 adds the additional abstract idea of recognizing the keyword based on a certain threshold level of confidence.</p> <p>Like claims 1 and 2, the subject matter of claim 3 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves determining how confident we are that spoken word is one we know and can recognize. At a certain point (or threshold), we become confident enough that a word has been spoken and we accept that it has been spoken. This happened so instantaneously and regularly, we don’t even think about it. Also like claims 1 and 2, a threshold value for confidence is not an improvement in computer capabilities, and claim 3 merely describes desired functions without providing any technical details. For example, the ’196 Patent provides no insight as to how the threshold value is calculated, nor goes it provide any examples of the claimed threshold value.</p> <p>Claim 3 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “recognizing the input speech as the preset call command if the computed confidence rate is higher than a threshold value.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. For</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>example, the '196 Patent does not meaningfully describe this threshold value, does not explain how to determine this threshold value, and does not provide any examples of this threshold value.</p> <p>At best, claim 3 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the '196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>4. The method of claim 2, wherein recognizing the input speech comprises:</p> <p>determining whether or not the input speech is an authentic call command based on the computed confidence rate; and</p> <p>recognizing the input speech as the preset call command if it is determined that the input speech is the authentic call command.</p>	<p>Like claim 2, this claim is directed to the abstract idea(s) of (1) recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword, and (2) recognizing the keyword based on how confident we are it has been spoken. <i>See supra</i> claims 1, 2. Claim 4 adds the additional abstract idea of recognizing the keyword if we think it is authentic.</p> <p>Like claims 1 and 2, the subject matter of claim 4 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves determining whether we think the speech is authentic. Also like claims 1 and 2, the authenticity of speech is not an improvement in computer capabilities, and claim 4 merely describes desired functions without providing any technical details.</p> <p>Claim 4 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “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.” But the '196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. For example, the '196 Patent does not meaningfully describe what is or is not an authentic call command, nor does it describe how to recognize speech if it is authentic.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>At best, claim 4 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>5. The method of claim 1, wherein determining the input speech is recognized as the preset call command further comprises:</p> <p>computing a log likelihood ratio of the input speech; and</p> <p>assuming the input speech to be the preset call command based on the computed log likelihood ratio.</p>	<p>Like claim 1, this claim is directed to the abstract idea of recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword. <i>See supra</i> claim 1. Claim 5 adds the additional abstract idea of recognizing the keyword based on how likely it is the keyword has been spoken.</p> <p>Like claim 1, the subject matter of claim 5 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves determining how confident we are that spoken word is one we know and can recognize. In other words, we are determining how likely it is that a particular word has been spoken. This is what humans do in normal conversation. Also like claim 1, the likelihood of a keyword is not an improvement in computer capabilities, and claim 5 merely describes desired functions without providing any technical details. A POSITA would understand that likelihood and/or confidence values are often stored as logarithms to save processing resources.</p> <p>Claim 5 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “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.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. For example, the ’196 Patent does not provide any description or explanation of the models used to calculate the claimed likelihood ratios.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p data-bbox="583 248 1850 280">At best, claim 5 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p data-bbox="583 354 1990 532">The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p data-bbox="583 605 1955 711">Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p data-bbox="111 743 558 1027">6. The method of 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.</p>	<p data-bbox="583 743 1976 922">Like claims 1 and 5, this claim is directed to the abstract idea(s) of (1) recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword, and (2) recognizing the keyword based on how likely it is the keyword has been spoken. <i>See supra</i> claims 1, 5. Claim 6 adds the additional abstract idea of using the keyword to determine how likely it is the keyword has been spoken.</p> <p data-bbox="583 963 1986 1182">Like claims 1 and 5, the subject matter of claim 6 can be performed in the human mind. Using what we know a word sounds like (a model) is exactly what a human uses to determine the probability that the word has been spoken. Also like claim 1, using the word or a model of the word to generate a probability that the word has been spoken is not an improvement in computer capabilities, and claim 6 merely describes desired functions without providing any technical details. As best as Defendants can tell because the term is not defined or explained in the ’196 Patent, a phoneme model is simply a model of what a word sounds like.</p> <p data-bbox="583 1255 1850 1287">At best, claim 6 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p data-bbox="583 1360 1955 1466">The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>7. The method of 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.</p>	<p>Like claims 1, 5 and 6, this claim is directed to the abstract idea(s) of (1) recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword, (2) recognizing the keyword based on how likely it is the keyword has been spoken, and (3) using the keyword to determine how likely it is the keyword has been spoken. <i>See supra</i> claims 1, 5, 6. Claim 6 adds the additional abstract idea of using sound other than the word to determine how likely it is that the keyword has been spoken.</p> <p>Like claims 1, 5 and 6, the subject matter of claim 7 can be performed in the human mind. Using what we know a word sounds like (a model) and what other sounds sound like (an anti model) is exactly what a human uses to determine the probability or likelihood that the word has been spoken. Also like claims 1, 5 and 6, using an anti model of the word to generate a probability that the word has been spoken is not an improvement in computer capabilities, and claim 7 merely describes desired functions without providing any technical details. As best as Defendants can tell because the term is not defined or explained in the ’196 Patent, an anti-phoneme model is simply a model of what a word doesn’t sound like.</p> <p>At best, claim 7 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>8. The method of claim 1, further comprising:</p> <p>determining, using a rule check unit, whether or not the recognized subsequent input speech corresponds to a target sentence.</p>	<p>Like claim 1, this claim is directed to the abstract idea of recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword. <i>See supra</i> claim 1. Claim 8 adds the additional abstract idea of recognizing speech after the keyword.</p> <p>Like claim 1, the subject matter of claim 8 can be performed in the human mind. To use an example above, we understand that “Andrew, please take out the trash” involves both a keyword (“Andrew”) and a subsequent target sentence (“please take out the trash”). Determining whether what comes after the keyword is a particular request or other sentence that needs your attention is what humans do in everyday conversation. This is simply describing hearing and understanding the speech of others. Also like claim 1, recognizing speech after the keyword is not an improvement in computer capabilities, and claim 8 merely describes desired functions without providing any technical details.</p> <p>Claim 8 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “determining, using a rule check unit, whether or not the recognized subsequent input speech corresponds to a target sentence.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. For example, the ’196 Patent does not provide any meaningful explanation of what the rule check unit is, or how the rule check unit determines whether speech is a target sentence. There are certainly no rules disclosed in the ’196 Patent that would allow something to determine whether speech is a target sentence.</p> <p>At best, claim 8 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>9. An apparatus for recognizing a speech of a speaker, the apparatus comprising:</p> <p>a processor,</p> <p>a preset call command recognition unit using the processor; and</p> <p>a continuous speech recognition unit using the processor,</p> <p>wherein the preset call command recognition unit is configured to receive an input speech spoken by the speaker;</p> <p>receive a subsequent input speech spoken, by the speaker, subsequently to the input speech;</p> <p>compare the input speech with a preset call command; and</p> <p>recognize the input speech as the preset call command according to a result of the comparison, 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</p>	<p>All of the Challenged Claims are directed to the abstract idea of recognizing a keyword and then listening for speech after the spoken keyword. This is what humans do every day. For example, when our name is spoken (e.g., “Andrew”), we hear it, and we then listen for what is spoken next, such as an instruction or a request (e.g., “Andrew, will you please take out the trash?”). The challenged claims are also directed to the abstract idea of recognizing the spoken keyword using (a) what we know the keyword sounds like, (b) what we know other words and other noise sound like, and (c) what we know silence sounds like. Again, this is what humans do in normal, everyday conversation. For example, we know when “Andrew” is spoken because it comports with what we know “Andrew” sounds like and because “Andrew” doesn’t sound like what we know other words like “Megan” and “John” sound like or what silence or extraneous noise like laughing sounds like.</p> <p>35 U.S.C. § 101 lists four broad categories of patent-eligible subject matter, but is subject to “an important implicit exception” that “[l]aws of nature, natural phenomena, and abstract ideas,” which form the “basic tools of scientific and technological work,” are not eligible for patent protection. <i>Alice Corp. Pty. Ltd. v. CLS Bank Int’l</i>, 573 U.S. 208, 216 (2014) (internal quotation marks omitted). “The ‘abstract ideas’ category embodies ‘the longstanding rule that an idea of itself is not patentable.’” <i>Id.</i> at 209 (quoting <i>Gottschalk v. Benson</i>, 409 U.S. 63, 67 (1972) (quoting <i>Rubber-Tip Pencil Co. v. Howard</i>, 87 U.S. 498 (1874))).</p> <p><i>Alice</i> set forth a two-step inquiry. “First, we determine whether the claims at issue are directed to one of those patent-ineligible concepts.” <i>Id.</i> at 217. In making this inquiry, the claims “must be considered as a whole.” <i>Alice</i>, 573 U.S. at 218 n.3 (quoting <i>Diamond v. Diehr</i>, 450 U.S. 175, 188 (1981)). And courts consider “what the patent asserts to be the focus of the claimed advance over the prior art.” <i>Solutran, Inc. v. Elavon, Inc.</i>, 931 F.3d 1161, 1168 (Fed. Cir. 2019) (cleaned up). “Instead of a definition [for abstract idea] ... courts ... examine earlier cases in which a similar or parallel descriptive nature can be seen—what prior cases were about, and which way they were decided.” <i>Amdocs (Isr.) Ltd v. Openet Telecom, Inc.</i>, 841 F.3d 1288, 1294 (Fed. Cir. 2016).</p> <p>If the claims recite an abstract idea, the second step requires a “search for an ‘inventive concept’— an element or combination of elements that is ‘sufficient to ensure that the patent in practice amounts to</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
<p>call command by the preset call command recognition unit, and</p> <p>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.</p>	<p>significantly more than a patent upon the [ineligible concept] itself.” <i>Id.</i> at 218 (quoting <i>Mayo Collaborative Servs. v. Prometheus Labs., Inc.</i>, 566 U.S. 66, 73 (2016)). “A claim that recites an abstract idea must include ‘additional features’ to ensure ‘that the claim is more than a drafting effort designed to monopolize the abstract idea.’” <i>Id.</i> at 221 (quoting <i>Mayo</i>, 566 U.S. at 78) (alterations omitted). The recited “additional features must be more than well-understood, routine, conventional activity.” <i>Ultramercial, Inc. v. Hulu, LLC</i>, 772 F.3d 1335, 1339 (Fed. Cir. 2013) (internal quotations omitted).</p> <p>Recognizing a keyword based on what it, other noises, and silence sounds like, and then listening for subsequent speech after recognizing the keyword, as claimed, is an abstract idea.</p> <p>First, the claim can be performed by the human mind alone and is, therefore, directed to an abstract idea. <i>See, e.g., CyberSource Corp. v. Retail Decisions, Inc.</i>, 654 F.3d 1366, 1373 (Fed. Cir. 2011) (a “method that can be performed by human thought alone” is an abstract idea); <i>Alice</i>, 573 U.S. at 208 (“method[s] of organizing human activity” are abstract ideas); <i>PersonalWeb Techs. LLC v. Google LLC</i>, 8 F.4th 1310, 1316 (Fed. Cir. 2021) (mental processes are “a telltale sign of abstraction”). As discussed above, humans recognize a keyword (such as their name) and once recognized, listen for whatever is said next. The human mind does so based on what it knows the keyword, what other words and noises sound like, as well as what silence sounds like.</p> <p>Moreover, the Challenged Claims have no particular concrete or tangible form or application. <i>See, e.g., Ultramercial, Inc. v. Hulu, LLC</i>, 772 F.3d 709, 715 (Fed. Cir. 2014) (an idea is abstract when it has “no particular concrete or tangible form,” and is “devoid of a concrete or tangible application.”); arguments and evidence related to Unit Limitations in ’196 Patent Invalidity Contentions Cover Pleading. Instead, the Challenged Claims merely describe desired functions without providing any technical details about tangible components, further suggesting the challenged claims are directed to an abstract idea. <i>See, e.g., Affinity Labs of Texas, LLC v. Amazon.com, Inc.</i>, 838 F.3d 1266 (Fed. Cir. 2016) (finding claims that did “no more than describe a desired function or outcome, without providing any limiting detail that confines the claim to a particular solution to an identified problem[.]” abstract); <i>IPA Technologies, Inc. v. Amazon.com, Inc.</i>, 307 F. Supp. 3d 356 (D. Del. 2018) (finding claims that recited “the basic steps that would be required” to achieve a particular aim abstract); <i>Univ. of Fla. Rsch. Found., Inc. v. Gen. Elec. Co.</i>, 916 F.3d 1363, 1368 (Fed. Cir. 2019) (finding claims abstract where they “fail[ed] to provide any technical details for the tangible components, ... instead predominately describ[ing] the system and methods in purely functional terms.”) Like</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>the claims, the specification uses primarily functional terminology to describe the alleged invention. <i>See, e.g., Pebble Tide LLC v. Arlo Techs.</i>, No. CV 18-1767-LPS, 2020 WL 509183, at *1 (D. Del. Jan. 31, 2020) (“This conclusion at Step One is supported by the fact that the representative claim lacks limiting technical details. Neither of the claims, nor for that matter the specification, explain[s] how the claimed invention’s components perform their recited functions. Rather, they describe those components in purely functional terms.”). Notably, the processor recited in passing in this claim is not described or disclosed in the specification. That the preset call recognition unit and the continuous speech recognition unit somehow use an undisclosed processor in some undisclosed way, without more, is not enough for a particular concrete or tangible form.</p> <p>“In the context of software-based inventions, <i>Alice/Mayo</i> step one often turns on whether the claims focus on the specific asserted improvement in computer capabilities or, instead, on a process that qualifies as an abstract idea for which computers are invoked merely as a tool.” <i>Trinity Info Media, LLC v. Covalent, Inc.</i>, 72 F.4th 1355, 1362–63 (Fed. Cir. 2023). There are no asserted improvements in computer capabilities disclosed in the ’196 Patent. Instead, the ’196 Patent is clear that its alleged advancement over the prior art is using call commands or keywords to wake up some sort of generic device or computer (that is not described in the patent). <i>See</i> ’196 Patent at 2:43-3:20. However, as discussed above, using a keyword or a call command to draw someone’s attention to what happens next (such as a request or command) is a human activity that can be performed in the human mind. Such ideas are not patent eligible. <i>See, e.g., Affinity Labs of Texas, LLC v. DIRECTV, LLC</i>, 838 F.3d 1253, 1257 (Fed. Cir. 2016) (if the “focus of the claimed advance over the prior art” shows that “the claim’s character as a whole is directed to” steps that “can be performed in the human mind, or by a human using a pen and paper,” the claim is directed to an abstract idea) (cleaned up).</p> <p>The Challenged Claims merely describe desired functions without providing any technical details, further suggesting the challenged claims are directed to an abstract idea. <i>See, e.g., Affinity Labs of Texas, LLC v. Amazon.com, Inc.</i>, 838 F.3d 1266 (Fed. Cir. 2016) (finding claims that did “no more than describe a desired function or outcome, without providing any limiting detail that confines the claim to a particular solution to an identified problem[.]” abstract); <i>IPA Technologies, Inc. v. Amazon.com, Inc.</i>, 307 F. Supp. 3d 356 (D. Del. 2018) (finding claims that recited “the basic steps that would be required” to achieve a particular aim abstract); <i>Univ. of Fla. Rsch. Found., Inc. v. Gen. Elec. Co.</i>, 916 F.3d 1363, 1368 (Fed. Cir. 2019) (finding claims abstract where they “fail[ed] to provide any technical details for the tangible components, ... instead predominately describ[ing] the system and methods in purely functional terms.”) Like the claims, the ’196 Patent specification uses primarily functional terminology to describe the alleged invention. <i>See, e.g., Pebble</i></p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p><i>Tide LLC v. Arlo Techs.</i>, No. CV 18-1767-LPS, 2020 WL 509183, at *1 (D. Del. Jan. 31, 2020) (“This conclusion at Step One is supported by the fact that the representative claim lacks limiting technical details. Neither of the claims, nor for that matter the specification, explain[s] how the claimed invention’s components perform their recited functions. Rather, they describe those components in purely functional terms.”).</p> <p>Courts considering similar claims have routinely found them abstract (and ineligible). <i>See, e.g., Content Extraction & Transmission LLC v. Wells Fargo Bank, N.A.</i>, 776 F.3d 1343 (Fed. Cir. 2014) (“concept of data collection, recognition, and storage” is abstract idea); <i>Enco Systems, Inc. v. DaVincia, LLC</i>, 447 F.Supp.3d 916 (E.D. Mo. 2020), <i>aff’d Enco Sys., Inc. v. DaVincia, LLC</i>, 845 F. App’x 953 (Fed. Cir. 2021) (finding claims directed to employing speech-to-text technology abstract); <i>Parus Holdings, Inc. v. Sallie Mae Bank</i>, 137 F. Supp. 3d 660, 672 (D. Del. 2015), <i>aff’d</i>, 677 F. App’x 682 (Fed. Cir. 2017) (finding a “computer and telecommunications network for receiving, sending and managing information from a subscriber to the network and from the network to a subscriber” including via speech recognition abstract); <i>MModal Servs. Ltd. v. Nuance Commc’ns, Inc.</i>, No. 1:18-CV-00901, 2019 WL 13059774, at *1 (N.D. Ga. May 13, 2019) (finding claims directed to a “free-form narration of a patient encounter provided by a clinician,’ provid[ing] ‘to a user the one or more options’ corresponding to particular semantic concepts, and ‘updat[e] a textual representation of the free-form narration’ based on the correlations” abstract); <i>In re TLI Communications LLC Patent Litigation</i>, 87 F.Supp.3d 773 (E.D. Va. 2015) (finding “taking, organizing, classifying, and storing photographs” is an abstract idea where claims could be performed by human thought alone, and further finding that limiting this idea to speech does not make the concept patent eligible); <i>Quad City Pat., LLC v. Zoosk, Inc.</i>, 498 F. Supp. 3d 1178, 1185 (N.D. Cal. 2020) (“First, detecting participant ‘speech, language, emotion, social intelligence, character and characteristics’ is routinely performed by humans when conducting face-to-face transactions. More importantly, the claims do not recite any novel computer implementation of such analysis. Instead, the claim ‘describe[s] a desired function or outcome’ of recognizing participant characteristics without limiting that function to any technical solution. <i>Affinity Labs of Tex.</i>, 838 F.3d at 1269. Finally, the limitation is not directed to improving computer functionality. Quad City does not identify any technical problem that is resolved by identifying participant characteristics. Instead, the benefits of recognizing participant characteristics by a computer appear to be the same as those of recognizing participant characteristics using human beings—namely, better marketing.”); <i>W. View Rsch., LLC v. Audi AG</i>, 685 F. App’x 923 (Fed. Cir. 2017) (“receiving or collecting data queries, analyzing the data query, retrieving and processing the information constituting a response to the initial data query, and generating a visual or audio response to the initial data query,” including involving speech recognition, is an abstract idea).</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>To the extent this abstract idea is considered more than one such ideas, the combination of these abstract ideas does not render claim 9 any less abstract. <i>See Sensormatic Elecs., LLC v. Wyze Labs, Inc.</i>, 484 F. Supp. 3d 161, 165 (D. Del. 2020), <i>aff'd</i>, No. 2020-2320, 2021 WL 2944838 (Fed. Cir. July 14, 2021) (“Courts often invalidate patents that are directed to a combination of abstract ideas.”) (citing cases); <i>Broadband iTV, Inc. v. Amazon.com, Inc.</i>, No. 6:20-CV-00921-ADA, 2022 WL 4703425, at *14 (“Combining two abstract ideas . . . is not inventive.”); <i>RecogniCorp. LLC v. Nintendo Co.</i>, 855 F.3d 1322, 1327 (Fed. Cir. 2017) (“Adding one abstract idea . . . to another abstract idea . . . does not render the claim non-abstract.”); <i>Consumer 2.0, Inc. v. Tenant Turner, Inc.</i>, 343 F. Supp. 3d 581, 590-91 (E.D. Va. 2018), <i>aff'd</i> 796 F. App’x 752 (Fed. Cir. 2020) (“[T]he mere combination of these abstract processes is insufficient to bring the patent-at-issue out of the realm of the abstract.”) (citing cases).</p> <p>Claim 9 of the ’196 Patent does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “configured to receive” speech, “compare the input speech with a preset call command,” “recognize the input speech as the preset call command according to a result of the comparison,” “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,” and “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.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. <i>See, e.g., Two-Way Media Ltd. v. Comcast Cable Commc’ns, LLC</i>, 874 F.3d 1329, 1337 (Fed. Cir. 2017) (finding the use of results-based functional language failed to “sufficiently describe how to achieve these results in a non-abstract way”).</p> <p>At best, Claim 9 seeks to apply the abstract idea(s) using a generic electronic “apparatus” or structures. But “[i]nstructing one to ‘apply’ an abstract idea and reciting no more than generic computer elements performing generic computer tasks does not make an abstract idea patent-eligible.” <i>Intellectual Ventures I LLC v. Capital One Bank (USA)</i>, 792 F.3d 1363, 1368 (Fed. Cir. 2015) (citing <i>Alice</i>, 134 S. Ct. at 2359-60). The ’196 Patent does not meaningfully describe the claimed device or any of the Unit Limitations in a meaningful way. Accordingly, there is nothing in the specification to suggest that the claimed apparatus, processor, or any of the Unit Limitations are more than generic, off the rack structures well known in the art. If any of them <i>were</i> something more, the ’196 Patent would need to explain that, and it does not. Again, the ’196 Patent is clear that its alleged advancement over the prior art is using call commands or keywords to wake up a device. There is no allegation of advancement or improvements to the device or Unit Limitations in the ’196 Patent.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p data-bbox="583 175 1978 435"><i>See, e.g., Enco</i> (“nowhere does the 191 Patent suggest its purpose is to specifically improve speech-to-text processing software, and nowhere does it explain how this is to be accomplished—across myriad software iterations—at a technical level. Simply put, there are references to things like ‘autoflush counters’ and computer-determined communication relays between speech-to-text processors and encoders, but nothing at all to explain how those functional operations are implemented at an inventive level. This is insufficient to create an inventive concept that transforms the patent into something more than a patent upon the ineligible concept itself.”).”).</p> <p data-bbox="583 500 1923 571">The ’196 Patent admits that speech recognition systems for single words as well as for many words were known, and that it was known to use recognition networks to recognize speech:</p> <p data-bbox="680 646 1978 786">Speech recognition systems using the above-described various access methods can be classified into isolated word speech recognition systems that recognize isolated words according to the type of vocalization, and continuous speech recognition systems that recognize speech made by continuously speaking words.</p> <p data-bbox="680 808 1923 1058">Of these speech recognition systems, an isolated word speech recognition system is a scheme for performing recognition in the sequence of Voice Activity Detection (VAD) feature extraction→pattern comparison→recognition, and is a technology which is suitable for small computational loads or small-scale speech recognition, but is unsuitable for commercialization because unregistered words are rejected only by confidence detection with the result that errors frequently occur in confidence detection depending on noisy environments, and because a special event such as a recording button for speech recognition is required.</p> <p data-bbox="680 1081 1978 1221">FIG. 1 is a diagram briefly showing the construction of a conventional continuous speech recognition network. The continuous speech recognition network employs a scheme that performs recognition by performing post-processing on a recognized word sequence using a language model. This scheme is mainly used in large-capacity speech recognition systems of ten thousand or more words.</p> <p data-bbox="680 1243 1978 1458">A continuous speech recognition system is a scheme for performing recognition in the sequence of feature extraction→pattern comparison (acoustic model comparison)→language model comparison→post-processing→recognition, is suitable for large-scale speech recognition systems in high-specification server level Personal Computers (PCs), and also enables a real-time large-scale speech recognition system to be implemented with the development of storage media and computational processing abilities. However, there are disadvantages because, in order to install such</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>a continuous speech recognition system in terminals which have gradually become lightweight, still as of yet insufficient storage capacity and complicated floating computation cause problems, and a large computational load and storage capacity for a large amount of data are required.</p> <p>'196 Patent at 1:44-2:16; <i>see also</i> Fig. 1; 6:4-9.</p> <p>This specification makes clear that the devices disclosed in the '196 Patent are generic, off-the-rack structures well-known in the art:</p> <p style="padding-left: 40px;">Therefore, 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.</p> <p>'196 Patent at 6:10-13.</p> <p>Moreover, this claim would preempt the broad subject matter of using an apparatus to recognize a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword. This further suggests that the claims are directed to an abstract idea and patent-ineligible subject matter. <i>See, e.g., Bancorp Servs., LLC v. Sun Life Assur. Co. of Canada (U.S.)</i>, 687 F.3d 1266, 1280 (Fed. Cir. 2012) (finding that claimed abstract idea would improperly preempt broad subject matter).</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants' Invalidity Contentions Cover Pleading, provided concurrently herewith, at each "List B" associated with a claim limitation included in this claim of the '196 Patent, and as charted in the corresponding claim charts submitted with Defendants' Invalidity Contentions.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>10. The apparatus of claim 9, wherein the preset call command recognition unit further comprises:</p> <p>an adaptive filter configured to receive the input speech and the subsequent input speech; and</p> <p>a confidence detection unit configured to compute a confidence rate of the assumed input speech and recognize the assumed input speech as the preset call command based on the computed confidence rate.</p>	<p>Like claim 9, this claim is directed to the abstract idea of recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword. <i>See supra</i> claim 9. Claim 10 adds the additional abstract ideas of (1) receiving speech, and (2) determining how confident we are the keyword has been spoken.</p> <p>Like claim 9, the subject matter of claim 10 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves receiving the spoken word and determining how confident we are that spoken word is one we know and can recognize. Also like claim 9, neither receiving speech nor determining confidence is an improvement in computer capabilities, and claim 10 merely describes desired functions without providing any technical details.</p> <p>Claim 10 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “computing ... a confidence rate of the input speech” and “recognizing ... the input speech as the preset call command based on the computed confidence rate.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. For example, the ’196 Patent does not explain how a confidence score is computed, nor does it explain how it is used to recognize input speech.</p> <p>At best, claim 10 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>11. The apparatus of claim 10, 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.</p>	<p>Like claims 9 and 10, this claim is directed to the abstract idea(s) of (1) recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword, (2) receiving speech, and (3) determining how confident we are the keyword has been spoken. Claim 11 adds the additional abstract idea of recognizing the keyword based on a certain threshold level of confidence.</p> <p>Like claims 9 and 10, the subject matter of claim 11 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves determining how confident we are that spoken word is one we know and can recognize. At a certain point (or threshold), we become confident enough that a word has been spoken and consider it done. Also like claims 9 and 10, a threshold value for confidence is not an improvement in computer capabilities, and claim 11 merely describes desired functions without providing any technical details.</p> <p>Claim 11 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “configured to recognize the assumed input speech as the preset call command if the computed confidence rate is higher than a threshold value.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. For example, the ’196 Patent does not meaningfully describe this threshold value, does not explain how to determine this threshold value, and does not provide any examples of this threshold value.</p> <p>At best, claim 11 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>12. The apparatus of claim 10, 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.</p>	<p>Like claims 9 and 10, this claim is directed to the abstract idea(s) of (1) recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword, (2) receiving speech, and (3) determining how confident we are the keyword has been spoken. Claim 12 adds the additional abstract idea of recognizing the keyword if we think it is authentic.</p> <p>Like claims 9 and 10, the subject matter of claim 12 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves determining whether we think the speech is authentic. Also like claims 9 and 10, the authenticity of speech is not an improvement in computer capabilities, and claim 12 merely describes desired functions without providing any technical details.</p> <p>Claim 12 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “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.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. For example, the ’196 Patent does not meaningfully describe what is or is not an authentic call command, nor does it describe how to recognize speech if it is authentic.</p> <p>At best, claim 12 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>13. The apparatus of claim 10, 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.</p>	<p>Like claims 9 and 10, this claim is directed to the abstract ideas of (1) recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword, (2) receiving speech, and (3) determining how confident we are the keyword has been spoken. Claim 13 adds the additional abstract idea of recognizing the keyword based on how likely it is the keyword has been spoken.</p> <p>Like claims 9 and 10, the subject matter of claim 13 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves determining how confident we are that spoken word is one we know and can recognize. In other words, we are determining how likely it is that a particular word has been spoken. This is what humans do in normal conversation. Also like claims 9 and 10, the likelihood of a keyword is not an improvement in computer capabilities, and claim 13 merely describes desired functions without providing any technical details.</p> <p>Claim 13 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “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.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. For example, the ’196 Patent does not provide any description or explanation of the models used to calculate the claimed likelihood ratios.</p> <p>At best, claim 13 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>14. The apparatus of 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.</p>	<p>Like claims 9, 10, and 13, this claim is directed to the abstract ideas of (1) recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword, (2) receiving speech, (3) determining how confident we are the keyword has been spoken, and (4) recognizing the keyword based on how likely it is the keyword has been spoken. <i>See supra</i> claims 9, 10, 13. Claim 14 adds the additional abstract idea of using the keyword to determine how likely it is the keyword has been spoken.</p> <p>Like claims 9, 10, and 13, the subject matter of claim 14 can be performed in the human mind. Using what we know a word sounds like (a model) is exactly what a human uses to determine the probability that the word has been spoken. Also like claims 9, 10, and 13, using the word or a model of the word to generate a probability that the word has been spoken is not an improvement in computer capabilities, and claim 14 merely describes desired functions without providing any technical details.</p> <p>At best, claim 14 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>15. The apparatus of claim 14, wherein said another log likelihood ratio is computed by</p>	<p>Like claims 9, 10, 13, and 14, this claim is directed to the abstract idea(s) of (1) recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword, (2) receiving speech, (3) determining how confident we are the keyword has been spoken, (4)</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
<p>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.</p>	<p>recognizing the keyword based on how likely it is the keyword has been spoken, and (5) using the keyword to determine how likely it is the keyword has been spoken. <i>See supra</i> claims 9, 10, 13, 14. Claim 15 adds the additional abstract idea of using sound other than the word to determine how likely it is that the keyword has been spoken.</p> <p>Like claims 9, 10, 13, and 14, the subject matter of claim 15 can be performed in the human mind. Using what we know a word sounds like (a model) and what other sounds sound like (an anti model) is exactly what a human uses to determine the probability or likelihood that the word has been spoken. Also like claims 9, 10, 13, and 14, using an anti model of the word to generate a probability that the word has been spoken is not an improvement in computer capabilities, and claim 15 merely describes desired functions without providing any technical details.</p> <p>At best, claim 15 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>16. The apparatus of claim 9, wherein the continuous speech recognition unit comprises:</p> <p>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</p>	<p>Like claim 9, this claim is directed to the abstract idea of recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword. <i>See supra</i> claim 9. Claim 16 adds the additional abstract idea of recognizing and understanding speech after the keyword.</p> <p>Like claim 9, the subject matter of claim 16 can be performed in the human mind. Recognizing and understanding speech after the keyword is no different than recognizing and understanding the keyword. Also like claim 9, recognizing and understanding speech is not an improvement in computer capabilities, and claim 16 merely describes desired functions without providing any technical details.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
<p>a rule check unit configured to determine whether the recognized subsequent input speech corresponds to a target sentence.</p>	<p>Claim 16 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “configured to recognize the subsequent input speech as the actual command if the input speech is recognized as the preset call command” and “configured to determine whether the recognized subsequent input speech corresponds to a target sentence.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved.</p> <p>At best, claim 16 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>17. A method for recognizing a speech of a speaker by a single device, the method comprising:</p> <p>receiving, by a preset call command recognition unit of the single device, an input speech spoken by the speaker;</p> <p>receiving, by the preset call command recognition unit, a subsequent input speech spoken,</p>	<p>All of the Challenged Claims are directed to the abstract idea of recognizing a keyword and then listening for speech after the spoken keyword. This is what humans do every day. For example, when our name is spoken (e.g., “Andrew”), we hear it, and we then listen for what is spoken next, such as an instruction or a request (e.g., “Andrew, will you please take out the trash?”). The challenged claims are also directed to the abstract idea of recognizing the spoken keyword using (a) what we know the keyword sounds like, (b) what we know other words and noise sound like, and (c) what we know silence sounds like. Again, this is what humans do in normal, everyday conversation. For example, we know when “Andrew” is spoken because it comports with what we know “Andrew” sounds like, what other words like “Megan” and “John sound like, what we know extraneous noise like laughing sounds like, and what we know silence sounds like.</p> <p>35 U.S.C. § 101 lists four broad categories of patent-eligible subject matter, but is subject to “an important implicit exception” that “[l]aws of nature, natural phenomena, and abstract ideas,” which form the “basic</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
<p>by the speaker, subsequently to the input speech;</p> <p>comparing, by the preset call command recognition unit, the input speech with a preset call command;</p> <p>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; and 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</p> <p>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</p>	<p>tools of scientific and technological work,” are not eligible for patent protection. <i>Alice Corp. Pty. Ltd. v. CLS Bank Int’l</i>, 573 U.S. 208, 216 (2014) (internal quotation marks omitted). “The ‘abstract ideas’ category embodies ‘the longstanding rule that an idea of itself is not patentable.’” <i>Id.</i> at 209 (quoting <i>Gottschalk v. Benson</i>, 409 U.S. 63, 67 (1972) (quoting <i>Rubber-Tip Pencil Co. v. Howard</i>, 87 U.S. 498 (1874))).</p> <p><i>Alice</i> set forth a two-step inquiry. “First, we determine whether the claims at issue are directed to one of those patent-ineligible concepts.” <i>Id.</i> at 217. In making this inquiry, the claims “must be considered as a whole.” <i>Alice</i>, 573 U.S. at 218 n.3 (quoting <i>Diamond v. Diehr</i>, 450 U.S. 175, 188 (1981)). And courts consider “what the patent asserts to be the focus of the claimed advance over the prior art.” <i>Solutran, Inc. v. Elavon, Inc.</i>, 931 F.3d 1161, 1168 (Fed. Cir. 2019) (cleaned up). “Instead of a definition [for abstract idea] ... courts ... examine earlier cases in which a similar or parallel descriptive nature can be seen—what prior cases were about, and which way they were decided.” <i>Amdocs (Isr.) Ltd v. Openet Telecom, Inc.</i>, 841 F.3d 1288, 1294 (Fed. Cir. 2016).</p> <p>If the claims recite an abstract idea, the second step requires a “search for an ‘inventive concept’— an element or combination of elements that is ‘sufficient to ensure that the patent in practice amounts to significantly more than a patent upon the [ineligible concept] itself.’” <i>Id.</i> at 218 (quoting <i>Mayo Collaborative Servs. v. Prometheus Labs., Inc.</i>, 566 U.S. 66, 73 (2016)). “A claim that recites an abstract idea must include ‘additional features’ to ensure ‘that the claim is more than a drafting effort designed to monopolize the abstract idea.’” <i>Id.</i> at 221 (quoting <i>Mayo</i>, 566 U.S. at 78) (alterations omitted). The recited “additional features must be more than well-understood, routine, conventional activity.” <i>Ultramercial, Inc. v. Hulu, LLC</i>, 772 F.3d 1335, 1339 (Fed. Cir. 2013) (internal quotations omitted).</p> <p>Recognizing a keyword based on what it, other noises, and silence sounds like, and then listening for subsequent speech after recognizing the keyword, as claimed, is an abstract idea.</p> <p>First, the claim can be performed by the human mind alone and is, therefore, directed to an abstract idea. <i>See, e.g., CyberSource Corp. v. Retail Decisions, Inc.</i>, 654 F.3d 1366, 1373 (Fed. Cir. 2011) (a “method that can be performed by human thought alone” is an abstract idea); <i>Alice Corp. Pty. v. CLS Bank Int’l</i>, 573 U.S. 208 (2014). (“method[s] of organizing human activity” are abstract ideas); <i>PersonalWeb Techs. LLC v. Google LLC</i>, 8 F.4th 1310, 1316 (Fed. Cir. 2021) (mental processes are “a telltale sign of abstraction”). As discussed</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
<p>accompanied by noise and the preset call command.</p>	<p>above, humans recognize a keyword (such as their name) and once recognized, listen for whatever is said next. The human mind does so based on what it knows the keyword, what other words and noises, as well as what silence sounds like.</p> <p>“In the context of software-based inventions, <i>Alice/Mayo</i> step one often turns on whether the claims focus on the specific asserted improvement in computer capabilities or, instead, on a process that qualifies as an abstract idea for which computers are invoked merely as a tool.” <i>Trinity Info Media, LLC v. Covalent, Inc.</i>, 72 F.4th 1355, 1362–63 (Fed. Cir. 2023). There are no asserted improvements in computer capabilities disclosed in the ’196 Patent. Instead, the ’196 Patent is clear that its alleged advancement over the prior art is using call commands or keywords to wake up some sort of generic device or computer (that is not described in the patent). <i>See</i> ’196 Patent at 2:43-3:20. However, as discussed above, using a keyword or a call command to draw someone’s attention to what happens next (such as a request or command) is a human activity that can be performed in the human mind. Such ideas are not patent eligible. <i>See, e.g., Affinity Labs of Texas, LLC v. DIRECTV, LLC</i>, 838 F.3d 1253, 1257 (Fed. Cir. 2016) (if the “focus of the claimed advance over the prior art” shows that “the claim’s character as a whole is directed to” steps that “can be performed in the human mind, or by a human using a pen and paper,” the claim is directed to an abstract idea) (cleaned up).</p> <p>The Challenged Claims merely describe desired functions without providing any technical details, further suggesting the challenged claims are directed to an abstract idea. <i>See, e.g., Affinity Labs of Texas, LLC v. Amazon.com, Inc.</i>, 838 F.3d 1266 (Fed. Cir. 2016) (finding claims that did “no more than describe a desired function or outcome, without providing any limiting detail that confines the claim to a particular solution to an identified problem[.]” abstract); <i>IPA Technologies, Inc. v. Amazon.com, Inc.</i>, 307 F. Supp. 3d 356 (D. Del. 2018) (finding claims that recited “the basic steps that would be required” to achieve a particular aim abstract); <i>Univ. of Fla. Rsch. Found., Inc. v. Gen. Elec. Co.</i>, 916 F.3d 1363, 1368 (Fed. Cir. 2019) (finding claims abstract where they “fail[ed] to provide any technical details for the tangible components, ... instead predominately describ[ing] the system and methods in purely functional terms.”) Like the claims, the ’196 Patent specification uses primarily functional terminology to describe the alleged invention. <i>See, e.g., Pebble Tide LLC v. Arlo Techs.</i>, No. CV 18-1767-LPS, 2020 WL 509183, at *1 (D. Del. Jan. 31, 2020) (“This conclusion at Step One is supported by the fact that the representative claim lacks limiting technical details. Neither of the claims, nor for that matter the specification, explain[s] how the claimed invention’s components perform their recited functions. Rather, they describe those components in purely functional terms.”).</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Courts considering similar claims have routinely found them abstract (and ineligible). <i>See, e.g., Content Extraction & Transmission LLC v. Wells Fargo Bank, N.A.</i>, 776 F.3d 1343 (Fed. Cir. 2014) (“concept of data collection, recognition, and storage” is abstract idea); <i>Enco Systems, Inc. v. DaVincia, LLC</i>, 447 F.Supp.3d 916 (E.D. Mo. 2020), <i>aff’d Enco Sys., Inc. v. DaVincia, LLC</i>, 845 F. App’x 953 (Fed. Cir. 2021) (finding claims directed to employing speech-to-text technology abstract); <i>Parus Holdings, Inc. v. Sallie Mae Bank</i>, 137 F. Supp. 3d 660, 672 (D. Del. 2015), <i>aff’d</i>, 677 F. App’x 682 (Fed. Cir. 2017) (finding a “computer and telecommunications network for receiving, sending and managing information from a subscriber to the network and from the network to a subscriber” including via speech recognition abstract); <i>MModal Servs. Ltd. v. Nuance Commc’ns, Inc.</i>, No. 1:18-CV-00901, 2019 WL 13059774, at *1 (N.D. Ga. May 13, 2019) (finding claims directed to a “free-form narration of a patient encounter provided by a clinician,’ provid[ing] ‘to a user the one or more options’ corresponding to particular semantic concepts, and ‘updat[e] a textual representation of the free-form narration’ based on the correlations” abstract); <i>In re TLI Communications LLC Patent Litigation</i>, 87 F.Supp.3d 773 (E.D. Va. 2015) (finding “taking, organizing, classifying, and storing photographs” is an abstract idea where claims could be performed by human thought alone, and further finding that limiting this idea to speech does not make the concept patent eligible); <i>Quad City Pat., LLC v. Zoosk, Inc.</i>, 498 F. Supp. 3d 1178, 1185 (N.D. Cal. 2020) (“First, detecting participant ‘speech, language, emotion, social intelligence, character and characteristics’ is routinely performed by humans when conducting face-to-face transactions. More importantly, the claims do not recite any novel computer implementation of such analysis. Instead, the claim ‘describe[s] a desired function or outcome’ of recognizing participant characteristics without limiting that function to any technical solution. <i>Affinity Labs of Tex.</i>, 838 F.3d at 1269. Finally, the limitation is not directed to improving computer functionality. Quad City does not identify any technical problem that is resolved by identifying participant characteristics. Instead, the benefits of recognizing participant characteristics by a computer appear to be the same as those of recognizing participant characteristics using human beings—namely, better marketing.”); <i>W. View Rsch., LLC v. Audi AG</i>, 685 F. App’x 923 (Fed. Cir. 2017) (“receiving or collecting data queries, analyzing the data query, retrieving and processing the information constituting a response to the initial data query, and generating a visual or audio response to the initial data query,” including involving speech recognition, is an abstract idea).</p> <p>To the extent this abstract idea is considered more than one such ideas, the combination of these abstract ideas does not render Claim 1 any less abstract. <i>See Sensormatic Elecs., LLC v. Wyze Labs, Inc.</i>, 484 F. Supp. 3d 161, 165 (D. Del. 2020), <i>aff’d</i>, No. 2020-2320, 2021 WL 2944838 (Fed. Cir. July 14, 2021) (“Courts often invalidate patents that are directed to a combination of abstract ideas.”) (citing cases); <i>Broadband iTV, Inc. v. Amazon.com, Inc.</i>, No. 6:20-CV-00921-ADA, 2022 WL 4703425, at *14 (“Combing two abstract ideas . . . is not inventive.”); <i>RecogniCorp. LLC v. Nintendo Co.</i>, 855 F.3d 1322, 1327 (Fed. Cir. 2017) (“Adding one</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>abstract idea . . . to another abstract idea . . . does not render the claim non-abstract.”); <i>Consumer 2.0, Inc. v. Tenant Turner, Inc.</i>, 343 F. Supp. 3d 581, 590-91 (E.D. Va. 2018), <i>aff’d</i> 796 F. App’x 752 (Fed. Cir. 2020) (“[T]he mere combination of these abstract processes is insufficient to bring the patent-at-issue out of the realm of the abstract.”) (citing cases).</p> <p>Claim 17 of the ’196 Patent does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “comparing . . . the input speech with a preset call command,” “determining . . . whether the input speech is recognized as the preset call command according to a result of the comparison,” and “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.” But the ’196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved. <i>See, e.g., Two-Way Media Ltd. v. Comcast Cable Commc’ns, LLC</i>, 874 F.3d 1329, 1337 (Fed. Cir. 2017) (finding the use of results-based functional language failed to “sufficiently describe how to achieve these results in a non-abstract way”).</p> <p>At best, Claim 17 seeks to apply the abstract idea(s) using a generic electronic “single device” or structures. But “[i]nstructing one to ‘apply’ an abstract idea and reciting no more than generic computer elements performing generic computer tasks does not make an abstract idea patent-eligible.” <i>Intellectual Ventures I LLC v. Capital One Bank (USA)</i>, 792 F.3d 1363, 1368 (Fed. Cir. 2015) (citing <i>Alice</i>, 134 S. Ct. at 2359-60). The ’196 Patent does not meaningfully describe the claimed device or any of the Unit Limitations in a meaningful way. Accordingly, there is nothing in the specification to suggest that the claimed device or any of the Unit Limitations are more than generic, off the rack structures well known in the art. If any of them <i>were</i> something more, the ’196 Patent would need to explain that, and it does not. Again, the ’196 Patent is clear that its alleged advancement over the prior art is using call commands or keywords to wake up a device. There is no allegation of advancement or improvements to the device or Unit Limitations in the ’196 Patent. <i>See, e.g., Enco</i> (“nowhere does the 191 Patent suggest its purpose is to specifically improve speech-to-text processing software, and nowhere does it explain how this is to be accomplished—across myriad software iterations—at a technical level. Simply put, there are references to things like ‘autoflush counters’ and computer-determined communication relays between speech-to-text processors and encoders, but nothing at all to explain how those functional operations are implemented at an inventive level. This is insufficient to create an inventive concept that transforms the patent into something more than a patent upon the ineligible concept itself.”).”).</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p data-bbox="583 245 1921 315">The '196 Patent admits that speech recognition systems for single words as well as for many words were known, and that it was known to use recognition networks to recognize speech:</p> <p data-bbox="680 386 1976 529">Speech recognition systems using the above-described various access methods can be classified into isolated word speech recognition systems that recognize isolated words according to the type of vocalization, and continuous speech recognition systems that recognize speech made by continuously speaking words.</p> <p data-bbox="680 548 1921 802">Of these speech recognition systems, an isolated word speech recognition system is a scheme for performing recognition in the sequence of Voice Activity Detection (VAD) feature extraction→pattern comparison→recognition, and is a technology which is suitable for small computational loads or small-scale speech recognition, but is unsuitable for commercialization because unregistered words are rejected only by confidence detection with the result that errors frequently occur in confidence detection depending on noisy environments, and because a special event such as a recording button for speech recognition is required.</p> <p data-bbox="680 821 1976 964">FIG. 1 is a diagram briefly showing the construction of a conventional continuous speech recognition network. The continuous speech recognition network employs a scheme that performs recognition by performing post-processing on a recognized word sequence using a language model. This scheme is mainly used in large-capacity speech recognition systems of ten thousand or more words.</p> <p data-bbox="680 984 1986 1308">A continuous speech recognition system is a scheme for performing recognition in the sequence of feature extraction→pattern comparison (acoustic model comparison)→language model comparison→post-processing→recognition, is suitable for large-scale speech recognition systems in high-specification server level Personal Computers (PCs), and also enables a real-time large-scale speech recognition system to be implemented with the development of storage media and computational processing abilities. However, there are disadvantages because, in order to install such a continuous speech recognition system in terminals which have gradually become lightweight, still as of yet insufficient storage capacity and complicated floating computation cause problems, and a large computational load and storage capacity for a large amount of data are required.</p> <p data-bbox="583 1382 1184 1414">'196 Patent at 1:44-2:16; <i>see also</i> Fig. 1; 6:4-9.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>This specification makes clear that the devices disclosed in the '196 Patent are generic, off-the-rack structures well-known in the art:</p> <p style="padding-left: 40px;">Therefore, 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.</p> <p>'196 Patent at 6:10-13.</p> <p>Moreover, this claim would preempt the broad subject matter of using a device to recognize a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword. This further suggests that the claims are directed to an abstract idea and patent-ineligible subject matter. <i>See, e.g., Bancorp Servs., LLC v. Sun Life Assur. Co. of Canada (U.S.)</i>, 687 F.3d 1266, 1280 (Fed. Cir. 2012) (finding that claimed abstract idea would improperly preempt broad subject matter).</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants' Invalidity Contentions Cover Pleading, provided concurrently herewith, at each "List B" associated with a claim limitation included in this claim of the '196 Patent, and as charted in the corresponding claim charts submitted with Defendants' Invalidity Contentions.</p> <p>Thus, either "individually" or "as an ordered combination," the elements of this claim do not add enough to "transform the nature of the claim" into a patent-eligible application." <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>18. The method of claim 17, wherein the single device is a portable device or an embedded system.</p>	<p>Like claim 17, this claim is directed to the abstract idea of recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword. <i>See supra</i> claim 17. Claim 18 only adds a well known and generic portable device or embedded system.</p> <p>Like claim 17, the subject matter of claim 18 can be performed in the human mind. Also like claim 17, claim 18 does not involve an improvement in computer capabilities.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>Claim 18 does not contain an inventive concept such that it does more than claim ineligible ideas. There is nothing in the '196 Patent to suggest the claimed a portable device or embedded system is something more than a well-known, off the shelf structure.</p> <p>At best, claim 18 seeks to apply the abstract idea(s) using a generic electronic “single device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the '196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim’ into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>19. The method of claim 17, wherein the single device, wherein determining whether the input speech is recognized as the preset call command further comprises:</p> <p>computing, by a confidence detection unit, a confidence rate of the input speech; and</p> <p>recognizing, by the confidence detection unit, the input speech as the preset call command based on the computed confidence rate.</p>	<p>Like claim 17, this claim is directed to the abstract idea of recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword. <i>See supra</i> claim 17. Claim 19 adds the additional abstract idea of recognizing the keyword based on how confident we are it has been spoken.</p> <p>Like claim 17, the subject matter of claim 19 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves determining how confident we are that spoken word is one we know and can recognize. Also like claim 17, determining confidence is not an improvement in computer capabilities, and claim 19 merely describes desired functions without providing any technical details.</p> <p>Claim 19 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “computing ... a confidence rate of the input speech” and “recognizing ... the input speech as the preset call command based on the computed confidence rate.” But the '196 Patent does not provide any meaningful explanation of how those</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p>functional steps are to be achieved. For example, the '196 Patent does not explain how a confidence score is computed, nor does it explain how it is used to recognize input speech.</p> <p>At best, claim 19 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p>The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the '196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p>Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>
<p>20. The method of claim 17, wherein determining whether the input speech is recognized as the preset call command further comprises:</p> <p>computing a log likelihood ratio of the input speech; and</p> <p>assuming the input speech to be the preset call command based on the computed log likelihood ratio.</p>	<p>Like claim 17, this claim is directed to the abstract idea of recognizing a keyword based on what we know it and other sounds and silence sound like, and then listening for speech after the spoken keyword. <i>See supra</i> claim 17. Claim 20 adds the additional abstract idea of recognizing the keyword based on how likely it is the keyword has been spoken.</p> <p>Like claim 17, the subject matter of claim 20 can be performed in the human mind. Deciding whether or not we recognize a spoken word involves determining how confident we are that spoken word is one we know and can recognize. In other words, we are determining how likely it is that a particular word has been spoken. Also like claim 17, the likelihood of a keyword is not an improvement in computer capabilities, and claim 20 merely describes desired functions without providing any technical details.</p> <p>Claim 20 does not contain an inventive concept such that it does more than claim ineligible ideas. Instead, it recites only broad and result-based aspirational and functional language such as “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.” But the '196 Patent does not provide any meaningful explanation of how those functional steps are to be achieved.</p>

Challenged Claims	Exception to Eligibility, and Basis Therefor
	<p data-bbox="583 245 1866 280">At best, claim 20 seeks to apply the abstract idea(s) using a generic electronic “device” or structures.</p> <p data-bbox="583 354 1992 532">The functions and generic structures recited in this claim were all well-known, routine, and conventional in the art, as shown in greater detail in this chart and the chart below and as further shown in each of the references listed in Defendants’ Invalidity Contentions Cover Pleading, provided concurrently herewith, at each “List B” associated with a claim limitation included in this claim of the ’196 Patent, and as charted in the corresponding claim charts submitted with Defendants’ Invalidity Contentions.</p> <p data-bbox="583 605 1955 711">Thus, either “individually” or “as an ordered combination,” the elements of this claim do not add enough to “transform the nature of the claim” into a patent-eligible application.” <i>Alice</i>, 134 S. Ct. at 2355 (quoting <i>Mayo</i>, 132 S. Ct. at 1297-98).</p>

Pursuant to the Court’s Standing Order Regarding Subject Matter Eligibility Contentions, Defendants provide the following chart identifying a description of the industry at the relevant time in which the Challenged Claims are alleged to be well understood, routine, and conventional, and the factual and legal basis therefore; and a description of how each element of each Challenged Claim, both individually and in combination with other elements of that claim, was well understood, routine, and conventional, in the relevant industry at the relevant time, and the legal and factual basis therefor.

Description of Industry, at Relevant Time, in Which the Claim is Well-Understood, Routine, and Conventional, and Basis Therefor
<p>The Challenged Claims cover system and method equivalents of listening for a keyword and upon recognizing that keyword, listening for what is said next. This is what humans do every day. For example, when our name is spoken (e.g., “Andrew”), we hear it, and we then listen for what is spoken next, such as an instruction or a request (e.g., “Andrew, will you please take out the trash?”). It is “a broad and familiar concept [in the realm of organizing human activity] that is untethered to any specific or concrete way of implementing it.” <i>Affinity Labs of Texas, LLC v. DIRECTV, LLC</i>, 838 F.3d 1253, 1258 (Fed. Cir. 2016).</p> <p>Humans have been using speech for thousands of years. Likewise, humans have used names for thousands of years. <i>See</i> https://prolingo.com/blog/-comparative-diversity-where-did-our-names-come-from/; https://www.britannica.com/topic/name/Family-names. Names are used to identify or alert a specific person (or other thing, like a dog). It is well known that humans use names to signal attention to whatever is coming next (“John, will you please pass me the sugar?”). Similarly, humans use names to get the attention of or wake up their pet companions (“Timber, let’s go!”).</p> <p>As discussed below, using computers to recognize speech was a logical progression of these basic human activities, and it was well-known well before KAIFI’s May 28, 2008 claimed priority date for the ’196 Patent, and earlier alleged invention date of January 2007. <i>See, e.g.</i>, Schmandt at xiii (in 1994, “As we enter the next millennium, speech unquestionably will become the primary means of communication between human beings and machines.”). Speech recognition systems were well developed by these dates.</p> <p>Also as discussed below, using a name or other particular word to wake up a device or speech recognition system was a logical progression of these basic human activities. This was also well-known well before KAIFI’s claimed priority and earlier invention dates for the ’196 Patent. In fact, this functionality existed in numerous commercial products by these dates. Perhaps most prevalently, Microsoft Speech Recognition used a wake word to wake up its speech recognition features, and was distributed along with each Microsoft XP and Vista operating system as well as freely downloadable from the Microsoft’s website years before KAIFI’s claimed priority and prior invention dates.</p> <p>Speech recognition had been studied in earnest for at least several decades before KAIFI’s claimed priority and prior invention dates. <i>See, e.g.</i> Young I (dated 1989, citing numerous speech recognition studies and papers from 1970s and 1980s); Rabiner I (dated 1985); Paul (dated 1990, “Automatic speech recognition has a long history of being a difficult problem-the first papers date from about 1950”). Several textbooks directed to speech recognition and the use of computers in speech recognition were published in the 1990s. <i>See</i> Schmandt (1994), Rabiner III at 6</p>

(1993, “Research in automatic speech recognition by machine has been done for almost four decades.”); *see also id.* at 6-9 (discussing the history of speech recognition).

The ’196 Patent discusses the goal of speech recognition, but does not describe how this goal is accomplished. Regardless, there were numerous ways to recognize speech by KAIFI’s claimed priority and prior invention dates.

As Schmandt explained in 1994, there are three basic components of any speech recognizer: (1) a speech representation, or the form into which the recognizer converts speech signals, (2) a set of templates or models “against which an input can be compared to determine what was spoken[,]” and (3) “[a] pattern matching algorithm to determine which template is most similar to ... speech input.” Schmandt at 132-33. Furthermore, “[t]o identify a word the recognizer must capture incoming speech and convert it to the chosen internal representation The pattern matcher then selects the template that most closely matches the input or rejects the utterance if no template is a close enough match.” *Id.* at 133.

By 1994, speaker-independent recognizers with small vocabularies (10-20 words) could recognize “single words spoken in isolation.” *Id.* at 138. For example, by 1994, “[s]everal commercially available speaker-independent recognizers accept the digits ‘0’ through ‘9’ plus ‘yes’ and ‘no.’” *Id.*; *see also id.* at 139 (“A keyword spotting recognizer can locate a few words in the midst of any amount of speech.”). In 1994, advanced speech recognizers used either dynamic time warping (which is based on dynamic programming) or hidden Markov models. *See id.* at 141-42.

Hidden Markov models were used by almost all large continuous speech recognition systems in the years just before KAIFI’s claimed priority and prior invention dates and, therefore, will be primarily discussed herein. *See* Gales at Abstract (“Hidden Markov Models (HMMs) provide a simple and effective framework for modelling time-varying spectral vector sequences. As a consequence, almost all present day large vocabulary continuous speech recognition (LVCSR) systems are based on HMMs.”); Lee III at 10 (“Hidden Markov modeling (HMM) is a powerful technique capable of robust modeling of speech.”), *id.* at 17 (“It was only in the past few years, however, that HMMs became the predominant approach to speech recognition, superseding dynamic time warping.”). As evidenced below, HMMs were well-understood, routine, and conventional at the relevant time. It was also well-understood, routine, and conventional to use HMMs to recognize single (or isolated) words or to recognize many words.

“The basic theory of hidden Markov models was published in a series of classic papers by Baum and his colleagues ([1]-[5]) in the late 1960s and early 1970s and was implemented for speech processing applications by Baker [6] at CMU, and by Jelinek and his colleagues at IBM ([7]-[13]) in the 1970s.” Rabiner III at 322; *see also* Lee III at 17 (“Hidden Markov models (HMM) were first described in the classic paper by Baum [Baum 72]. Shortly afterwards, they were extended to automatic speech recognition independently at CMU [Baker 75a] and IBM [Bakis 76, Jelinek 76].”).

As Rabiner I explains, in 1985:

In the literature a wide variety of approaches have been proposed to recognize isolated words, based on standard statistical-pattern recognition techniques. The most successful of these has been the template-based recognizer approach, which uses Dynamic Programming (DP) as the method for comparing patterns. Although the template-based approach using DP has been very successful, alternative recognition strategies have been studied because of:

1. The high computational cost of the DP approach;
2. The difficulties in extending the DP recognition paradigm to more difficult problems-e.g., connected words, continuous speech;
3. The desire to use a robust parametric model, rather than the nonparametric template, to represent the speech;
4. The desire to use speech units other than words in some circumstances-e.g., syllables, demisyllables, phonemes.

Rabiner I at 1211-12. Rabiner I explains that due to these above reasons, different approaches to speech recognition had been proposed including using Hidden Markov Models, commonly referred to in the literature as HMMs. *See id.* at 1212. By 1985, “[w]ork performed at IBM, CMU, and more recently at Phillips has used continuous HMMs” *Id.*

In the Rabiner I paper, the authors (employees at AT&T) “extend previous work on isolated-word recognition based on hidden Markov models by replacing the discrete symbol representation of the speech signal with a continuous Gaussian mixture density.” *Id.* at 1211. The authors explain in detail the HMM recognizer they built as well as the procedures for training the HMM. *See id.* at 1212-1222. The authors ran several experiments to determine proper parameters for the HMM recognizer. *See id.* at 1222-1228. The authors also discuss computational consideration for the HMM recognizer. *Id.* at 1230-1231. The authors conclude that “We have tested the HMM recognizer, in a speaker independent mode, on a vocabulary of the ten digits, and shown that the error rate of the system is smaller than that obtained from the discrete density (VQ-based) HMM recognizer, and comparable to that of the multiple template-based DTW recognizer.” *Id.* at 1232.

In 1990, Paul observed that:

Automatic speech recognition has a long history of being a difficult problem-the first papers date from about 1950. During this period, a number of techniques, such as linear-time-scaled word-template matching, dynamic-time-warped word-template matching, linguistically motivated approaches (find the phonemes, assemble into words, assemble into sentences), and hidden Markov models (HMM), were used. Of all of the available techniques, HMMs are currently yielding the best performance.

Paul at 41. Paul notes that “A.A. Markov first used Markov models to model letter sequences in Russian [.]” *Id.* at 42 (citing A.A. Markov article from 1913). Like Rabiner I, Paul explains the HMM theory, model and training in detail. *See id.* at 41-49.

Normally we speak in continuous sentences of words called sentences. In continuous speech, there is considerable variation due to the sentence structure (prosodics) and interaction between adjacent words (cross-word coarticulation). In addition, no clear acoustic markers delineate the word boundaries. Therefore, it is simpler to recognize isolated words. The word pronunciations are much more uniform and the boundaries are much easier to locate than in continuous speech. Therefore, isolated-word recognition will be described first.

Id. at 49-50. Paul goes on to describe a HMM system for single isolated words, as well as one for continuous speech recognition (*see id.* at 50-59). He concludes that “[g]reat strides in speech recognition have been made by using hidden Markov models.” *Id.* at 59.

See also Schmandt at 144-47 (in 1994, discussing HMMs and their use in speech recognition systems under development), 151-52 (describing details of three advanced speech recognizers that use HMMs); Rabiner III, chapter 6 (Theory and Implementation of Hidden Markov Models).

Around the same time, researchers were considering topologies and computational cost/resources for HMMs. For example, in 1989, Young published a paper describing “Token Passing: a Simple Conceptual Model for Connected Speech Recognition Systems[.]” Young I at 1. Young I teaches a token passing model for “isolated word recognition.” Young I explains that “each vocabulary word is represented by a word model which is a finite state network,” as shown in Figure 1 below. Young I, 2-3.

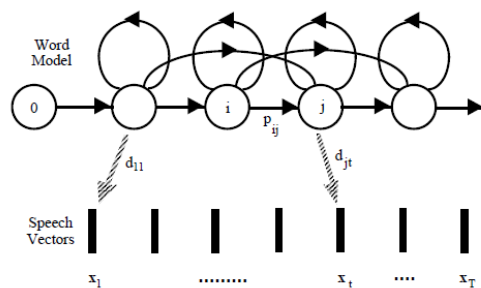


Figure 1: Structure of a Word Model

Young I provides a method for computing similarities between a word model and the unknown speech. Specifically, Equation 3 shown below defines a value for how similar the word model is to the unknown speech, which is inversely proportional to the minimum cost alignment shown in Equation 2.

Each possible sequence of states through the model $\mathbf{i} = i_0, i_1, \dots, i_T$, where i_0 is the initial state, represents one possible alignment of the model with the speech signal. The total cost $S(\mathbf{i})$ of this alignment is given by

$$S(\mathbf{i}) = \sum_{\tau=1}^T (p_{i_{\tau-1}, i_{\tau}} + d_{i_{\tau}}(\tau)) \quad (1)$$

In speech recognition, the *similarity* between a model and the unknown speech is assumed to be inversely proportional to the minimum cost alignment. Let $s_j(t)$ be the minimum cost alignment between the segment $\mathbf{x}_1 \dots \mathbf{x}_t$ of the unknown speech and the model starting in state i_0 and ending in state j . By the principle of optimality, this cost can be computed by the recursion

$$s_j(t) = \min_i \{s_i(t-1) + p_{ij}\} + d_j(t) \quad (2)$$

thereby giving an efficient way of computing the required minimum value of $S(\mathbf{i})$, that is

$$S_{min}(\mathbf{i}) = \min_j \{s_j(T)\} \quad (3)$$

where j is any allowable final state of the word model.

With respect to the equations described above, Young I explains that “all of the above is entirely straightforward and well-known.” Young I, Sec. 2.1, p. 2.

Young I further provides an alternative algorithm for evaluating the minimum cost alignment in Equation 2 above using token passing:

Let each state of a word model be capable of holding a moveable *token* where each token, for the moment at least, holds a single alignment cost so that at time t , the token in state j holds the value $s_j(t)$. In terms of these tokens, the algorithm implied by equation 2 can be reformulated (somewhat informally) as follows:

Initialisation:

Each model initial state holds a token with value 0;
All other states hold a token with value ∞

Algorithm:

```
for t:= 1 to T do
  for each state  $i$  do
    Pass a copy of the token in state  $i$  to all connecting
    states  $j$ , incrementing its  $s$  value by  $p_{ij} + d_j(t)$ ;
  end;
  Discard the original tokens;
  for each state  $i$  do
    find the token in state  $i$  with the smallest  $s$ 
    value and discard the rest
  end;
end;
```

Termination:

Examine all final states, the token with the smallest s value
gives the required minimum matching score.

Young I explains that the use of token passing “leads to much simpler and more powerful generalization.” Young I, 3. As shown in the algorithm above and Figure 3 below, at a given state, a token captures information regarding the path by which the token arrived at that state including a value s , corresponding to how similar the unknown speech is to that specific path in the word model. If there are multiple paths for arriving at a state, then each path will have a token with a different s value, and the algorithm specifies that the token with the smallest s value is kept while the rest are discarded. In other words, token passing requires that the paths that are less likely to correspond to the unknown speech are discarded, thereby minimizing the beam search space and simplifying the process as word recognition continues.

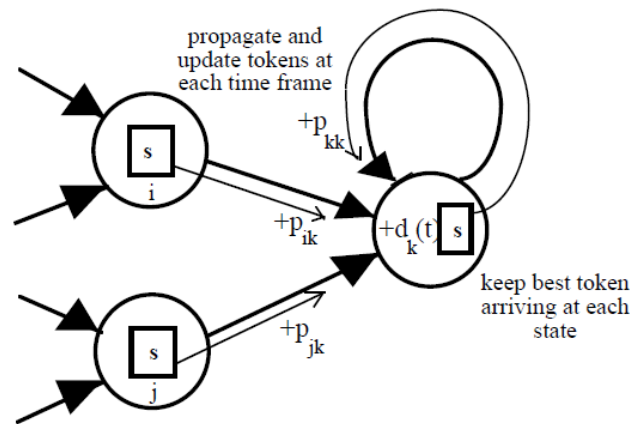


Figure 3: Basic Token Passing Algorithm

Young goes on to provide an equation for determining the similarity between a word model and unknown speech in the context of hidden Markov model (HMM)-based pattern matching:

As noted above, the correspondence between the token passing model and hidden Markov models is obvious and direct. For a HMM with transition probabilities a_{ij} and output probabilities $b_j(\mathbf{x})$, if each transition cost p_{ij} is set equal to $-\log a_{ij}$ and each local cost $d_j(t)$ is set equal to $-\log b_j(\mathbf{x})$ then equation 2 becomes

$$s_j(t) = \max_i \{s_i(t-1) + \log a_{ij}\} + \log b_j(\mathbf{x}) \quad (4)$$

and $s_j(t)$ now denotes the maximum log probability of the model being in state j after generating the sequence $\mathbf{x}_1 \dots \mathbf{x}_T$. Equation 4 is the standard Viterbi decoder equation for a HMM [Levinson *et al* 1983].

As explained in Young I, “the correspondence between the token passing model and hidden Markov models is obvious and direct,” such that a POSITA would understand that token passing is also used in the context of HMMs as an alternative algorithm for evaluating Equation 4, similar to the token passing algorithm described with respect to Equation 2 above. Young I, 4. *See also* Young I, 2 (“Indeed, the similarity between this model and a hidden Markov model is particularly self-evident. However, the distinction is maintained here purposefully to emphasize that token passing is not restricted to HMM’s.”)

Numerous other prior references evidence that token passing for HMMs was well-understood, routine, and conventional as of KAIFI's claimed priority date. For example, Konopka '188 (published 2008, priority to 1999) teaches that "token passing is a well-known approach to tracking the best word hypotheses through an HMM." Konopka '188, [0052]. Konopka '188 explains that this "token passing technique provides a highly memory efficient solution to an otherwise 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." Konopka '188, [0052].

As another example, Hogenhout '750 (published 2005) discusses various token passing approaches known in the art, and proposes variations on those methods. *See generally* Hogenhout '750. Like the other references discussed herein, Hogenhout '750 discusses concerns with computational expense, resources, and efficiency, and proposes ways to use token passing to decrease these burdens. *See, e.g., id.* at 19-20.

Lahti '711 is additional evidence that token passing with HMMs was well-understood, routine, and conventional at the relevant time. Lahti '711 teaches that "[m]ost speech recognition systems use Viterbi search algorithm which builds a search through a network of Hidden Markov Models (HMMs) and maintains most likely path score at each state in this network for each frame or time step." Lahti '711 at [0002]. In Lahti '711's speech recognition system:

Token passing is used to transfer score information between states. Each state of a HMM (at time frame t) holds a token comprising information on partial log probability. A token represents partial match between observation sequence (up to time t) and the model. A token passing algorithm propagates and updates tokens at each time frame and passes the best token (having the highest probability at time t-1) to next state (at time t). At each time frame, the log probability of a token is accumulated by corresponding transition probabilities and emission probabilities. The best token scores are thus found by examining all possible tokens and selecting the ones having the best scores. As each token is passing through a search tree (network), it maintains a history recording its route. For more details on token passing and token scores, reference is made to "Token passing: a Simple Conceptual model for Connected Speech Recognition Systems", Young, Russell, Thornton, Cambridge University Engineering Department, Jul. 31, 1989, which is incorporated herein as reference.

Id. at [0027].

Catchpole '214 (published 2006, priority to 2004) is also additional evidence that token passing with HMMs was well-understood, routine, and conventional at the relevant time.

This speech recognition technology was so advanced at the relevant time there was popular and readily available portable toolkit for building and manipulating hidden Markov models called the Hidden Markov Model Toolkit (HTK). 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/>.

As set forth in detail in the HTK claim chart attached to Defendants concurrently-served Invalidity Contentions, the Hidden Markov Model Toolkit used token passing.

Prior art from KoreaPowervoice (the '196 Patent's original assignee) also shows that speech recognition was well understood, routine, and conventional at the relevant time. For example, Hee-Souk '469 (published February 2006) teaches:

The preprocessing step extracts and processes a feature parameter for increasing the degree of discrimination for the speaker in the system using the voice information of each speaker input through the input device controller in the main processor 30. For example, pitch, formant, and LPCC (Linear Predictive Cepstral Coefficient) based on the speaker's vocalization characteristics and vocal tract model characteristics, etc. Coefficient) and PLP (Perceptual Linear Predictive) are mainly used.

Then, the normalization and grouping of the parameters is performed, which algorithms grouping any number of input features into a certain number of groups so that an infinite number of input patterns can be mapped to a finite representative value of a finite number of finite numbers. K-means algorithm, LBG algorithm, and the like.

In the recognition step, a pattern classification algorithm is mainly used. The VQ (Vector Quantization) method for measuring the distance between the pre-configured codebook patterns and the distance through the quantization operation of the input pattern and the DTW (Dynamic Time) using the time matching technique Warning) and NN (Neural Network) algorithms that show relatively good performance in nonlinear classification by reflecting the characteristics of human neural structure, and HMM (Hidden Markov) that analyzes temporal dynamic characteristics as stochastic static characteristics by probability modeling technique. The algorithm is mainly used.

The decision rule step includes a cohort model technique that compares the average similarity of the most probable speakers with similarity with the corresponding speaker through an arbitrary user model and determines whether the user is determined by a predetermined threshold. The world model method that analyzes the characteristics of many speakers as one common model in advance and compares and determines the similarity of the speakers is mainly used.

In the present invention, a mixed model Viterbi Searching algorithm is designed in order to increase discrimination between speakers and minimize false recognition rate using a HMM-based speaker identification system having a relatively high recognition rate. Through the non-deterministic algorithm, the recognition performance is maximized.

The mixed model Viterbi search algorithm is performed by estimating a maximum probability score for arbitrary input data through a model parameter of a speaker generated in a learning process, and a non-built speaker for the decision rule step. The identity of the person is determined by comparing the probability estimates with respect to the model (ie, the world model, the cohort model, etc.).

Hee-Souk '469 (english translation) at 6. None of the above is described of being new/novel in Hee-Souk '469. And, notably, Hee-Souk '469 provides significantly more detail as to how it performs speech recognition than the '196 Patent.

As another example from the original assignee, and in comparison to the '196 Patent, Hee-souk '695 alleges improvements to speech recognition technology. Hee-souk '695's disclosure is very detailed with technical explanations of its speech recognition system, including dozens of equations. This is more of the type of disclosure one would expect to see in a patent application that has allegedly improved speech recognition in the 2000s (after decades of earlier work).

As discussed in the chart above, the '196 Patent's purported advance over the prior art is the idea of using a call command or wake word to wake up some sort of generic device or computer (that is not described in the patent). *See* '196 Patent at 2:43-3:20. As evidenced by published prior art, several products that were available on the market prior to the 196 Patent's priority date, and even pop culture, this idea was also well-understood, routine, and conventional at the relevant time.

As discussed above, humans have used keywords such as a name to wake up or alert pets and humans for many years. As such, this concept appears throughout pop culture. *See, e.g.,* https://www.reddit.com/r/comics/comments/cnkor/what_we_say_to_dogs_and_what_they_hear_thanks/?rdt=46770 (Gary Larson cartoon regarding dog names):



Using this concept to wake up or alert an electronic device/computer is thus no surprise.

For example, in Star Trek, users of the Enterprise's computer used the wake word/call command "computer" to wake up the computer and then initiate tasks. See, e.g., <https://www.youtube.com/watch?v=IPLiaYD1IUo> ("Computer, make this a metal table"); see also <https://www.cnet.com/home/smart-home/amazon-echos-new-wake-word-is-a-gift-to-trekkies/>; <https://www.youtube.com/watch?v=DmHZ6dR4HhA>; <https://www.youtube.com/watch?v=x74qzr0z3B0>. This is exactly what the '196 Patent inventors purport to have invented in 2009.

As another example, in Night Rider, Michael Knight speaks to his car KITT, often saying KITT before asking the car to do something. <https://www.youtube.com/watch?v=KRItKmTH6Hw> ("KITT, just keep driving"). Again, this is exactly what the '196 Patent inventors purport to have invented in 2009; see also Rabiner III at I (discussing science fiction wonders).

Like in the pop culture, the idea of using a wake word/keyword/onword/ or call command to wake up or alert a computing device is prevalent in the prior art.

As Schmandt explained in 1994:

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 Ramachandran explained in 1996:

■ *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. Likewise, Rudnický described using wake/call/attention words in 1990:

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).

Rudnick at 14-15. And 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; *see generally* Manos (discussing many keyword spotting applications and systems).

As another example, Kepuska I (2006) presents a solution for recognizing a wake-up-word using dynamic time warping. As Kepuska I explains:

The presented technology provides a cutting-edge solution for human-machine interaction through the utilization of a properly solved Wake-Up- Word (WUW) SR problem. This paradigm-shift provides the basis for development of SR systems with truly “Voice Activated” capabilities, impacting all SR based technologies and the way in which humans interact with computers. This shift is a radical

departure from the current “push-to-talk” paradigm currently applied to all speech-to-text or speech-recognition applications. To be able to achieve this goal, a significantly more accurate pattern classification and scoring technique is required, which in turn provides SR systems enhanced performance for correct recognition (i.e., minimization of false rejection) as well as correct rejection (i.e., minimization of false acceptance). A revolutionary and innovative classification and scoring technique is used that is a significant enhancement over an earlier method presented in reference [1]. The solution in reference [1] has been demonstrated to meet the stringent requirements of the WUW-SR task. Advanced solution of [1] is a novel technique that is model and algorithm independent. Therefore, it could be used to significantly improve performance of existing recognition algorithms and systems. Reduction of error rates of over 40% are commonly observed for both false rejections and false acceptance. In this paper the architecture of the WUW-SR based system as interface to current SR applications is presented. In this system WUW-SR is used as a gateway for truly Voice Activated applications utilizing the current solution without “push-to-talk” paradigm. The technique has been developed with hardware optimization in mind and therefore has the ability to run as a “background” application on a standard Windows-based PC platform.

Kepuska I at 1.

Bou-Ghazale (2000) also teaches this idea. Bou-Ghazale is “designed to...recognize a pre-defined activation command.” Bou-Ghazale, Abstract. “Upon recognizing that the activation command has been spoken, the voice activation system turns the device on or switches to the main speech recognition system to receive further commands from the user.” Bou-Ghazale, 1735.

As another example, Konopka '188 also teaches the idea of using a wake word/keyword/onword/call command to wake up or alert a computing device. In Konopka '188:

A search is performed in a network for an attention word based on a first context including a first set of models, grammars, and lexica. Upon finding the attention word, a Switch is performed to a second context to search for an open-ended user request. The second context includes a second set of models, grammars, and lexicons and the open-ended user request does not follow a predetermined format.

Konopka '188 at Abstract.

Additionally, Cohen '776, Tran '140, Qu '786, and Millman '413 all each disclose the idea of using a wake word/keyword/onword/call command to wake up or alert a computing device. They provide additional evidence of why this claimed feature was well-understood, routine, and conventional.

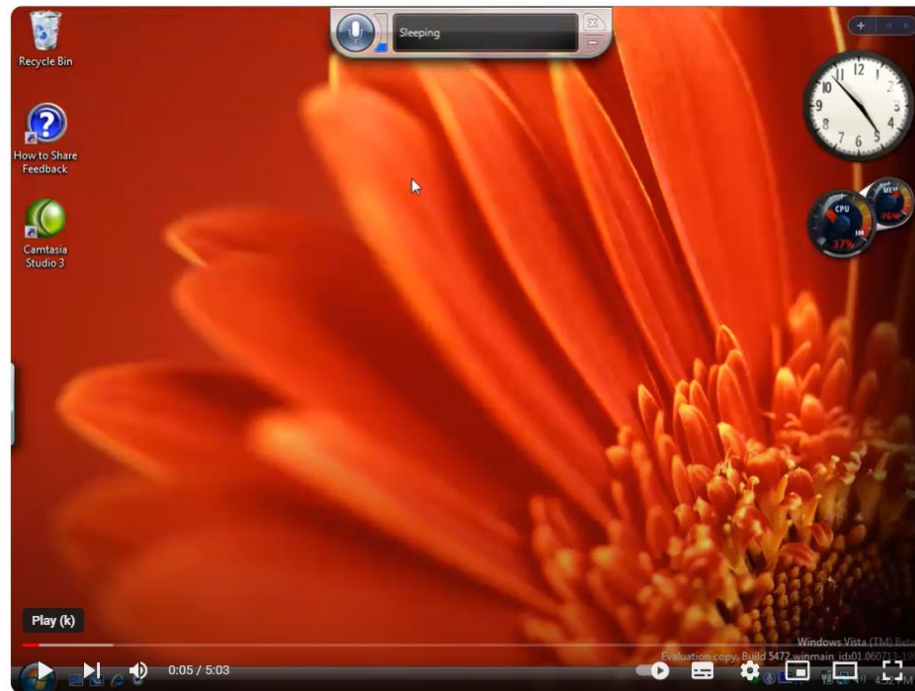
The product prior art also shows why this feature was well understood, routine, and conventional. As set forth in detail in the claim charts attached to Defendants concurrently-served Invalidity Contentions, there were at least three products available on the market at the relevant time that used a wake word/keyword/onword/call command to wake up or alert a computing device: Microsoft Speech Recognition, IBM's ViaVoice, and Dragon's NaturallySpeaking. Upon recognition of the particular wake word, each of these products performed continuous speech recognition.

For example, in 2007, the New York Times featured an article explaining how a user could use a wake word to turn on and turn off Microsoft speech recognition:

* “Start listening”/”Stop listening.” These commands tell your PC to start and stop listening to you. That’s important, because you don’t want it to interpret everything you say. It would not be so great if it tried to act when you said to your roommate, “Hey, Chris, close the window.”

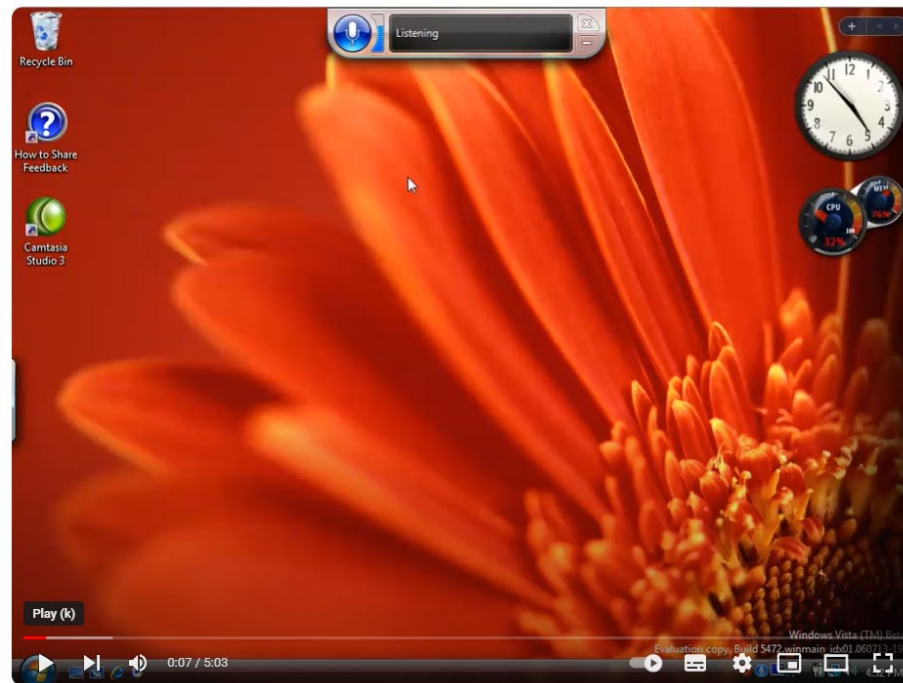
<https://archive.nytimes.com/pogue.blogs.nytimes.com/2007/03/01/telling-your-computer-what-to-do/>; *see also*
<https://www.youtube.com/watch?v=kX8oYoYy2Gc> (showing computer receiving “start listening” input speech from a speaker);
<https://www.youtube.com/watch?v=bU9we3TdrUA> (same).

To illustrate, before the wake word/preset call command (e.g., “start listening”) is spoken, Microsoft Speech Recognition is in a sleep state:



<https://www.youtube.com/watch?v=bU9we3TdrUA>.

After the wake word/preset call command (e.g., “start listening”) is spoken, Microsoft Speech Recognition switches to a listening mode:



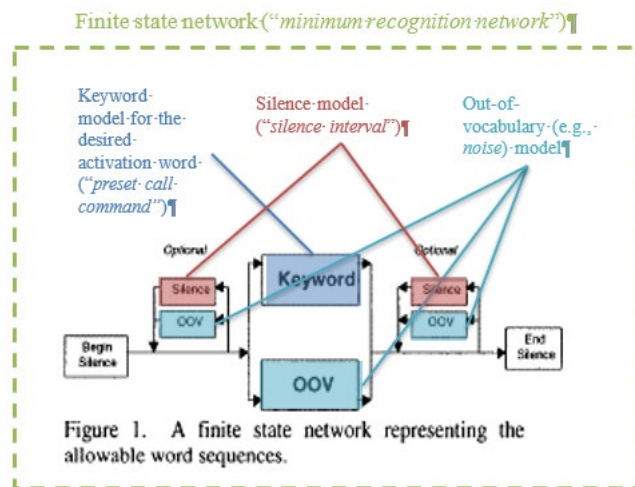
<https://www.youtube.com/watch?v=bU9we3TdrUA>. In that listening mode, Microsoft Speech Recognition listens for commands from the user and then performs those commands.

The other product prior art discussed herein offered similar features.

Regarding the claimed “minimum recognition network,” this limitation renders the claims indefinite as discussed in the concurrently served Invalidity Contentions Cover Pleading. To the extent this language can be understood, it seems to mean a network that looks for only the call command. *See* ’196 Patent, 6:60-63. Numerous examples of recognizing a single word or command are discussed above as well as in the claim charts attached to the Invalidity Contentions, and therefore this limitation is also routine, conventional and well-known.

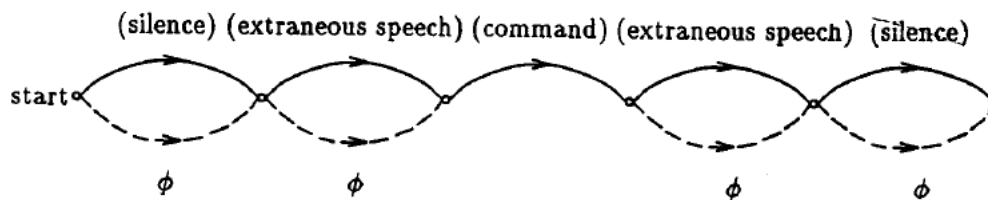
Networks composed of a silence interval accompanied by noise and the preset call command were also well-known, routine, and conventional at the relevant time. For example, Bou-Ghazale teaches that its speech recognition system (“token passing unit”) uses “a minimum recognition network composed of a silence interval accompanied by noise and the preset call command.” Specifically, the speech recognition system uses a “finite state network” that includes a keyword model for the desired activation word (“preset call command”), a model for out-of-vocabulary words (“noise”), and a model for silence (“silence interval”): “The finite state network consists of optional silence models at the beginning and

end, and a parallel network of the activation keyword and OOV model,” as shown in Figure 1 below. Bou-Ghazale, Sec. 2.2, 1736; see also id. (“the recognition system uses a finite state network, as shown in Figure 1, to represent the allowable sequences of silence, keyword, and out-of-vocabulary words”) (listing examples of allowable word sequences, including “[Silence Keyword OOV Silence]” and “[Silence OOV Keyword OOV Silence]”).



Rabiner III also teaches that networks of silence, noise, and a particular keyword or command were well known and conventional. See Rabiner III at 490:

For such cases it was necessary to use some type of keyword spotting technique to find the command words embedded within the sentence. A very powerful keyword spotter was devised [3] based on the concept of training an “extraneous speech” pattern from the spoken input with extraneous speech, and on recognizing the input using connected word recognition techniques (as discussed in Chapter 7) based on a sentence grammar of the form:



The confidence and likelihood elements of the claims were also well known, routine, and conventional. For example, Bou-Ghazale teaches that the voice activation system (“preset call command recognition unit”) comprises a post-processing component. Bou-Ghazale, Sec. 2, 1735. The post-processing component renders obvious the claimed “confidence detection unit” because it verifies whether the activation word has been found in the input speech by determining (i.e., detecting) a confidence rate for the input speech, called the “relative log-likelihood score.” Bou-Ghazale, Abstract, 1735, Sec. 2.3-3.1, 1736. As explained in detail in the claim charts attached to the Invalidity Contentions, at least Konopka ’188, Lee I, Lee II, and Lam ’664 show that the claimed confidence and likelihood elements were well known, routine, and conventional.

Finally, the claimed rule check and keyword spotting units (which are not explained or disclosed in the ’196 patent) appear to simply describe speech recognition as it was well known in the prior art as discussed in all the speech recognition prior art identified above and in the Defendants’ Invalidity Contentions. Recognizing a target sentence and searching for keywords is exactly what speech recognition systems do.

In sum, at the relevant time, the speech recognition industry was quite developed, having existed for decades. Numerous commercial products that incorporate the alleged novelty of the ’196 Patent (using a word to wake up a computer or speech recognition system) were being sold and distributed years before KAIFI’s alleged priority date.

In addition to the discussion above, below is additional analysis and discussion of how each Challenged Claim, both individually and in combination with the other elements of that claim, was well understood, routine, and conventional.

Regarding *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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for the preamble functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Every single one of the references or systems discussed above discloses the preamble functionality. As just a few examples, Microsoft Speech Recognition, Bou-Ghazale, and Konopka ’188 all disclose methods and apparatuses for recognizing speech of a speaker, such as a user.

Regarding *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”), this functionality was well understood, routine, and conventional for all

the reasons shown in Defendants' invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to "List B" for the receiving input speech functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

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.

When using Microsoft Speech Recognition, a user/speaker's computer "receiv[es], by a preset call command recognition unit, an input speech spoken by the speaker." Microsoft Speech Recognition includes a "a preset call command unit," for example, that initially searches only for the command "start listening."

Bou-Ghazale teaches that "[t]he voice activation system continuously listens to and processes incoming speech[,]" including speech spoken by "the user." Thus, "receiving...an input speech spoken by the speaker" was well understood, routine, and conventional. Further, Bou-Ghazale's "voice activation system" that "continuously listens to and processes incoming speech" is the claimed preset call command recognition unit because it is "designed to...recognize a pre-defined activation command." Thus, "receiving, by a preset call command recognition unit, an input speech spoken by the speaker" was well understood, routine, and conventional.

Regarding *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"), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants' invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to "List B" for the receiving subsequent speech functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Both Microsoft Speech Recognition and Bou-Ghazale receive subsequent input speech in the same way they receive the input speech, as discussed above.

Additionally, Rabiner III explains a general model for speech recognition that "begins with a user creating a speech signal (speaking) to accomplish a given task. The spoken output is first recognized in that the speech signal is decoded into a series of words that are meaningful according to the syntax, semantics and pragmatics of the recognition task." Rabiner III at 3; *see also id.* at Fig. 1-1, chapter 3 (discussing processing received speech). Thus, speech recognition systems must receive speech in order to be in a position to recognize it.

Regarding *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"), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants' invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to "List B" for the comparing functionality show that it was well understood, routine, and conventional

at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

First, this is what speech recognition systems do. All of the speech recognition systems discussed above and in the Invalidity Contentions compare sound that is received to sound that is known, for example, using a network of Markov models.

Second, as shown above, Microsoft Speech Recognition is in a sleep state before the wake word/preset call command (e.g., “start listening”) is spoken. After the wake word/preset call command (e.g., “start listening”) is spoken, Microsoft Speech Recognition switches to a listening mode. Between the sleep and listening modes, Microsoft Speech Recognition compares the input speech with the preset call command.

See also Rabiner III at chapter 4 (discussing comparison techniques).

Regarding 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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for the result of the comparison functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

First, this is what speech recognition systems do. All of the speech recognition systems discussed above and in the Invalidity Contentions compare sound that is received to sound that is known, for example, using a network of Markov models. Based on the result of that comparison, the speech recognition systems determine whether the sound is recognized (or not).

Second, as shown above, Microsoft Speech Recognition is in a sleep state before the wake word/preset call command (e.g., “start listening”) is spoken. After the wake word/preset call command (e.g., “start listening”) is spoken, Microsoft Speech Recognition switches to a listening mode. Between the sleep and listening modes, Microsoft Speech Recognition compares the input speech with the preset call command, and determines that the wake word has been spoken based on the result of that comparison.

Regarding 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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for activating functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

This is the core idea of wake words. Once a system (like a speech recognition system) recognizes a wake word, the system wakes up or turns on.

As one example, and as discussed above, Konopka ’188 is evidence that this feature is well known, routine, and conventional. In Konopka ’188:

A search is performed in a network for an attention word based on a first context including a first set of models, grammars, and lexica. Upon finding the attention word, a Switch is performed to a second context to search for an open-ended user request. The second context includes a second set of models, grammars, and lexicons and the open-ended user request does not follow a predetermined format.

Konopka ’188 at Abstract.

Microsoft Speech Recognition and Bou-Ghazale also show this limitation was well known, routine, and conventional, as discussed above.

Regarding 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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for token passing functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

As Konopka ’188 explains:

Token Passing is a well-known approach to tracking the best word hypotheses through an HMM. As is known in the art, in connected word recognition systems, the last model state for the state sequence with the highest Viterbi score can be easily found once the processing of all frames of an utterance is completed. However, this does not necessarily provide the best state (or word) sequence. To find the best state sequence, it is required to perform “back tracing”. The traditional way of doing this is to let each state contain a pointer back to the previously best state for each frame. Back tracing can then be performed by following the pointers back, starting with the last model state for the state sequence with the highest Viterbi score. This means that if a system uses N states over T discrete time steps, the number of back pointers required is typically NT . This quickly becomes a high number and therefore leads to high memory requirements. Various methods have been proposed to minimize the memory requirements associated with storing such back-pointers, whereof some are based on the idea of passing “tokens” around to the various states instead of allocating memory on a per-state basis. In accordance with one embodiment of the invention, instead of storing one token pointer in each state, the speech decoder 216 uses two arrays S_1 and S_2 to hold the token pointers for each state. Array S_1 keeps the token pointers for each state and the previous frame, and S_2 keeps the token pointers for each state and the current frame. When each state i “looks back” to find the previously best state j , two things can happen. If the previous best state j is a member of the same acoustic model as i , the token pointer for state j in S_1 is copied into position i in S_2 . If this is not the case, a new token is created and stored in position i in S_2 . The new token gets the same contents as token i in S_1 , and in the token history, a reference to model m , $i \in m$, is added. Once all states have been processed for the current frame, the pointers to structures S_1 and S_2 , are swapped, and the process is repeated for the following frame. Thus, this token passing technique provides a highly memory efficient solution to an otherwise 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.

Konopka '188 at [0052]. Thus, Konopka '188's speech recognition module includes a token passing unit that uses token passing with its HMMs and Konopka '188's speech recognition module (which “functions as a Hidden-Markov Model (HMM)-based continuous speech recognizer that has the ability to reject “unmodeled events”, e.g. out-of vocabulary events, disfluencies, environmental noise, etc[.]” *id.* at [0039]) is a token passing unit as claimed.

Konopka '188 further teaches 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. For example, Konopka '188 teaches that (i) the recognition system utilizes a finite state network for recognizing speech based on HMMs that consists of models for the keyword, for out-of-vocabulary words (e.g., extraneous speech and background noises), and silence, (ii) and the recognition system “assumes” the activation word is included in the input speech according to the result string with the highest likelihood score.

For example, Konopka '188 teaches:

- “Next, generally, the speech recognition module 204 functions as a Hidden-Markov Model (HMM)-based continuous speech recognizer that has the ability to reject “unmodeled events”, e.g. out-of vocabulary events, disfluencies, environmental noise, etc. The speech

recognition module 204 is under the control of the natural language interface module 222 and can switch between different acoustic models and different grammars based on the context of the speech, as determined by the natural language interface control module 206. The speech recognition module 204 may be entirely conventional, although the speech recognition module 204 has several features which are advantageous for use in the NLICS 102. Furthermore, memory usage in the speech recognition module 204 has been optimized so that the memory requirement is mainly a reflection of the amount of acoustic speech model data used. A more detailed description follows of the speech recognition module 204 and the natural language interface control module 206.” *Id.* at [0039].

- “Additionally, two garbage-modeling techniques are utilized. Garbage filler models are stored with the acoustic models module 220 to model background noises as well as disfluencies and “silences”. These models are utilized by the speech decoder 216 in the rejection of out-of-vocabulary (oov) events. The speech decoder 216 also rejects out-of-vocabulary (oov) events using an online garbage calculation. It then returns the N-best candidates if their scores are very close. Such out-of-vocabulary rejection is also well understood in the art.” *Id.* at [0039].
- “The models used by the speech decoder 216 are stored in the acoustic models module 220. These models may comprise context-dependent or independent phonetic models, sub word models or whole word models, e.g. monophones, biphones and/or triphones. In one embodiment, the speech decoder 216 may dynamically switch between different models, e.g., the speech decoder 216 may switch between models based on triphones and monophones. This is in contrast to known systems, where there are a fixed number of states and Gaussians per state, i.e. the architecture of the respective phonemes is fixed. In contrast, a selection between models based on monophones, biphones, and triphones, as well as varying the architecture of these phonemes, e.g., the number of states and the number of Gaussians per state for each type of phoneme (monophone, biphone, and triphone) may be varied for optimization in space, speed, and accuracy. As is well understood in the art, the received utterances are analyzed with the models, e.g., using a Viterbi algorithm, and scores are assigned representing how well the utterance fits the given models. Furthermore, the models used by the speech decoder 216 are under direct control by the natural language interface control module 206, which is described further below.” *Id.* at [0041].
- “Additionally, two garbage-modeling techniques are utilized. Garbage filler models are stored with the acoustic models module 220 to model background noises as well as disfluencies and “silences”. These models are utilized by the speech decoder 216 in the rejection of out-of-vocabulary (oov) events. The speech decoder 216 also rejects out-of-vocabulary (oov) events using an online garbage calculation. It then returns the N-best candidates if their scores are very close. Such out-of-vocabulary rejection is also well understood in the art.” *Id.* at [0042].
- “In some embodiments, the rejection techniques have been improved compared to those known in the art. The basic principle behind HMM-based speech recognition systems is that an utterance is compared with a number of speech models (from the acoustic models module 220) in order to find the model that best matches the utterance. This implies that the output of the speech recognition module 204 will be a reference to the model (e.g. word) with the best match. However, this causes problems in cases where no models exist that represent the words spoken. In such cases, an HMM-based system will typically still attempt to find the closest match between utterances and models and report the results. In many cases this is unwanted, as any sound that is picked up by an open microphone will cause a reference to a model to be emitted. To avoid this effect, it is sometimes preferred to determine whether the utterance is contained within in-vocabulary words or not. For example, if a Viterbi score passes a threshold, the utterance is determined to be an in-vocabulary word. If the Viterbi score of the utterance does not exceed the threshold, then the utterance is deemed out-of vocabulary. Such a Viterbi score is generated using the Viterbi algorithm. This algorithm calculates a single best state sequence through an HMM and its corresponding

probability, given an observation sequence. However, experiments have shown that this is not a very accurate rejection scheme. Instead, many systems rely on comparing the Viterbi score with another Viterbi score that is obtained by reprocessing the utterance through an alternative HMM whose task is to represent all out-of-vocabulary events or filler sounds, i.e., using garbage models. A garbage score can then be defined as the difference between the logarithms of each of the two Viterbi scores divided by the number of frames in the utterance according to equation 1 below. The garbage score reveals whether the utterance had a closer match with the word models or the out-of-vocabulary models. Many variants have been proposed as to how to reject out-of-vocabulary events. One observation is that periods of silence in an utterance typically produce high Viterbi scores even for models that are supposed to model high-energy parts-of-speech. To some extent this can be avoided by providing an additional feature representing the energy of the speech signal in the feature extraction module 202. However, this still leads to incorrect garbage score measurements. If there is silence in the beginning or ending of an utterance and this beginning or ending silence is not being modeled, it has been observed that the garbage scores are indeed affected. The feature extraction module 202 performs speech detection such that the beginning and ending silences should not be included in the sample forwarded to the speech decoder 216 of the speech recognition module 204. However, finding the beginning and ending of an utterance becomes a complex task for utterances that begin or end with low-energy sounds. An example of a group of sounds where this is a problem is the fricative. Fricatives are characterized as broadband, low energy noise, e.g. “white noise”. A fricative, as known in the art, is a sound, as exemplified by such phonemes as “th”, “sh”, etc. The feature extraction module 202 attempts to solve this problem by making its best efforts to find the beginning and ending samples To guarantee that low-energy sounds are included in the speech sample, the feature extraction module 202 includes a number of extra samples in the beginning and ending of the utterance. In cases where there is no low-energy sound in the beginning or ending of an utterance, this implies that silence will be prepended and appended to the speech sample, assuming that the utterance was spoken in isolation, and hence the garbage scores in the speech decoder 216 become skewed. To solve this problem, in one embodiment, each model is preceded and followed by a single-state silence model that “consumes” the frames of silence passed along from the feature extraction module 202. The speech decoder 216 then finds the sequence of models with the closest match and optimally aligns the silence models as well as the word-models with the utterance. Now the start and end indices for the beginning and ending silence portions of the utterance can be obtained and removed. Furthermore, the best matching word models are now kept and reprocessed without the preceding and following silence models, using only the pure-speech portion of the utterance. Next, the out-of vocabulary HMMs process the same portion of the utterance and the garbage scores can be calculated as,

$$s = \frac{w - g}{n - m}, \quad (1)$$

where w is the logarithm of the Viterbi score for the acoustic models of in-vocabulary words without preceding or following silence models and where no silence is included in the utterance. Similarly, g is the logarithm of the corresponding score for the out-of-vocabulary HMM models. Also, n is the total number of frames in the utterance and m is the number of frames that were consumed by the preceding and following silence models. In summary, using this rejection technique, the system is better able to accurately isolate the speech portion of the utterance. This has the effect of better isolating in-vocabulary words and rejecting out-of vocabulary events that begin or end with low energy sounds, such as fricatives, in comparison to conventional rejection schemes.” *Id.* at [0043]-[0044].

- “Next, the speech recognition module 204 applies acoustic hidden Markov models (HMM) and an N-gram grammar to the incoming feature vectors (as specified by the natural language interface) to derive an in-vocabulary (IV) Viterbi (likelihood) score (Step 408). Then, the feature data is reprocessed using models of OOV events, e.g., an ergodic bank of monophone models, to derive an out-of-vocabulary (OOV) Viterbi score (Step 410). The garbage score is calculated from the IV and OOV scores, e.g., the garbage score equals $[\text{Ln}(\text{IV score}) - \text{Ln}(\text{OOV score})] / \text{number of frames}$ (Block 411). A low score indicates a garbage utterance. The N-best transcribed text string(s) and corresponding garbage score(s) are passed to the natural language interface module 222 (Step 412). The natural language interface module 222 parses the incoming string(s) using a set of probabilistic context-free grammar (PCFG) rules as well as device context information for an attention utterance (Step 414). As described above, the natural language interface module 222 requires an attention strategy, e.g., the receipt of an attention word (i.e., Mona) that is unique to the user, or speaker identification coupled with allowable grammar rules.” *Id.* at [0080].

When searching for the attention word, Konopka ’188’s finite state network is a “minimum” recognition network because it is searching for only the attention word. This is consistent with the example of a “*minimum recognition network*” in the ’196 Patent which recognizes only the preset call command. ’196 Patent, 6:60-63.

As discussed above, Bou-Ghazale also teaches using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command.

The HTK toolkit also uses token passing. The HTK Book at 183:

For an unknown input utterance with T frames, every path from the start node to the exit node of the network which passes through exactly T emitting HMM states is a potential recognition hypothesis. Each of these paths has a log probability which is computed by summing the log probability of each individual transition in the path and the log probability of each emitting state generating the corresponding observation. Within-HMM transitions are determined from the HMM parameters, between-model transitions are constant and word-end transitions are determined by the language model likelihoods attached to the word level networks.

The job of the decoder is to find those paths through the network which have the highest log probability. These paths are found using a *Token Passing* algorithm. A token represents a partial path through the network extending from time 0 through time t . At time 0, a token is placed in every possible start node.

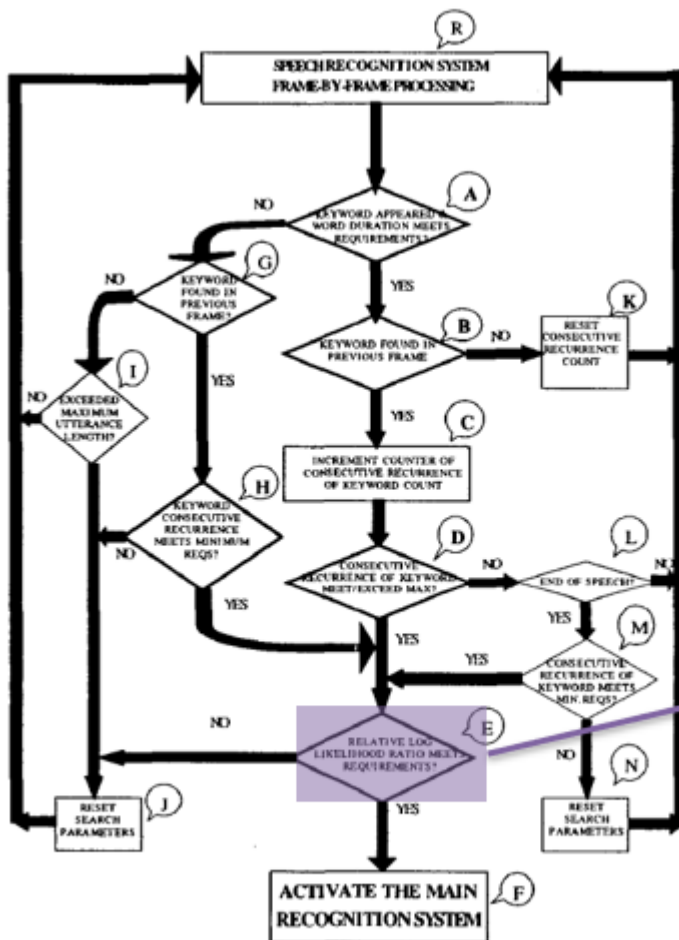
Regarding *wherein determining whether the input speech is recognized as the preset call command further comprises computing, by a confidence detection unit, a confidence rate of the input speech, wherein the preset call command recognition unit further comprises ... a confidence detection unit configured 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”), this functionality was well understood, routine, and conventional for all the reasons shown in

Defendants' invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to "List B" for confidence detection unit functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Bou-Ghazale teaches that the voice activation system ("preset call command recognition unit") comprises a post-processing component. Bou-Ghazale, Sec. 2, 1735. The post-processing component renders obvious the claimed "confidence detection unit" because, as will be explained below, it verifies whether the activation word has been found in the input speech by determining (i.e., detecting) a confidence rate for the input speech, called the "relative log-likelihood score." Bou-Ghazale, Abstract, 1735, Sec. 2.3-3.1, 1736. Further, Bou-Ghazale's "post-processing component" is a "unit" as claimed. See also '196 Patent, 5:65-67 ("the confidence detection unit 433 is a component").

Bou-Ghazale teaches that the post-processing component determines a "relative log-likelihood score" for the input speech ("computing...a confidence rate for the input speech") to determine whether the input speech is recognized as the activation word. Bou-Ghazale, Sec. 2.3-3.1, 1736. As explained with respect to limitation [1d], Bou-Ghazale's post-processing component verifies whether the activation word has been found in the input speech ("determining whether the input speech is recognized as the preset call command") based on whether the input speech meets the criteria and confidence measurements associated with the activation word. Bou-Ghazale, Sec. 2.3, 1736 ("the post-processing component decodes the result string on a frame-synchronous basis and applies a set of criteria and confidence measures to decide whether the activation word was a part of the utterance"), ("The activation word is declared to be found only after it meets a set of statistics derived from training data such as...the confidence level associated with the keyword."). For example, with respect to one such criteria illustrated by Block E in Figure 2 below, Bou-Ghazale teaches that its post-processing component determines a "relative log-likelihood score" (also called a "relative log likelihood ratio"²) for the input speech and compares it to the "confidence level associated with the keyword," where the confidence level is represented as a threshold value, η . Bou-Ghazale, Sec. 2.3, 1736, Sec. 3.1, 1736 ("Criterion 6: The relative log-likelihood score should be equal to or greater than a threshold, η , given by: $\eta = (\text{Score}_{\text{keyword}} - \text{Score}_{\text{OOV}}) / |\text{Score}_{\text{keyword}}|$, where the accumulated log-likelihood score for the keyword model is denoted by $\text{Score}_{\text{keyword}}$, and the log-likelihood score for the OOV model is denoted by $\text{Score}_{\text{OOV}}$."), Sec. 3.2, 1736 (explaining that "block E is associated with criterion 6"). A POSITA would have understood that Bou-Ghazale's "relative log-likelihood score" is a "confidence rate" because it is compared to another confidence rate, the "confidence level associated with the keyword" represented by the threshold value, η . Bou-Ghazale, Sec. 2.3-3.1, 1736.

² In Bou-Ghazale, the terms "log-likelihood score" and "log likelihood ratio" are used interchangeably.



Computing a relative-log-likelihood score ("confidence-rate")

Figure 2. Flow graph of the post-processing algorithm

Bou-Ghazale, Figure 2 (annotated).

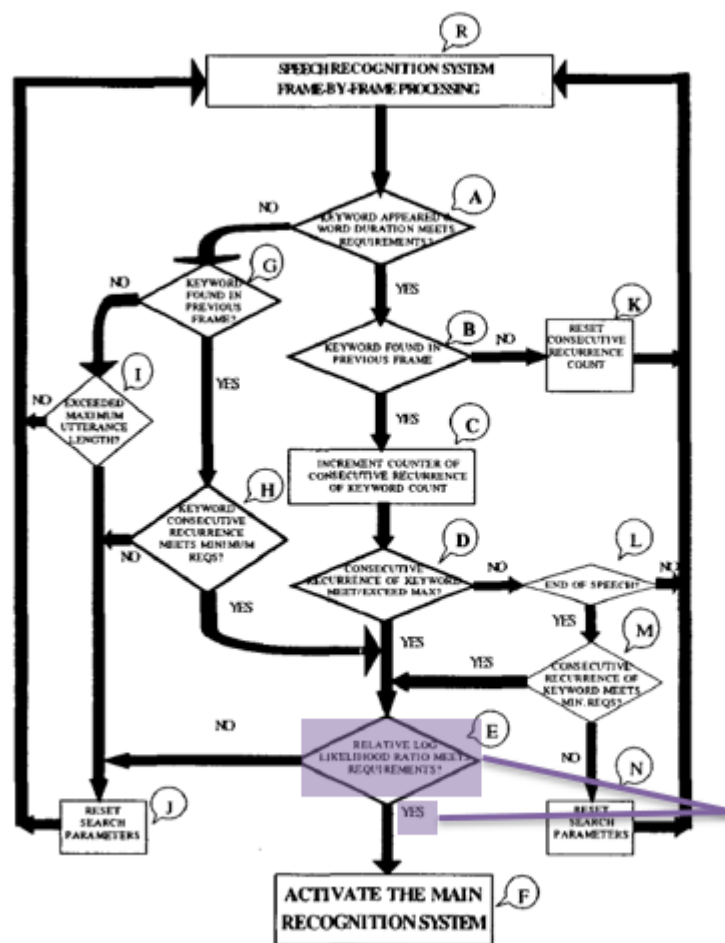
Thus, Bou-Ghazale's teaching of determining whether the activation word has been found in the input speech by determining, using the post-processing component of the voice activation system, a log-likelihood score for the input speech that is compared to a confidence level for the keyword renders obvious "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."

Regarding 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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for recognizing based on confidence functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Bou-Ghazale teaches that the post-processing component (“confidence detection unit”) “declares that the activation word has been found” (i.e., recogniz[ed]) based on the relative log-likelihood score of the input speech (“computed confidence rate”). Bou-Ghazale, Sec. 3.2, Figure 2, 1737.

Bou-Ghazale explains that “the activation word is declared to be found only after it meets a set of statistics derived from training data such as...the confidence level associated with the keyword.” Bou-Ghazale, Sec. 2.3, 1736. More specifically, Bou-Ghazale explains that the “relative log-likelihood score should be equal to or greater than a threshold, η ,” where the threshold, η is the confidence level for the activation word and the relative log-likelihood score is a “confidence rate” determined for the input speech, as explained above for limitation [2.0]. Bou-Ghazale, Sec. 2.3-3.1, 1736. As shown in Figure 2 below, if the “relative log likelihood ratio meets requirements,” then the post-processing component “declares that the activation word has been found” and activates the main speech recognition system. Bou-Ghazale, Sec. 3.2, 1736-37 (“[i]n the path traversing A, B, C, D, E, F, the keyword has met all the necessary requirements, and hence the post-processor declares that the activation word has been found”), Sec. 5, 1738 (“...once the activation word is recognized, the device turns on or switches to the main speech recognition system...”).



Declaring
 ("recognizing")
 that the
 activation word
 is found based
 on the relative
 log-likelihood
 score

Figure 2. Flow graph of the post-processing algorithm

Bou-Ghazale, Figure 2 (annotated).

Thus, Bou-Ghazale’s teaching of the post-processing component declaring that the activation word has been found in the input speech based on the relative log-likelihood score of the input speech renders obvious “recognizing, by the confidence detection unit, the input speech as the preset call command based on the computed confidence rate.”

Regarding 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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In

particular, the claim charts corresponding to “List B” for threshold functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Bou-Ghazale’s post-processing component determines a relative log-likelihood score (also called a ratio) for the input speech (“computed confidence rate”) and “declares that the activation word has been found” in the input speech if the relative log-likelihood score (“computed confidence rate”) is “equal to or greater than a threshold, η .” Bou-Ghazale, Sec. 3.1-3.2, Figure 2, 1736-37, Sec. 5, 1738.

Regarding *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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for determine authenticity functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Any reference that uses confidence to decide whether to recognize speech is determining whether the input speech is authentic based on the confidence.

Also, Bou-Ghazale teaches that its post-processing component (“confidence detection unit”) “declares that the activation word has been found” (i.e., recognized) based on whether the relative log-likelihood score of the input speech (“computed confidence rate”) is “equal to or greater than a threshold, η .” Bou-Ghazale, Sec. 3.1-3.2, 1736-37. Bou-Ghazale explains that the threshold value, η may be adjusted to achieve lower “false alarm rates”—i.e., a lower number of commands that are falsely recognized as the activation word. Bou-Ghazale, Sec. 4.4, 1738 (“As the threshold value is increased, the false rejection rate increases and the false alarm rate decreases.... For achieving very low false alarm rates of 0.26% for isolated utterances and 0% for embedded utterances, η can be set to 0.007.”). Because Bou-Ghazale describes false alarm rates as a “function of η ,” a POSITA would have understood that when the post-processing component compares the relative log-likelihood score of the input speech to the threshold, η , it is determining whether or not the input speech is a “false alarm” or false command. Bou-Ghazale, Tables 2-3. If the relative log-likelihood score is equal to or greater than the threshold, η , then the post-processing component determines that the input speech is an “authentic command” and not a false alarm, and “declares that the activation word has been found.” Bou-Ghazale, Sec. 3.1-3.2, 1736-37.

Thus, Bou-Ghazale’s teaching of determining whether or not the activation word is a false alarm and is found based on whether the relative log-likelihood score of the input speech meets a threshold, η renders obvious “wherein recognizing the input speech comprises: determining whether or not the input speech is an authentic call command based on the computed confidence rate.”

Regarding *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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for recognizing based on authenticity functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Any reference that uses confidence to decide whether to recognize speech is determining whether the input speech is authentic based on the confidence.

Bou-Ghazale’s post-processing component determines whether or not the activation word is an “authentic command” or a false alarm based on whether the relative log-likelihood score of the input speech meets the threshold requirement, η , as shown by the inquiry in Block E of Figure 2 below. Bou-Ghazale, Sec. 3.1-3.2, 1736-37. If the relative log-likelihood score is equal to or greater than the threshold, η , then the post-processing component determines that the input speech is an “authentic command” and not a false alarm, and “declares that the activation word has been found” (i.e., recognized) as shown by the “yes” path following Block E of Figure 2 below. Bou-Ghazale, Sec. 3.1-3.2, 1736-37.

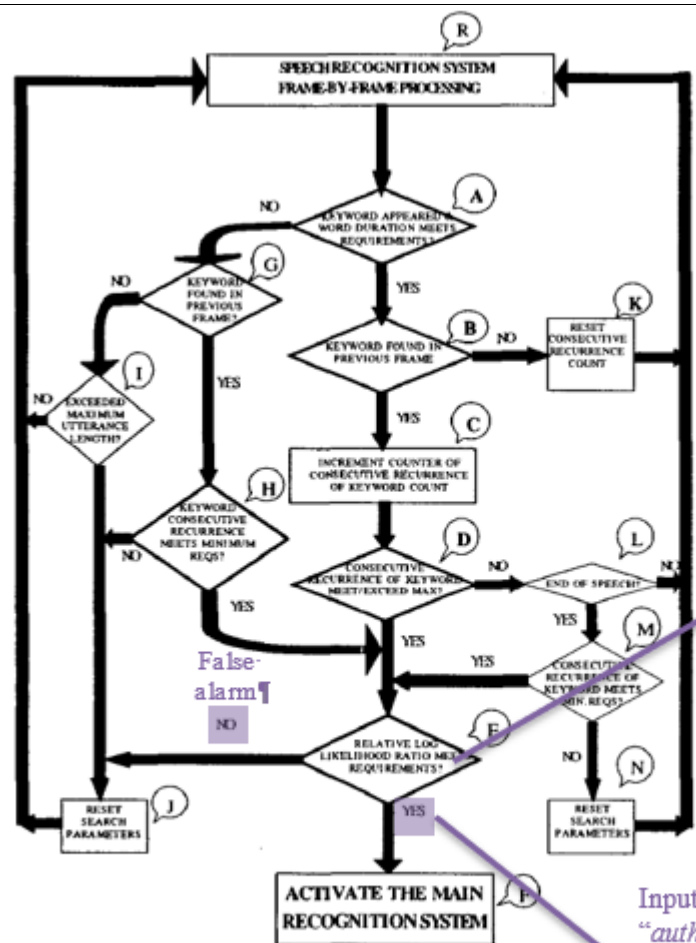


Figure 2. Flow graph of the post-processing algorithm

Determining-
whether-or-not
the-input-
speech-is-an-
"authentic-
command"-or-
a-false-alarm

Input-speech-is-an-
"authentic-
command"-and-
activation-word-is-
declared-found

Bou-Ghazale, Figure 2 (annotated).

Regarding 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"), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants' invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to "List B" for computing log likelihood ratio functionality

show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Konopka '188 teaches using HMM and garbage models to determine a “score” which is a log likelihood ratio equal to “[Ln(IV score)–Ln(OOV score)]/number of frames,” where Ln(IV score) is a natural logarithm of the “in-vocabulary (IV) Viterbi (likelihood) score” and Ln(OOV score) is a natural logarithm of the “out-of-vocabulary (OOV) Viterbi score.” Konopka '188, [0080], Figure 1. Konopka '188 explains that “a low score indicates a garbage utterance,” therefore the text strings with the highest garbage score indicate the best text strings that most likely match Konopka’s attention word. Konopka '188, [0080]. Konopka '188 also explains that it was known in the art to determine the best text strings based on just a likelihood (Viterbi) score for the in-vocabulary words instead of based on both the in-vocabulary and the out-of-vocabulary scores. Konopka '188, [0043] (“if a Viterbi score passes a threshold, the utterance is determined to be an in-vocabulary word....Such a Viterbi score is generated using the Viterbi algorithm. This algorithm calculates a single best state sequence through an HMM and its corresponding probability, given an observation sequence.). Accordingly, a POSITA would have understood that the best text strings could be determined by computing a logarithm of the in-vocabulary score (“log likelihood ratio of the input speech”).

Regarding *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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for assuming based on computed log likelihood ratio functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Bou-Ghazale’s speech recognition component (“token passing unit”) assumes the input speech to be the activation word when it passes the result string with the highest likelihood score (“log likelihood ratio”) to the post-processing component.

Regarding *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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for phoneme model functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

First, HMMs are or can be configured as phoneme models. “A phoneme is a unit of speech, the set of which defines all the sounds from which words can be constructed in a particular language.” Schmandt at 14; *see also id.* at 133 (“Hidden Markov Models ... are a statistical representation based on the various ways words or phonemes may be pronounced.”); Lee III at 35. Many references discussed above teach HMMs and, therefore, teach phoneme models.

Lee I teaches a “voice code verification” system and process, that uses “a phoneme model...modeled using HMM having states” and an “anti-phoneme model.” Lee I, 29 (“This is triphone, which is the model used in the proposed system. Such a phoneme model is modeled using HMM having states.”), 33, 38 (“The proposed anti-phoneme model is based on Korean phonetic theory.”). In particular, Lee I describes a two-step process, where in the first step “recognition is performed via a conventional Viterbi beam search algorithm that segments the test utterance into the N-best strings of a phoneme hypotheses.” Lee I, 30. As was well-known in the art, the Viterbi algorithm uses models to analyze the input speech, such as acoustic models which, Lee I explains, “represent general characteristics of speech” and “can be modeled as a word unit or a phoneme unit.” Lee I, 29; *see also* Konopka ’188, (“As is well understood in the art, the received utterances are analyzed with the models, e.g., using a Viterbi algorithm, and scores are assigned representing how well the utterance fits the given models.”).

Regarding *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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for subtracting functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Konopka ’188 discloses computing log likelihood ratios of the input speech and subtracting one log likelihood ratio from the other. For example, Konopka ’188 teaches “[a] garbage score [that] can then be defined as the difference between the logarithms of each of the two Viterbi scores divided by the number of frames in the utterance according to equation 1 below. The garbage score reveals whether the utterance had a closer match with the word models or the out-of-vocabulary models.” Konopka ’188 at [0043]. Equation 1 is shown below:

- “Next, the out-of vocabulary HMMs process the same portion of the utterance and the garbage scores can be calculated as,

$$s = \frac{w - g}{n - m}, \quad (1)$$

where w is the logarithm of the Viterbi score for the acoustic models of in-vocabulary words without preceding or following silence models and where no silence is included in the utterance. Similarly, g is the logarithm of the corresponding score for the out-of-vocabulary HMM models.

Also, n is the total number of frames in the utterance and m is the number of frames that were consumed by the preceding and following silence models. In summary, using this rejection technique, the system is better able to accurately isolate the speech portion of the utterance. This has the effect of better isolating in-vocabulary words and rejecting out-of vocabulary events that begin or end with low energy sounds, such as fricatives, in comparison to conventional rejection schemes.” *Id.* at [0044].

Regarding *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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for rule check functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Hee-souk ’469 teaches “Decision Rule Processing.”

Tran ’140 teaches a set of phoneme rules. *See* Tran ’140 at Fig. 31.

Regarding *a processor* (“processor functionality”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for processor functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. Processors were clearly well known, understood, routine, and conventional.

Regarding *a preset call command recognition unit using the processor* (“PCCRU using processor functionality”) and *a continuous speech recognition unit using the processor* (“CSRU using processor functionality”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for PCCRU using processor and CSRU using processor functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. Preset call command recognition units and speech recognition units would need a processor to process speech.

Regarding *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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for filter

functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Takiguchi '832 teaches that an “adaptive filter” can be used to adjust the frequency range in order to more effectively reduce extemporaneous noises and improve speech recognition. Takiguchi '832, [0026]-[0027] (“it is an aspect of the present invention to provide a speech enhancement technique which is effective for an extemporaneous noise without a noise interval and also for unknown extemporaneous noises”). In particular, Takiguchi '832 describes a “signal enhancement device designed to enhance a target signal...by controlling a filter coefficient of an adaptive filter to be applied to the reference signal to reduce the noise signal.” Takiguchi '832, [0027]. Takiguchi '832 explains that the “reference signal” to which the “adaptive filter” is applied to is obtained from “input signals” including speech utterances. Takiguchi '832, [0086] (explaining that the “signal synthesis unit 31 also adds the input signal m2 to the input signal m1 after inverting the input signal m2... and outputs the resultant signal as the reference signal...The adaptive filter unit 33 adjusts the amplitude of the reference signal $r(\omega, n)$ in accordance with the filter coefficient $w(\omega)$.”), [0094] (“input signals m1(t) and m2(t) were created using utterance data”). Accordingly, Takiguchi '832’s “adaptive filter” receives “input speech” in the form of reference signals.

Rabiner III also teaches an adaptive filter is well known and conventional:

5.7.2.1 Signal-Enhancement Preprocessing

When the adverse condition is due to additive noise alone, one can use well-established speech-enhancement methods to suppress the noise before applying the recognition algorithm. One of the most widely studied signal-enhancement methods is adaptive noise cancellation using two signal sources [25]. Figure 5.27 illustrates the principle of the two-input noise cancelling approach. The technique adaptively adjusts a set of filter coefficients so that subtracting the filtered noise input from the primary (speech plus noise) input leads to an output signal with minimum energy. This technique, however, requires that the noise component in the corrupted signal and the noise reference have high coherence. Inside a passenger car, it was found [26] that if the two microphones are located at a distance greater than 50 cm, the only coherent noise component is that of the engine. To cancel 90% of the noise energy, the two microphones cannot be more than 5 cm apart, which makes it almost impossible to prevent speech from being included in the noise reference. Therefore, one can only expect cancellation of noise related to the fundamental frequency of the engine revolution rate if the two microphones are spaced at a reasonable distance. This lack of noise coherence was also observed in the sound field of a fighter aircraft cockpit.

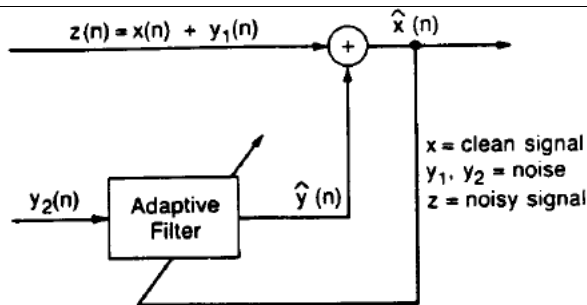


Figure 5.27 Schematic of the two-input noise cancelling approach.

Rabiner III at 310.

See also Rabiner III at 306-07 (discussing adverse noise conditions).

Regarding *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”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for keyword spotting functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

This is what speech recognition systems do. All of the speech recognition systems above disclose spotting keywords.

Regarding *wherein the single device is a portable device or an embedded system* (“portable or embedded functionality”), this functionality was well understood, routine, and conventional for all the reasons shown in Defendants’ invalidity contentions corresponding to this functionality. In particular, the claim charts corresponding to “List B” for portable or embedded functionality show that it was well understood, routine, and conventional at the time of the claimed invention. Those contentions and the charts for the List B references are hereby incorporated by reference. In addition, the following are particular examples of how this element was well understood, routine, and conventional.

Microsoft Speech Recognition is software embedded on a PC.

Bou-Ghazale and many other references teach speech recognition systems incorporated in portable devices.

Thus, for the reasons above as well as those in Defendants' Invalidity Contentions, the elements of each Challenged Claim, both individually and in combination, were well understood, routine, and conventional in the relevant industry at the relevant time.