
IN THE
UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON WEB SERVICES, INC.
AND AMAZON.COM SERVICES LLC,

Petitioners

- vs. -

ZENTIAN LIMITED

Patent Owner

Patent No. 10,971,140

Issued: April 6, 2021

Inventor: Mark Catchpole

Title: SPEECH RECOGNITION CIRCUIT USING PARALLEL PROCESSORS

Inter Partes Review No. IPR2023-01197

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 10,971,140**

Mail Stop Patent Board
Patent Trial and Appeal Board
P.O. Box 1450
Alexandria, VA 22313-1450

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. SUMMARY OF THE '140 PATENT	1
A. Description of the '140 Patent	1
B. Summary of Unpatentability	2
C. Priority Date of the Challenged Claims	6
D. Level of Skill of a POSITA	6
III. REQUIREMENTS FOR IPR UNDER 37 C.F.R. § 42.104	7
A. Grounds for Standing Under 37 C.F.R. § 42.104(a)	7
B. Identification of Challenged Claims Under 37 C.F.R. § 42.104(b) and Relief Requested	7
C. Claim Construction Under 37 C.F.R. § 42.104(b)(3)	8
IV. SHOWING OF ANALOGOUS PRIOR ART	8
A. <i>Jiang</i>	8
B. <i>Chen</i>	9
C. <i>Lucke</i>	9
D. Robinson	9
E. <i>Wrench</i>	10
V. GROUND 1: CLAIMS 1-3, 5, AND 7-8 ARE OBVIOUS OVER <i>JIANG</i> IN VIEW OF <i>CHEN</i>	10
A. Obviousness of <i>Jiang</i> Combined with <i>Chen</i>	10
1. Motivation to Combine Jiang and Chen	10

TABLE OF CONTENTS (Continued)

	Page
2. Obviousness of Modifying Electronic Hardware and Software to Be Circuitry	12
B. Claim 1	13
1. Claim 1(Pre): “A speech recognition circuit comprising:”	13
2. Claim 1(a): “one or more clusters of processors, each of the one or more clusters of processors comprising: a plurality of processors;”	15
a) <i>Jiang</i> Teaches a Multiprocessor System	15
b) The ’140 Patent’s Description of a “Cluster”	15
c) <i>Chen</i> Teaches Claim 1(a)	16
d) Motivation to Combine <i>Chen</i> with <i>Jiang</i> for Claim 1(a)	18
3. Claim 1(b): “and [one or more clusters of processors, each of the one or more clusters of processors comprising:] an acoustic model memory storing acoustic model data”	20
a) <i>Jiang</i> Teaches an Acoustic Model Memory Storing Acoustic Model Data	20
b) <i>Chen</i> Teaches One or More Clusters of Processors Each Comprising Memory	23
c) Motivation to Combine <i>Chen</i> with <i>Jiang</i> for Claim 1(b)	25
4. Claim 1(c): “wherein each of the plurality of processors is configured to compute a probability using the acoustic model data in the acoustic model memory”	26

TABLE OF CONTENTS (Continued)

	Page
a) <i>Jiang</i> Teaches “comput[ing] a probability using the acoustic model data”	26
b) <i>Jiang</i> in Combination with <i>Chen</i> Teaches “each of the plurality of processors [is configured to compute] ... using the acoustic model data in the acoustic model memory”	29
c) Motivation to Combine <i>Chen</i> with <i>Jiang</i> for Claim 1(c).....	32
5. Claim 1(d): “[wherein] the speech recognition circuit is configured to generate an initial score for an audio sample”	32
a) <i>Jiang</i> Teaches an “audio sample”	32
b) <i>Jiang</i> Teaches “the speech recognition circuit is configured to generate an initial score”	33
6. Claim 1(e): “[wherein] the initial score is used to determine whether to continue processing to determine a final score via processing a larger amount of model data than that was processed to generate the initial score”	37
a) Prosecution History of the ’140 Patent Regarding Claim 1(e)	37
b) <i>Jiang</i> Teaches Claim 1(e).....	39
C. Claim 2:.....	45
1. Claim 2: “...wherein the probability is an input to an evaluation of a state transition of a model of states”	45
D. Claim 3.....	48

TABLE OF CONTENTS (Continued)

	Page
1. Claim 3: “...wherein the model is a Hidden Markov Model”	48
E. Claim 5.....	49
1. Claim 5: “...further comprising: a buffer for storing one or more feature vectors coupled to at least one of the plurality of processors of the one or more clusters of processors”	49
F. Claim 7.....	50
1. Claim 7: “...the one or more clusters of processors comprises a first cluster of processors and a second cluster of processors;	50
G. Claim 8.....	53
1. Claim 8: “...the one or more processors comprises a first cluster of processors; and	53
VI. GROUND 2: CLAIMS 1-3, 5, AND 7-8 ARE OBVIOUS OVER <i>JIANG</i> IN VIEW OF <i>CHEN</i> IN FURTHER VIEW OF <i>LUCKE</i>	54
A. Overview of Ground 2	54
B. Claim 1.....	55
1. Claim 1(Pre): “A speech recognition circuit comprising:”	55
2. Claim 1(a): “one or more clusters of processors, each of the one or more clusters of processors comprising: a plurality of processors;”	55
3. Claim 1(b): “and [one or more clusters of processors, each of the one or more clusters of processors comprising:] an acoustic model memory storing acoustic model data”	55

TABLE OF CONTENTS (Continued)

	Page
4. Claim 1(c): “wherein each of the plurality of processors is configured to compute a probability using the acoustic model data in the acoustic model memory, wherein:”	55
5. Claim 1(d): “the speech recognition circuit is configured to generate an initial score for an audio sample; and	55
6. Claim 1(e): “the initial score is used to determine whether to continue processing to determine a final score via processing a larger amount of model data than that was processed to generate the initial score”	55
C. Claim 2.....	58
1. Claim 2: “...the probability is an input to an evaluation of a state transition of a model of states”	58
D. Claim 3.....	58
1. Claim 3: “...the model is a Hidden Markov Model”	58
E. Claim 5.....	58
1. Claim 5: “...a buffer for storing one or more feature vectors coupled to at least one of the plurality of processors of the one or more clusters of processors”	58
F. Claim 7.....	59
1. Claim 7: “...the one or more clusters of processors comprises a first cluster of processors and a second cluster of processors;.....	59
G. Claim 8.....	59
1. Claim 8: “...the one or more processors comprises a first cluster of processors; and	59

TABLE OF CONTENTS
(Continued)

	Page
VII. GROUND 3: CLAIM 4 IS OBVIOUS OVER <i>JIANG</i> IN VIEW OF <i>CHEN</i> IN FURTHER VIEW OF <i>ROBINSON</i>	60
A. Claim 4: “...wherein the probability is computed from a Gaussian mixture model and one or more feature vectors”	60
VIII. GROUND 4: CLAIM 4 IS OBVIOUS OVER <i>JIANG</i> IN VIEW OF <i>CHEN</i> IN FURTHER VIEW OF <i>LUCKE</i> IN FURTHER VIEW OF <i>ROBINSON</i>	61
A. Claim 4.....	61
IX. GROUND 5: CLAIM 6 IS OBVIOUS OVER <i>JIANG</i> IN VIEW OF <i>CHEN</i> IN FURTHER VIEW OF <i>WRENCH</i>	61
A. Claim 6: “...a search controller coupled to the at least one of the plurality of processors of the one or more clusters of processors, the search controller capable of controlling the at least one of the plurality of processors of the one or more clusters of processors to initiate speech recognition processing in at least a subset of processors of the one or more clusters of processors”	61
X. GROUND 6: CLAIM 6 IS OBVIOUS OVER <i>JIANG</i> IN VIEW OF <i>CHEN</i> IN FURTHER VIEW OF <i>LUCKE</i> IN FURTHER VIEW OF <i>WRENCH</i>	63
A. Claim 6.....	63
XI. THE BOARD’S DISCRETION UNDER 35 U.S.C. § 314(a) and § 325(d).....	63
XII. CONCLUSION.....	65
XIII. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8(A)(1)	66
A. Real Party-In-Interest.....	66
B. Related Matters	66

TABLE OF CONTENTS
(Continued)

	Page
C. Lead and Back-Up Counsel	66
D. Service Information	67
CLAIM LISTING APPENDIX	68

TABLE OF AUTHORITIES

	Page(s)
CASES	
<i>Apple Inc. v. Fintiv, Inc.</i> , IPR2020-00019 (PTAB, Mar. 20, 2021)	64
<i>General Plastic Industrial Co., Ltd. v. Canon Kabushiki Kaisha</i> , IPR2016-01357, Paper 19 (PTAB Sept. 6, 2017).....	63
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303 (Fed. Cir. 2005) (<i>en banc</i>)	8
STATUTES AND RULES	
35 U.S.C. § 102	8, 9, 10
35 U.S.C. § 314(A)	63
35 U.S.C. § 315(c)	1, 7
35 U.S.C. § 325(D)	63
OTHER AUTHORITIES	
37 C.F.R. § 42.100(b)	8
37 C.F.R. § 42.104	7, 8
37 C.F.R. § 42.122	7
83 Fed. Reg. 197 (Oct. 11, 2018).....	8

EXHIBIT LIST (37 C.F.R. § 42.63(E))

Exhibit	Description
1001	U.S. Patent No. 10,971,140 (“’140 Patent”)
1002	Prosecution History for the 10,971,140 Patent (“’140 File History”)
1003	Declaration of Christopher Schmandt (“Dec.”)
1004	U.S. Patent No. 6,374,219 to Jiang (“Jiang”)
1005	U.S. Patent No. 5,428,803 to Chen et al. (“Chen”)
1006	Hsiao-Wuen Hon, <i>A survey of hardware architectures designed for speech recognition</i> , Carnegie Mellon University, 1991 (“Hon”)
1007	Ph.D. Thesis of Mosur Ravishankar (“Ravishankar”)
1008	U.S. Patent Application Publication No. 2001/0053974 to Lucke et al. (“Lucke”)
1009	U.S. Patent No. 5,983,180 to Robinson (“Robinson”)
1010	U.S. Patent No. 5,036,539 to Wrench Jr., et al. (“Wrench”)
1011	Frederick Jelinek, <i>Statistical Methods for Speech Recognition</i> , The MIT Press, 1997 (“Jelinek”)
1012	Christopher Schmandt, <i>Voice Communication with Computers</i> , Van Nostrand Reinhold, 1994 (“Schmandt”)
1013	Lawrence Rabiner and Biing-Hwang Juang, <i>Fundamentals of Speech Recognition</i> , Prentice Hall PTR, 1993 (“Rabiner”)
1014	Richard Klevans and Robert Rodman, <i>Voice Recognition</i> , Artech House, 1997 (“Klevans”)
1015	U.S. Patent No. 7,120,582 to Young et al. (“Young”)
1016	John Holmes and Wendy Holmes, <i>Speech Synthesis and Recognition</i> , 2nd Edition, Taylor & Francis, 2001 (“Holmes”)
1017	U.S. Patent No. 4,926,488 to Nadas et al. (“Nadas”)

Petition for *Inter Partes* Review of
U.S. Patent No. 10,971,140

Exhibit	Description
1018	U.S. Patent No. 6,182,037 to Maes (“ <i>Maes</i> ”)
1019	U.S. Patent Application Publication No. 2003/0220796 to Aoyama et al. (“ <i>Aoyama</i> ”)
1020	U.S. Patent No. 6,092,045 to Stubley et al. (“ <i>Stubley</i> ”)
1021	U.S. Patent No. 6,151,574 to Lee et al. (“ <i>Lee</i> ”)
1022	U.S. Patent No. 6,580,814 to Ittycheriah et al. (“ <i>Ittycheriah</i> ”)
1023	U.S. Patent No. 6,452,348 to Toyoda (“ <i>Toyoda</i> ”)
1024	William A. Wulf and C.G. Bell, <i>C.mmp—A multi-mini-processor</i> , Carnegie-Mellon University, 1972 (“ <i>Wulf</i> ”)
1025	Lee D. Erman, Richard D. Fennell, Victor R. Lesser, and D. Raj Reddy, <i>System Organizations for Speech Understanding: Implications of Network and Multiprocessor Computer Architectures for AI</i> , IEEE Transactions on Computers, Vol. C-25, No. 4, April 1976 (“ <i>Erman</i> ”)
1026	U.S. Patent No. 6,393,481 to Deo et al. (“ <i>Deo</i> ”)
1027	U.S. Patent No. 6,615,338 to Tremblay et al. (“ <i>Tremblay</i> ”)
1028	U.S. Patent No. 5,922,076 to Garde (“ <i>Garde</i> ”)
1029	Lawrence R. Rabiner, <i>A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition</i> , Proceedings of the IEEE, Vol. 77, No. 2, February 1989 (“ <i>Rabiner 89</i> ”)
1030	[RESERVED]
1031	[RESERVED]
1032	Zentian, Ltd. v Apple Inc. WDTX Waco (6:22-cv-122) Scheduling Order “ <i>Scheduling Order</i> ”
1033	Declaration of June Ann Munford (“ <i>Munford</i> ”)
1034	[RESERVED]
1035	[RESERVED]

Petition for *Inter Partes* Review of
U.S. Patent No. 10,971,140

Exhibit	Description
1036	<i>[RESERVED]</i>
1037	<i>[RESERVED]</i>
1038	<i>[RESERVED]</i>
1039	<i>[RESERVED]</i>
1040	<i>[RESERVED]</i>
1041	<i>[RESERVED]</i>
1042	<i>[RESERVED]</i>
1043	Declaration of Les Atlas, Ph.D.
1044	<i>Zentian Ltd. v. Amazon.com Services LLC, et al.</i> , No. 6:22-cv-123 (W.D. Tex.) Mar. 22, 2023 Amended Scheduling Order

This Petition raises grounds of unpatentability identical to those in already-instituted IPR2023-00037. This Petition neither adds to nor alters any arguments the Board has already considered, will not expand the instituted grounds, and is accompanied by a timely motion for joinder under 35 U.S.C. § 315(c).

I. INTRODUCTION

Petitioners Amazon.com Services LLC and Amazon Web Services, Inc. (collectively, “Petitioners” or “Amazon”) request *Inter Partes* Review of Claims 1-8 (collectively, the “Challenged Claims”) of USPN 10,971,140 (“’140 Patent”). (Ex. 1001). The purportedly distinguishing features of the Challenged Claims—using clustered processors and determining whether to continue processing speech using a larger amount of model data based on an initial score—were well-known before the priority date of the ’140 Patent, and the Challenged Claims are obvious over the prior art as detailed herein. Accordingly, IPR of the Challenged Claims should be instituted.

II. SUMMARY OF THE ’140 PATENT

A. Description of the ’140 Patent

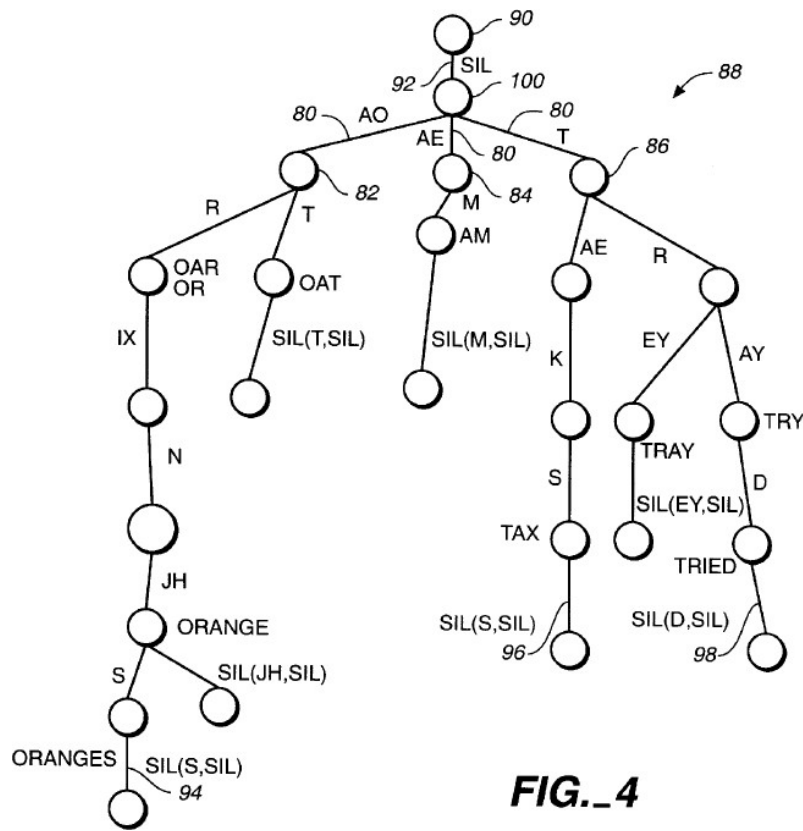
The ’140 Patent is directed to a speech recognition circuit utilizing clustered processors to perform speech recognition using an acoustic model stored in memory. ’140 Patent, 1:18-20, 5:6-29, Fig. 2. The claimed circuit generates an initial score for an audio sample, used to determine which possible word paths in a lexical tree should be “pruned” from its search. ’140 Patent, 6:21-35, 8:59-9:4.

B. Summary of Unpatentability

Claim 1's speech recognition circuit generally recites an acoustic model memory, well-known speech recognition steps associated with generating scores to inform processing of an audio sample, and one or more clusters of processors for performing the speech recognition steps. The proposed combination for Ground 1, Claim 1 relies on *Jiang* and *Chen*. *Jiang* is applied to teach the well-known speech recognition system comprising an acoustic model memory and speech recognition steps. *Chen* is applied to teach the well-known clusters of processors. Performing speech recognition using clusters of processors would have been obvious at the time of the '140 Patent. Indeed, *Jiang* already teaches a multi-processor configuration, and clusters of processors was known to "create a high performance parallel processing computer system." *Jiang*, 4:60-65, 6:39-42; *Chen*, 5:9-13.

Jiang teaches a speech recognition system comprising a CPU 21, "which may include one or more processors[.]" *Jiang*, 6:39-42, 4:60-65. *Jiang* teaches utilizing "phonetic speech unit models, such as hidden Markov models, which represent speech units to be detected by system 60." *Jiang*, 7:34-36, 7:36-38 (teaching the stored phonetic models "includes HMMs which represent phonemes"). A Hidden Markov Model (HMM) comprises a plurality of states for determining word or sub-word units (e.g., phonemes) that can be organized into a lexical tree. A "conventional technique" for reducing computation requirements, referred to in

Jiang as a “prefix tree,” associates phonetic models, such as those modeled by HMMs, with the branches of the prefix tree. *Jiang*, 2:47-59; *Ravishankar*, 33, 67-71 (generally describing the known benefits of the lexical tree structure) (Ex. 1007); *Dec.*, 84 (opining a POSITA would have understood *Jiang*’s “prefix tree” to be a “lexical tree”). In *Jiang*, the prefix tree is representative of phonemes, and the prefix tree comprises a plurality of nodes connected by phoneme branches. *Jiang*, 2:53-57. A score is assigned to each node of the prefix tree as the tree is traversed. *Jiang*, 8:16-24. *Jiang*’s speech recognition circuit computes “probability score[s]” for each state of an acoustic model stored in memory, such as the HMM, and then uses the output probabilities to search the prefix tree in a “known manner.” *Jiang*, 2:3-15, 1:19-32, 8:16-51.



Jiang, Fig. 4, 1:19-2:15, 9:28-42. Using initial generated scores, *Jiang's* speech recognition system prunes branches from the search, continuing to process (using additional data from the acoustic model) only branches whose score is sufficiently high. *Jiang*, 2:25-46, 8:52-64, 10:51-11:15.

Jiang does not teach clusters of processors. In related art, *Chen* teaches a “parallel processing architecture for connecting together an extendible number of clusters of multiple numbers of processors to create a high performance parallel processing computer system.” *Chen*, 5:9-13. As illustrated in annotated Fig. 4 below, *Chen*’s clusters 100a-d each includes a “Cluster Shared Memory” 104a-d accessible to each of the plurality of processors 102a-d in the respective cluster.

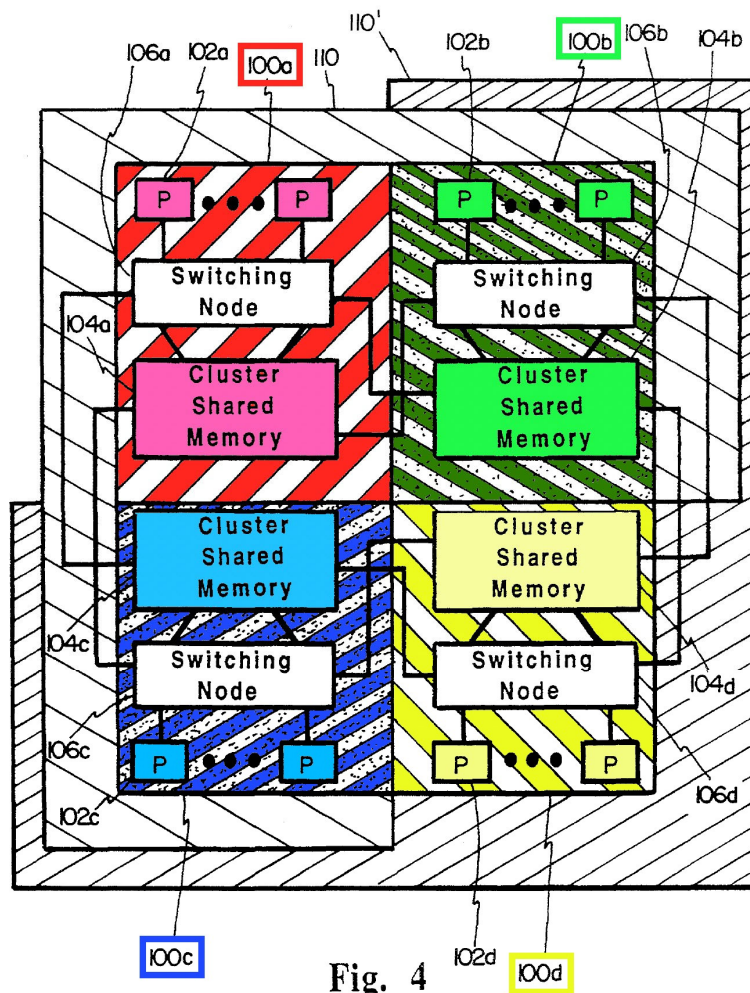


Fig. 4

Chen, Fig. 4 (annotated), 9:5-43. Thus, the speech recognition circuit of *Jiang* modified to utilize *Chen*'s architecture of a plurality of clustered processors renders obvious every limitation of the Challenged Claims, as detailed in the mapping below.

C. Priority Date of the Challenged Claims

The '140 Patent claims priority to a GB application filed February 4, 2002. '140 Patent, (22), (63), (30). For purposes of this Petition, Petitioners apply February 4, 2002, as the priority date for the Challenged Claims.

D. Level of Skill of a POSITA

A POSITA at the time of the '140 Patent would have had a master's degree in computer engineering, computer science, electrical engineering, or a related field, with at least two years of experience in the field of speech recognition, or a bachelor's degree in the same fields with at least four years of experience in the field of speech recognition. Additional education or experience might substitute for the above requirements. *Dec.*, 24-26.¹

¹ All citations to "*Dec.*" are to paragraph numbers in Ex. 1003, Declaration of Christopher Schmandt. Submitted herewith is the Declaration of Les Atlas, Ph.D., a qualified technical expert. EX1043, ¶¶ 4-8. Dr. Atlas has reviewed the opinions in the Declaration of Christopher Schmandt (EX1003 in IPR2023-00037). Dr. Atlas agrees with and adopts those opinions as his own. EX1043, ¶¶ 25-26. For

III. REQUIREMENTS FOR IPR UNDER 37 C.F.R. § 42.104

A. Grounds for Standing Under 37 C.F.R. § 42.104(a)

Petitioners certify the '140 Patent is available for IPR and Petitioners are not barred or estopped from requesting IPR challenging the claims of the '140 Patent. Because Petitioners file an accompanying motion for joinder within one month of institution in IPR2023-00037, this petition is timely under 35 U.S.C. § 315(c) and 37 C.F.R. § 42.122.

B. Identification of Challenged Claims Under 37 C.F.R. § 42.104(b) and Relief Requested

In view of the prior art and evidence presented, the Challenged Claims of the '140 Patent are unpatentable and should be cancelled. 37 C.F.R. § 42.104(b)(1). Based on the prior art identified below, IPR of the Challenged Claims should be granted. 37 C.F.R. § 42.104(b)(2).

Proposed Grounds of Unpatentability	
<u>Ground 1:</u>	Claims 1-3, 5, and 7-8 are obvious over <i>Jiang</i> and <i>Chen</i>
<u>Ground 2:</u>	Claims 1-3, 5, and 7-8 are obvious over <i>Jiang</i> , <i>Chen</i> , and <i>Lucke</i>
<u>Ground 3:</u>	Claim 4 is obvious over <i>Jiang</i> , <i>Chen</i> , and <i>Robinson</i>
<u>Ground 4:</u>	Claim 4 is obvious over <i>Jiang</i> , <i>Chen</i> , <i>Lucke</i> , and <i>Robinson</i>
<u>Ground 5:</u>	Claim 6 is obvious over <i>Jiang</i> , <i>Chen</i> , and <i>Wrench</i>

consistency with Apple's Petition, citations herein to Dr. Atlas's opinion will be citations to EX1003 containing those opinions as expressed by Mr. Schmandt.

Proposed Grounds of Unpatentability	
Ground 6:	Claim 6 is obvious over <i>Jiang, Chen, Lucke, and Wrench</i>

Sections V-X identify where each element of the Challenged Claims is found in the prior art. 37 C.F.R. § 42.104(b)(4). The exhibit numbers of the supporting evidence relied upon to support the challenges are provided above and the relevance of the evidence to the challenges raised are provided in Sections V-X. 37 C.F.R. § 42.104(b)(5). Exhibits 1001–1044 are also attached.

C. Claim Construction Under 37 C.F.R. § 42.104(b)(3)

In this proceeding, claims are interpreted under the same standard applied by Article III courts (i.e., the *Phillips* standard). 37 C.F.R. § 42.100(b); 83 Fed. Reg. 197 (Oct. 11, 2018); *Phillips v. AWH Corp.*, 415 F.3d 1303, 1312 (Fed. Cir. 2005) (*en banc*). Petitioners apply the plain and ordinary meaning of all claim terms. Petitioners do not waive any argument in any litigation that claim terms in the '140 Patent are indefinite or additional terms need construction.

IV. SHOWING OF ANALOGOUS PRIOR ART

A. *Jiang*

Jiang was neither cited nor considered during prosecution of the '140 Patent. *Jiang* was filed February 20, 1998, and issued April 16, 2002, qualifying as prior art under 35 U.S.C. § 102(e). *Jiang* describes a speech recognition circuit utilizing an

HMM acoustic model. *Compare Jiang* (Ex. 1004), 6:18-28, 7:29-54, with '140 Patent, 1:30-44, 6:41-48; Dec., 62. Thus, *Jiang* is analogous art to the '140 Patent.

B. *Chen*

Chen was neither cited nor considered during prosecution of the '140 Patent. *Chen* issued June 27, 1995, qualifying as prior art under 35 U.S.C. § 102(b). *Chen* describes a parallel processing computing platform utilizing clustered processors. *Compare Chen* (Ex. 1005), 5:9-17, 9:5-43, Fig. 4, with '140 Patent, 3:13-24, 5:6-36, Fig. 2; Dec., 63. Thus, *Chen* is analogous art to the '140 Patent.

C. *Lucke*

Lucke was neither cited nor considered during prosecution of the '140 Patent. *Lucke* was filed March 12, 2001, and published December 20, 2001, qualifying as prior art under 35 U.S.C. § 102(e). *Lucke* describes a speech recognition circuit utilizing an HMM acoustic model. *Compare Lucke* (Ex. 1008), 0002-0005, 0008, 0025, 0079 with '140 Patent, 1:30-44, 6:41-48; Dec., 64. Thus, *Lucke* is analogous art to the '140 Patent.

D. *Robinson*

Robinson was neither cited nor considered during prosecution of the '140 Patent. *Robinson* issued November 9, 1999, qualifying as prior art under 35 U.S.C. § 102(b). *Robinson* describes a speech recognition circuit comparing feature vectors

to a speech model. *Compare Robinson* (Ex. 1009), 8:28-55, 9:4-16, with '140 Patent, 4:52-67, 5:6-26; *Dec.*, 65. Thus, *Robinson* is analogous art to the '140 Patent.

E. *Wrench*

Wrench was neither cited nor considered during prosecution of the '140 Patent. *Wrench* issued July 30, 1991, qualifying as prior art under 35 U.S.C. § 102(b). *Wrench* describes a speech recognition circuit comprising a plurality of processors directed by a single controller. *Compare Wrench* (Ex. 1010), 4:40-58, 6:38-7:7, Fig. 2 with '140 Patent, 5:46-52, 10:3-30, Fig. 2; *Dec.*, 66. Thus, *Wrench* is analogous art to the '140 Patent.

V. GROUND 1: CLAIMS 1-3, 5, AND 7-8 ARE OBVIOUS OVER *JIANG* IN VIEW OF *CHEN*

A. Obviousness of *Jiang* Combined with *Chen*

1. Motivation to Combine *Jiang* and *Chen*

As discussed, *Jiang* and *Chen* are analogous art to the '140 Patent, and a POSITA would have been familiar with both references. A POSITA would have found it obvious and been motivated to combine *Chen*'s cluster of processors with shared memory with *Jiang*'s speech recognition circuit using an HMM acoustic model to generate scores, as taught by *Jiang*. *Dec.*, 67. Computing platforms providing processors in a cluster were extremely well-known in the art prior to the '140 Patent, as taught by *Chen* and numerous other references discussed by Mr. Schmandt. *Dec.*, 67. Prior to the priority date of the '140 Patent, connecting a

plurality of processors in parallel was a well-known technique to increase the processing power available to a computing circuit without becoming expensive to produce. *Dec.*, 67. It was also well-known that speech recognition was a computationally-intensive task, requiring expensive, sophisticated processors (for the time) if performed in real-time on a solitary, single-core processor. *Dec.*, 67. Well before the priority date of the '140 Patent, clusters of processors were used to recognize speech. *Id.*

The combination of the disclosures of *Jiang* and *Chen* would have constituted use of a known technique (e.g., *Chen*'s clustering of processors with memory) to improve a similar device (*Jiang*'s multiprocessor speech recognition circuit) in the same way. *Dec.*, 67. A POSITA would have recognized utilizing a clustered processor architecture with shared memory would have enabled the speech recognition circuit to efficiently process branches of the lexical tree in parallel. *Dec.*, 67. A POSITA would have understood a cluster of processors, as described in *Chen*, was usable and desirable in a speech recognition system with multiple processors, such as *Jiang*'s circuit, and would have been a low-cost, simple-to-implement improvement. Specifically, a POSITA would have recognized that the increased computing resources provided by *Chen*'s cluster of processors would have increased the speed of the system, enabling a reduced cost system that recognized speech at a faster rate and/or higher level of accuracy. *Dec.*, 67.

There would have been a reasonable expectation of success for adding *Chen*'s clusters to *Jiang*'s system because *Chen* and *Jiang* each contemplates a hardware architecture using multiple signal processors; therefore, the more efficient processing architecture of *Chen* would have been simply substituted for *Jiang*'s architecture to achieve a compatible improvement in processing speed and power. *Dec.*, 67. Given the ubiquitous use of pluralities of processors in speech recognition circuits, a POSITA would have had a reasonable expectation of success, without undue experimentation, in modifying *Jiang*'s circuit to use *Chen*'s clusters of processors. *Dec.*, 67. Additional reasons for the combination are provided in the mapping.

2. *Obviousness of Modifying Electronic Hardware and Software to Be Circuitry*

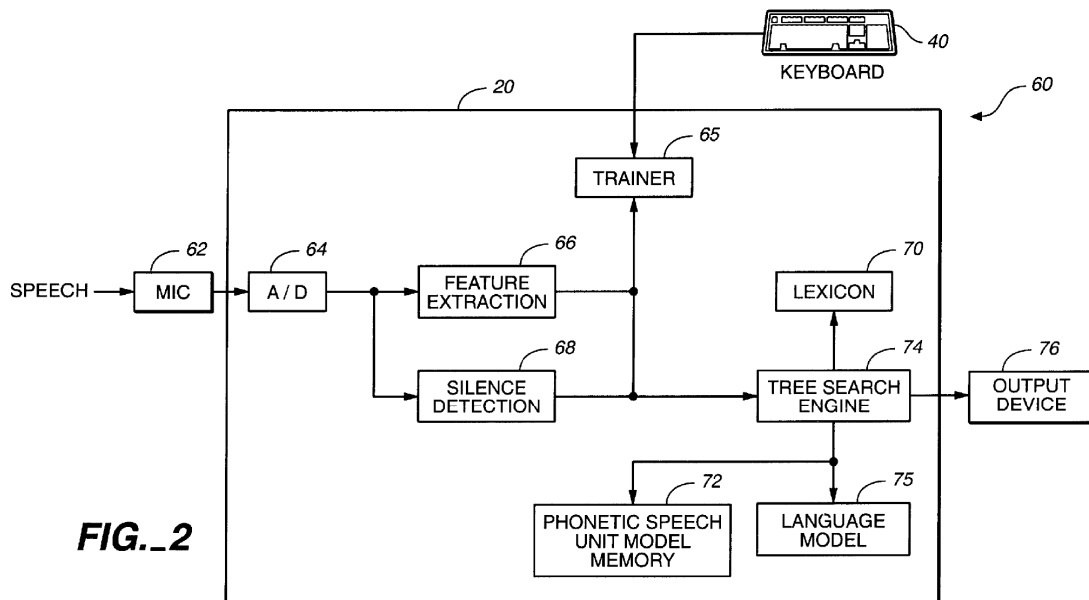
To the extent Patent Owner argues the various elements cited from *Jiang*, *Chen*, *Lucke*, *Robinson*, or *Wrench* must be embodied as application-specific “circuitry,” it would have been obvious to a POSITA to implement any of the computerized techniques or physical architectures described performing similar functions to the '140 Patent's claims using “circuitry.” *Dec.*, 69. A POSITA would have recognized each of the references independently describes computerized methods performed by electrical hardware components including computer processors, memory, a power source, and electrical connections therebetween, thus teaching or at least rendering obvious “circuitry” performing the recited methods.

Dec., 69. Specifically, to the extent any particular description refers to computerized methods stored as software, a POSITA would have physically implemented those computerized methods using electronic circuitry. *Dec.*, 69. Therefore, a POSITA would have found it obvious and been motivated to embody the various electronic components disclosed in the references as “circuitry” without undue experimentation. *Dec.*, 69.

B. Claim 1

1. Claim 1(Pre): “A speech recognition circuit comprising:”

To the extent the preamble is limiting, *Jiang* teaches or at least renders obvious a speech recognition circuit, as claimed. *Jiang*, 1:15-18, 6:18-61, 4:51-5:3, Abstract, Figs. 1, 2, 7; *Dec.*, 70. *Jiang* teaches a speech recognition system 60 that includes a microphone 62, A/D converter 64, and program modules 65-75.



Jiang, Fig. 2, 6:18-28. *Jiang* specifically teaches its speech recognition system may be implemented on a suitable computing environment such as the personal computer illustrated in Fig. 1.

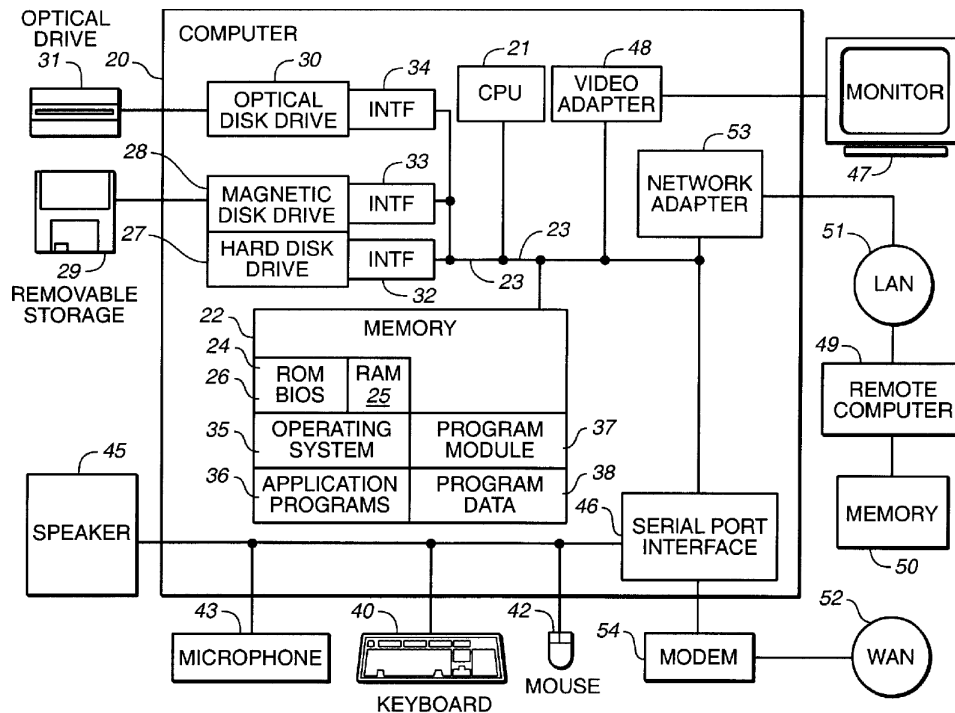


FIG. 1

Jiang, Fig. 1, 4:51-5:3.

Per Mr. Schmandt, a POSITA would have understood *Jiang*'s speech recognition system implemented on a computer teaches or at least renders obvious a "speech recognition circuit." See Section VI.B.2; *Dec.*, 70. *Jiang* already teaches "those skilled in the art will appreciate that the invention may be practiced with other computer system configurations...." *Jiang*, 4:60-62. Therefore, per Mr. Schmandt, given that it is generally obvious to configure computerized functions in circuitry,

and further given *Jiang* expressly teaches its invention may be employed in a variety of “computer system configurations,” it would have been obvious to a POSITA to implement *Jiang*’s speech recognition functionality as circuitry. *Dec.*, 70.

2. ***Claim 1(a): “one or more clusters of processors, each of the one or more clusters of processors comprising: a plurality of processors;”***

a) ***Jiang Teaches a Multiprocessor System***

Jiang in view of *Chen* teaches or at least renders obvious this limitation. *Jiang* teaches its speech recognition system may be implemented on various computing systems including specifically “**multiprocessor systems**, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, and the like.” *Jiang*, 4:51-5:3, *see also* 6:39-42 (“...CPU 21 [] may include one **or more** processors...”). However, *Jiang* does not specifically discuss *clusters* of processors. As discussed below in Section V.B.2.c, *Chen* teaches clusters of processors, with each cluster comprising a plurality of processors and memory. *Chen*, 9:5-43.

b) **The ’140 Patent’s Description of a “Cluster”**

The ’140 Patent describes a “cluster” as at least including processing architectures comprising a group of processors and a memory. *’140 Patent*, 3:14-19 (“Thus a group of lexical tree processors and a partial lexical memory form a cluster.”), 3:53-57, 5:14-21, Fig. 2. Computing systems arranged in a clustered

architecture were widely known and broadly used for a wide array of computing applications, including specifically speech recognition, prior to the '140 Patent. *Dec.*, 72 (citing examples of clustered speech recognition systems described by *Hon*, *Tremblay*, *Erman*, *Wulf*, *Garde*, and *Deo*). As discussed by *Hon*, computing platforms performing speech recognition, such as the MARS multiprocessor machine, benefitted from an architecture comprising a plurality of “clusters” of processing elements, providing desirably improved scalability and flexibility to speech recognition applications. *Hon*, pp. 24-25 (Ex. 1006); *Dec.*, 72 (further noting the link between *Jiang*’s “multiprocessor” system and *Hon*’s description of the MARS “clustered” system as “multiprocessor”).

c) *Chen* Teaches Claim 1(a)

An example of a clustered computing platform available in the art prior to the '140 Patent is described by *Chen*, teaching a “parallel processing computer system” comprising a plurality of “clusters” of processors. *Chen*, 9:5-19. As seen below in annotated Fig. 4, each cluster 100a-d includes two or more processors 102a-d “that are symmetrically connected to a cluster shared memory 104a via a connection node 106a.”

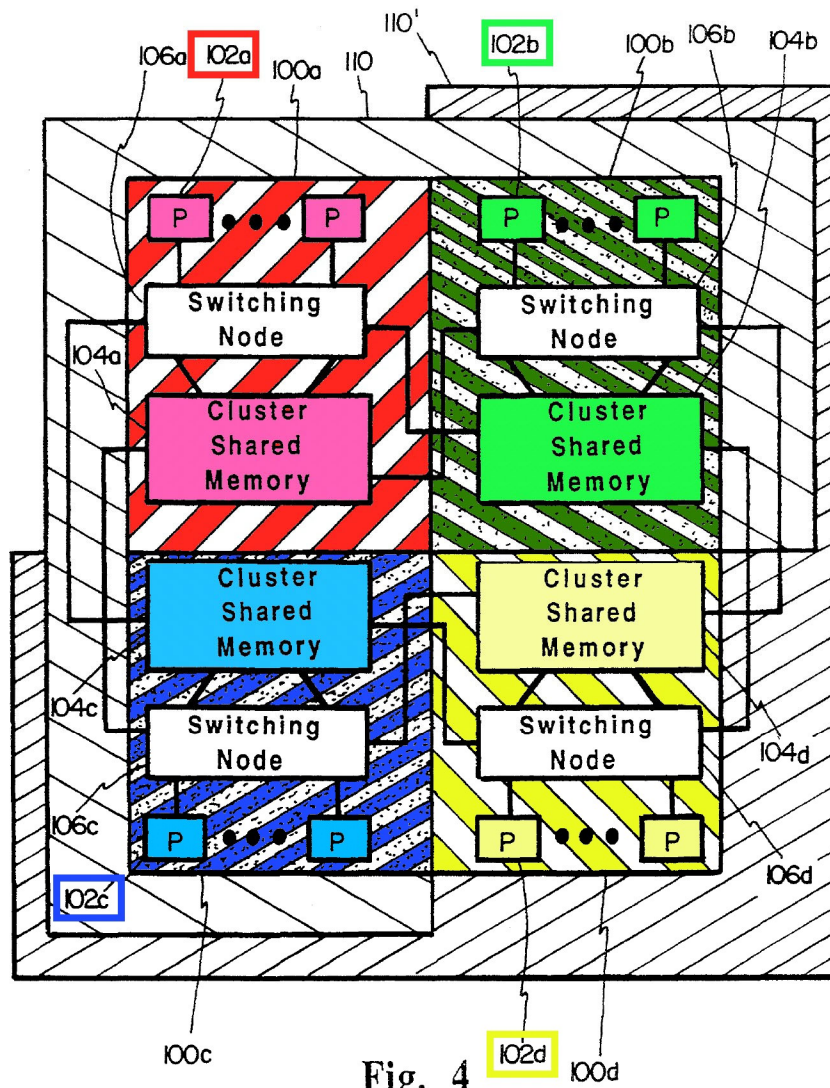


Fig. 4

Chen, Fig. 4 (annotated), 9:5-19; *Dec.*, 73.

The resulting computing architecture is described by *Chen* as “**an extendible number of clusters of multiple numbers of processors to create a high performance parallel processing computer system.**” *Chen*, 5:9-17. Thus, *Chen* teaches the claimed “one or more clusters of processors,” namely *Chen*’s clusters 100a, 100b, 100c, and 100d. *Chen*, 9:13-14, Fig. 4. *Chen* further teaches the claimed “each of the one or more clusters of processors comprising: a plurality of

processors[,]” namely clusters 100a-d each comprising its respective processors 102a, 102b, 102c, and 102d. *Chen*, Abstract, 5:9-17, 6:17-21, 9:5-43, 10:14-35 (envisioning that *Chen*’s clustered configuration may be formed using commercially available microprocessors), Figs. 4, 5A, 6A-B; *Dec.*, 73.

d) Motivation to Combine *Chen* with *Jiang* for Claim 1(a)

A POSITA would have been motivated and found it obvious to implement the speech recognition techniques taught by *Jiang* on the clustered processor computing platform of *Chen* to expedite the search process by leveraging shared memory amongst clustered processors. *Dec.*, 74. *Chen* expressly motivates the combination, stating that existing multiprocessor systems “might be retrofitted to take advantage of the unified parallel processing architecture of the present invention.” *Chen*, 10:21-25. Because *Jiang* is a multiprocessor system (*Jiang*, 4:60-65, 6:39-42), and because *Chen* expressly motivates **retrofitting** multiprocessor systems with *Chen*’s clustered processors, a POSITA would have been expressly motivated to modify *Jiang* as discussed herein.

Each of the one or more clusters in the modified *Jiang-Chen* circuit would have been configured by a POSITA to include a plurality of processors described by *Chen* performing the techniques described regarding the tree search engine 74 described by *Jiang*, replicating the tree search engine’s functionality amongst each of the plurality of *Chen*’s processors. Specifically, a POSITA would have recognized

that utilizing a processor architecture comprising one or more clusters of processors as taught by *Chen* would have enabled the speech recognition circuit of *Jiang* to efficiently process state sequences of the HMMs in parallel. *Id.* Doing so would have resulted in improved memory management and parallelism, allowing the pruning threshold to be relaxed. *Dec.*, 74 (*citing Jiang*, 2:25-46, describing pruning thresholds, error rate, and the dependency of each on computational savings).

A POSITA would have recognized employing clustered processors in addition to *Jiang*'s envisioned silence node approach would have furthered the goal of *Jiang* of providing a speech recognition circuit that is both accurate and efficient. *Dec.*, 74. *Jiang* specifically explains that a pruning threshold must be calibrated to ensure the computation time required keeps pace with the rate at which new frames of speech data are received. *Jiang*, 2:19-46, 11:20-24; *Dec.*, 74. Particularly in large-vocabulary, continuous recognition systems such as *Jiang*'s, limited computational resources require the pruning threshold to be set at a high level, reducing the search space at the cost of an increased error rate. *Jiang*, 2:19-46, 11:20-24; *Dec.*, 74. Thus, a POSITA would have recognized that improving *Jiang*'s computational resources would have desirably improved its speech recognition accuracy. *Id.*

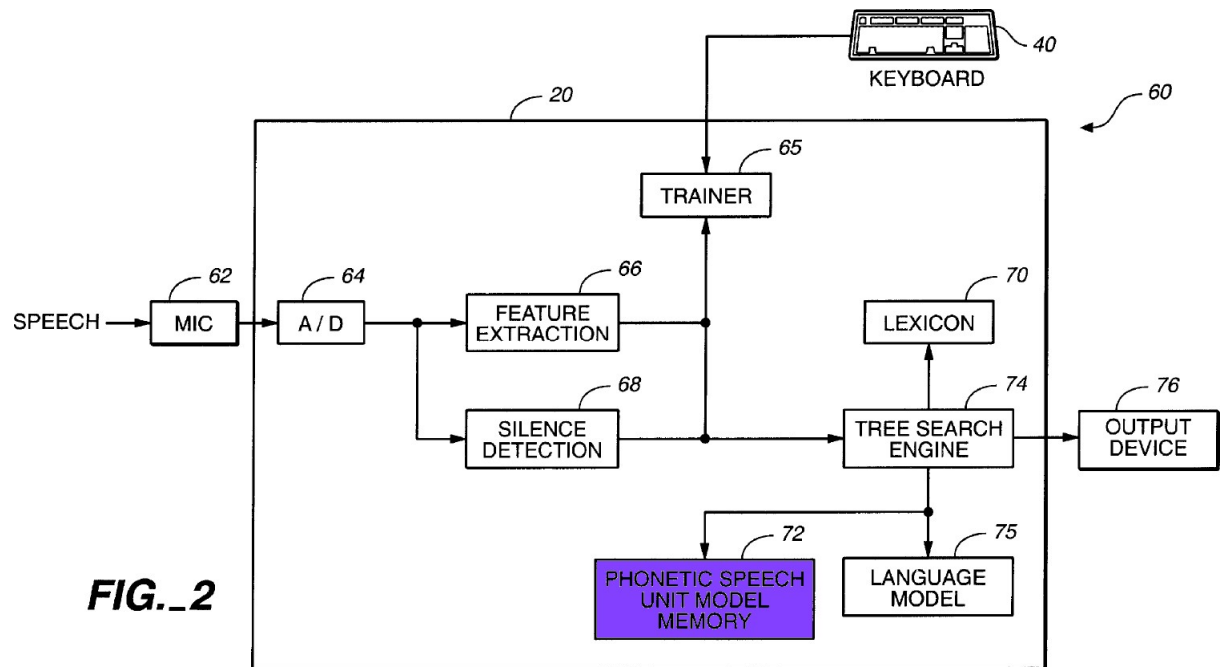
As a result of the relaxed pruning threshold enabled by *Chen*'s hardware, *Jiang*'s speech recognition system would have yielded a decreased error rate

resulting from over-pruning, improving the accuracy of the system. *Id.* Further, a POSITA would have recognized both *Chen* and *Jiang* contemplate a hardware architecture using multiple signal processors; therefore, the more efficient processing architecture of *Chen* would have been simply substituted for *Jiang*'s to achieve a compatible improvement in processing speed and power, leading a POSITA to a reasonable expectation of success for the combination. *Dec.*, 74. A POSITA would have been further motivated by *Hon*'s express praise for the scalability and flexibility provided by clustered-processing computer platforms to speech recognition tasks. *Dec.*, 71-72 (*citing Hon*, pp. 24-25, (Ex. 1006)).

3. ***Claim 1(b): “and [one or more clusters of processors, each of the one or more clusters of processors comprising:] an acoustic model memory storing acoustic model data”***

a) ***Jiang Teaches an Acoustic Model Memory Storing Acoustic Model Data***

Jiang in view of *Chen* teaches Claim 1(b). *Jiang* teaches an acoustic model memory, namely a “model memory” 72 (also referred to as “Phonetic Speech Unit Model Memory” in Fig. 2), storing acoustic model data, namely “phonetic speech unit models” such as Hidden Markov Models (HMMs). *Jiang*, 7:29-54, 6:18-28.



Jiang, Fig. 2 (annotated).

Jiang teaches the “phonetic speech unit models” (i.e., HMMs) are stored in speech unit model memory 72:

... tree search engine 74 accesses information stored in the phonetic speech unit model memory 72. **Memory 72 stores phonetic speech unit models, such as hidden Markov models**, which represent speech units to be detected by system 60. In one preferred embodiment, the phonetic models stored in memory 72 include HMMs which represent phonemes. Based upon the **HMMs stored in memory 72**, tree search engine 74 determines a most likely phoneme represented by the codeword received from feature extraction module 66, and hence representative of the utterance received by the user of the system.

Jiang, 7:32-43, 2:53-60 (referring to an HMM as a “phonetic model” when the HMM is modeling phonemes), 1:43-44 (“HMMs have been employed to model

observes sequences of speech spectra”), 1:19-21 (“the most successful current speech recognition systems employ probabilistic models known as hidden Markov models (HMMs).”). Per *Jiang*, a phoneme is a spoken sub-word unit of speech. *Jiang*, 1:50-53, 2:53-56. *Jiang* characterizes speech recognition processing via HMMs as utilizing “acoustic” signals and vectors. *Jiang*, 1:58–2:15. Thus, *Jiang* teaches that HMMs are phonetic speech unit models, and as such, a POSITA would have recognized that HMMs teach “acoustic models,” as claimed. *Dec.*, 75.

Additionally, per Mr. Schmandt, HMMs were well known at the ’140 Patent’s priority date for representing an “acoustic” model. *Dec.*, 75. In *Ravishankar*, HMMs are described as “a set of states connected by transitions” expressly characterized as an “acoustic model.” *Ravishankar*, 28-29, 38-40, 41 (“Acoustic modelling in Sphinx-II is based on hidden Markov models...”). Similarly, *Jelinek* teaches the “usual acoustic model employed in speech recognizers” is “the *hidden Markov model*,” while *Schmandt* teaches “Hidden Markov Models are a powerful representation of speech and consequently are used in many speech recognition systems currently under development” and may be used to “represent not only the phones of each word but also the probabilities of transitioning from one word to another.” *Jelinek*, 21 (Ex. 1011); *Schmandt*, 60 (Ex. 1012); *Dec.*, 75 (noting the ’140 Patent, 1:30-44, 6:41-48 describes HMMs as representing acoustic models and that

the “acoustic model memory 23 stores a Hidden Markov Model for acoustically modelling words as lexical trees.”).

Therefore, *Jiang* teaches or renders obvious “an acoustic model memory storing acoustic model data,” as claimed, namely “phonetic speech unit model memory 72” storing phonetic models including HMMs representing phonemes. *Jiang*, 7:35-38.

b) *Chen* Teaches One or More Clusters of Processors Each Comprising Memory

As discussed for Claim 1(a) (Section V.B.2.c), *Jiang*’s multiprocessor system as modified to include *Chen*’s clusters of processors teaches a speech recognition circuit comprising one or more clusters of processors. Additionally, *Chen* teaches each cluster includes a memory. Specifically, each cluster 100a-d includes two or more processors 102a-d that are symmetrically connected to a cluster shared memory 104a-d via connection nodes 106a-d, as illustrated in annotated Fig. 4 below:

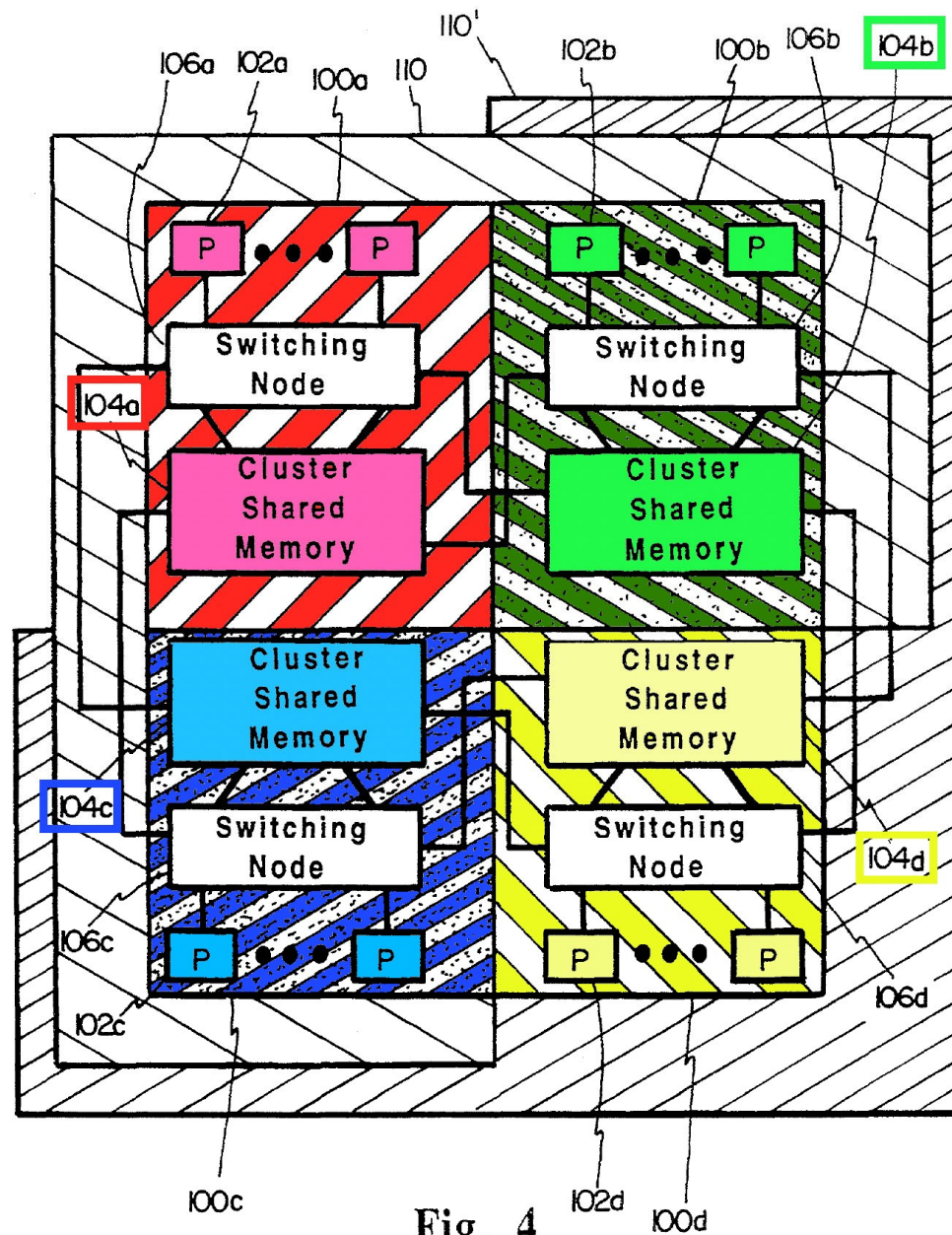


Fig. 4

Chen, Fig. 4 (annotated), 9:5-19; *Dec.*, 76. The resulting computing architecture described by *Chen* provides multiple processors in “physically separable clusters, each cluster having a common cluster shared memory that is symmetrically accessible by all of the processors in that cluster.” *Chen*, 5:9-17. Thus, *Chen* teaches one or more clusters of processors (as mapped for Claim 1(a)), each of the one or

more clusters of processors comprising a memory (the respective cluster shared memory 104). *Chen*, Abstract, 5:9-17, 6:17-21, 9:5-43, 10:14-35 (envisioning that *Chen's* clustered configuration may be formed using commercially available memory), Figs. 4, 5A, 6A-B; *Dec.*, 76.

c) Motivation to Combine *Chen* with *Jiang* for Claim 1(b)

As discussed in Section V.B.2.d, a POSITA would have found it obvious and been motivated to implement the speech recognition techniques taught by *Jiang* using the clustered processor computing platform of *Chen* to expedite the search process by leveraging shared memory amongst clustered processors. *Dec.*, 76. Specifically, a POSITA would have stored at least a portion of the acoustic model data (HMM acoustic model from *Jiang's* model memory 72) in the shared cluster memories 104a-d of each of the clusters 100a-d taught by *Chen* for access by processors 102a-d during speech recognition, as will be further discussed below for Claim 1(c). *Dec.*, 76.

The modification to *Jiang* to include multiple clusters of processors each with a memory (as taught by *Chen*) would have reduced the time required to recognize a user's speech, speeding the system's response and enabling a more relaxed pruning threshold. *Dec.*, 74 (*citing Jiang*, 2:25-46, describing pruning thresholds, error rate, and the dependency of each on computational settings). As a result of the more relaxed pruning threshold enabled by *Chen's* hardware, *Jiang's* speech recognition

system would have yielded a decreased error rate, improving the accuracy of the system. *Id.* Further, a POSITA would have recognized both *Chen* and *Jiang* contemplate a hardware architecture using multiple signal processors. Therefore, the more efficient processing architecture of *Chen* would have been simply substituted for *Jiang*'s to achieve a compatible improvement in processing speed and power, leading to a reasonable expectation of success for the combination. *Dec.*, 74. A POSITA would have been further motivated by *Hon*'s express praise for the scalability and flexibility provided by clustered-processing computer platforms to speech recognition tasks. *Dec.*, 72 (citing *Hon*, 24-25).

4. Claim 1(c): “wherein each of the plurality of processors is configured to compute a probability using the acoustic model data in the acoustic model memory”

Jiang in combination with *Chen* teaches or renders obvious Claim 1(c).

a) *Jiang* Teaches “comput[ing] a probability using the acoustic model data”

The term “probability” or its variants is only used in claims 1, 2, and 4 of the '140 Patent. Claim 2 recites that the “probability is an input to an evaluation of a state transition of a model of states[,]” where, per claim 3, the “model” at least includes an HMM.

Jiang teaches computing a probability using the acoustic model data (i.e., the HMM data). As discussed below, *Jiang* teaches computing a “probability score” for each state of an HMM. Specifically, *Jiang* teaches speech recognition circuitry

performing a process “to determine a most likely state sequence for which the overall probability is maximum, given the transition probabilities between states and the observation probabilities.” *Jiang*, 1:28-32, 1:33-39. *Jiang* describes its process as follows:

In order to find the most likely sequence of states corresponding to a **sequence of acoustic vectors**, the Viterbi algorithm is employed. The Viterbi algorithm performs a computation which starts at the first frame and proceeds one frame at a time, in a time-synchronous manner. **A probability score is computed** for each state in the state sequences (i.e., the HMMs) being considered. Therefore, a cumulative probability score is successively computed for each of the possible state sequences as the Viterbi algorithm analyzes the acoustic signal frame by frame. By the end of an utterance, **the state sequence (or HMM or series of HMMs) having the highest probability score computed by the Viterbi algorithm provides the most likely state sequence for the entire utterance.** The most likely state sequence is then converted into a corresponding spoken subword unit, word, or word sequence.”

Jiang, 2:1-15; *Dec.*, 77.

Jiang additionally teaches that a most likely phoneme is determined based on the HMMs stored in the phonetic speech unit model memory 72:

Upon receiving the codewords from feature extraction module 66, and the boundary detection signal provided by silence detection module 68, tree search engine 74 **accesses information stored in the phonetic speech unit model memory 72. Memory 72 stores phonetic speech**

unit models, such as hidden Markov models, which represent speech units to be detected by system 60. In one preferred embodiment, the **phonetic models stored in memory 72 include HMMs which represent phonemes. Based upon the HMMs stored in memory 72, tree search engine 74 determines a most likely phoneme represented by the codeword received from feature extraction module 66, and hence representative of the utterance received by the user of the system.**

Jiang, 7:30-44. *Jiang* thus teaches determining a most likely phoneme based on the stored HMMs. *Dec.*, 78-82.

Per Mr. Schmandt, *Jiang* teaches or renders obvious that this “most likely phoneme” is determined by computing probability scores for each HMM state, as discussed in *Jiang*, 2:1-15. *Id.* Mr. Schmandt opines that a POSITA reading the collective teachings of *Jiang* would have understood that the most likely phoneme determined from the HMMs is determined using the probability score computations discussed at *Jiang*, 2:1-15. *Jiang* teaches that the determination of the most likely phoneme (*Jiang*, 7:30-45) and use of the HMMs in a prefix tree for determining the word (8:16-52) is “all done in a known manner.” *Jiang*, 8:51. Per Mr. Schmandt, given that HMMs were widely used as of the ’140 Patent’s priority date for computing a probability of an utterance, such as a phoneme, and further given that *Jiang* describes in detail computing a probability score using HMMs, *Jiang* teaches computing a probability (i.e., the *Jiang* probability score) using the acoustic model

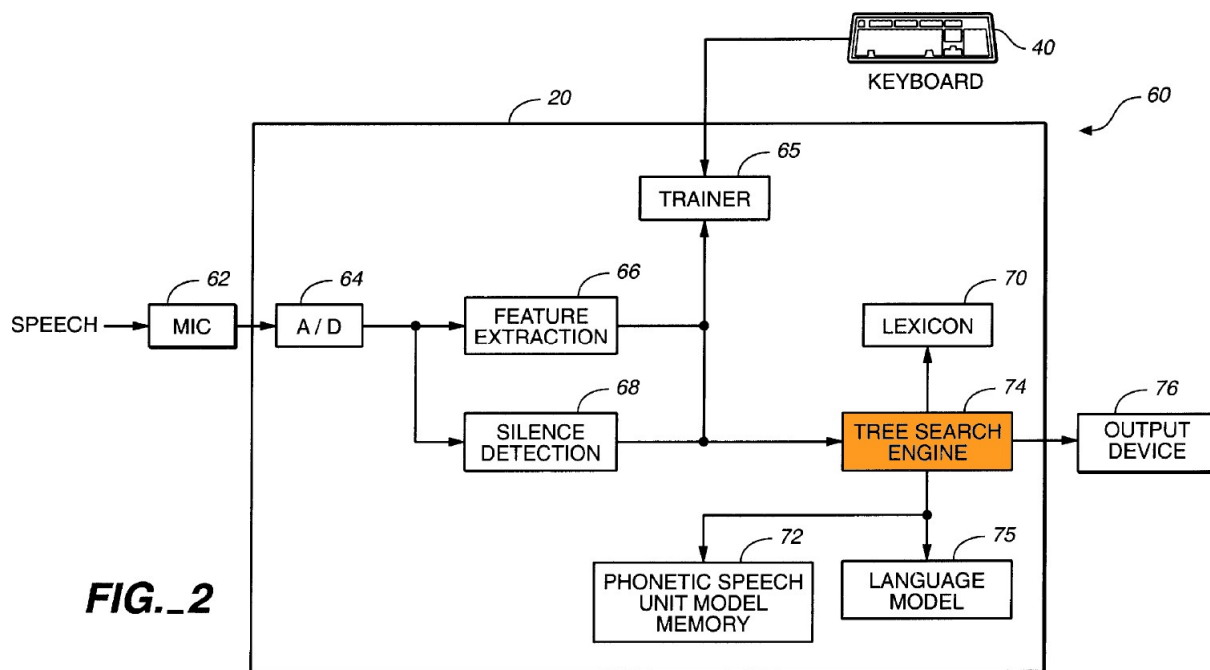
data (i.e., the *Jiang* HMMs) in the acoustic model memory (i.e., *Jiang*'s phonetic speech unit model memory 72). *Dec.*, 78-82 (citing *Ravishankar*, 28, 33 (Ex. 1007)). Claim 2 of the '140 Patent further explains the claimed "probability." *See* '140 *Patent*, Claim 2; Section V.C (Mapping for Claim 2). Like the '140 Patent, *Jiang* teaches an HMM as a model of states (*see* Claim 3 of the '140 Patent), where the probability for each state transition is computed. *Jiang*, 1:19-33 (discussing transition probabilities for each state).

b) *Jiang* in Combination with *Chen* Teaches "each of the plurality of processors [is configured to compute] ... using the acoustic model data in the acoustic model memory"

As mapped in Section V.B.4.a, *Jiang* teaches computing a probability score for each state of the HMM. *Jiang* further teaches a "processor," as claimed, performs the probability score computation using the acoustic model data, i.e., the HMM data, stored in the acoustic model memory 72.

Jiang explains that probabilities determined from extracted feature vectors are provided to a tree search engine 74 using HMM information stored in the phonetic speech unit model memory 72 (the mapped "acoustic model data in the acoustic model memory"). *Jiang*, 7:29-54. "**Memory 72** stores phonetic speech unit models, such as **hidden Markov models**, which represent speech units to be detected by system 60." *Jiang*, 7:33-36; *see also* Section V.B.3. "Based upon the HMMs stored in memory 72, tree search engine 74 determines a most likely phoneme represented

by the” series of extracted feature vectors, and “hence representative of the utterance received by the user of the system.” *Jiang*, 7:38-43. The “tree search engine 74 is preferably implemented in CPU 21 (**which may include one or more processors**) or may be performed by **a dedicated speech recognition processor** employed by personal computer 20.” *Jiang*, 6:39-42, Figs. 1-2; *Dec.*, 77. Fig. 2 illustrating computer 20 having tree search engine 74 is provided below:



Jiang, Fig. 2 (annotated).

Jiang thus teaches that tree search engine 74 (1) accesses HMM information stored in phonetic speech unit model memory 72 (*Jiang*, 7:30-36); and (2) computes a probability (*Jiang*, 2:1-15, 7:30-42, 8:44-51). Therefore, tree search engine 74, which is implemented in CPU including a processor, performs a probability

computation. *Jiang* thus teaches a “processor” configured to compute a probability using stored acoustic data, as claimed. *Dec.*, 77.

Jiang as modified to include *Chen*’s clustered processing further teaches or at least renders obvious that “**each** of the plurality of processors” is configured to compute a probability. Specifically, *Chen* teaches that for each cluster of processors, the plurality of processors in each cluster **may perform processing in parallel, such that each processor of each of the clusters** performs the intended operation. *Chen*, 5:10-17, 18:64-19:9; *Dec.*, 83. Once *Jiang* is modified, a POSITA would have recognized that *each* of the plurality of processors taught by the modified *Jiang* and (as discussed above regarding Claim 1(a)) provided in each cluster as modified by *Chen* would have been utilized to perform the probability computations expressly taught by *Jiang* for its single-processor embodiment. *Dec.*, 83. Specifically, in the proposed modified system, each of *Chen*’s processors 102a-d in respective clusters 100a-d would have computed probability scores as taught by *Jiang* using the acoustic model data stored in shared cluster memories 104a-d of each of the clusters 100a-d, as discussed above regarding Claim 1(b). *Id.* A POSITA would have recognized it would have been advantageously efficient to use each of the clustered processors of *Jiang-Chen* to perform speech recognition. *Id.* Thus, the speech recognition circuit taught by *Jiang* in view of *Chen* teaches *each* of a plurality of

processors is configured to compute a probability using acoustic model data in the acoustic model memory, as claimed.

c) Motivation to Combine *Chen* with *Jiang* for Claim 1(c)

A POSITA would have been motivated to modify *Jiang* to include *Chen*'s clustered processors, as discussed above for Claims 1(a) and 1(b). As discussed, *Jiang* teaches a multiprocessor speech recognition system in which a lexical tree search is used to recognize spoken words. *See* Section VI.C.1. *Jiang* modified by *Chen* teaches or renders obvious an HMM acoustic model, accessible to its plurality of processors, used to perform a tree search to determine a most likely phoneme represented by features extracted from a user's utterance. *Jiang*, 7:30-47.

5. Claim 1(d): “[wherein] the speech recognition circuit is configured to generate an initial score for an audio sample”

Jiang teaches this limitation. The '140 Patent describes the “initial score” at least includes “scores for paths being processed” and later used to “determine whether a global pruning strategy should be applied” based on a pruning threshold or the best scores of other paths. '140 Patent, 6:21-35, 8:59–9:4; Dec., 86; Section VI.C.6 (discussing the '140 Patent's File History).

a) *Jiang* Teaches an “audio sample”

Jiang teaches an “audio sample,” as claimed, namely a digital sample from an acoustic signal. *Jiang*, 1:58-67. *Jiang* teaches that a microphone provides an analog system to an A/D converter, which “samples” the signal at a rate of approximately

16kHz into “a plurality of digital samples.” *Jiang*, 6:46–7:1. After further processing, the sampled signal is then provided to a “tree search engine” 74 as a plurality of “codewords” for comparison with the acoustic model stored in memory 72. *Jiang*, 7:16-47. A tree search is then performed to determine the word hypotheses best represented by the codewords. *Jiang*, 7:5-15, 2:1-15, 10:1-11; *Dec.*, 78-82, 85.

Per Mr. Schmandt, the acoustic signal output by the microphone, sampled by the A/D converter, divided into frames comprising a plurality of digital samples by the feature extraction module 66, and encoded into feature vectors (or later, codewords) by the feature extraction module 66, is an “audio sample,” as claimed. *Dec.*, 85. The feature vectors (or codewords) are provided to tree search engine 74 for scoring using the HMMs, as further discussed in the following section. *Jiang*, 7:29-47, 8:16-24; *Dec.*, 85, 87.

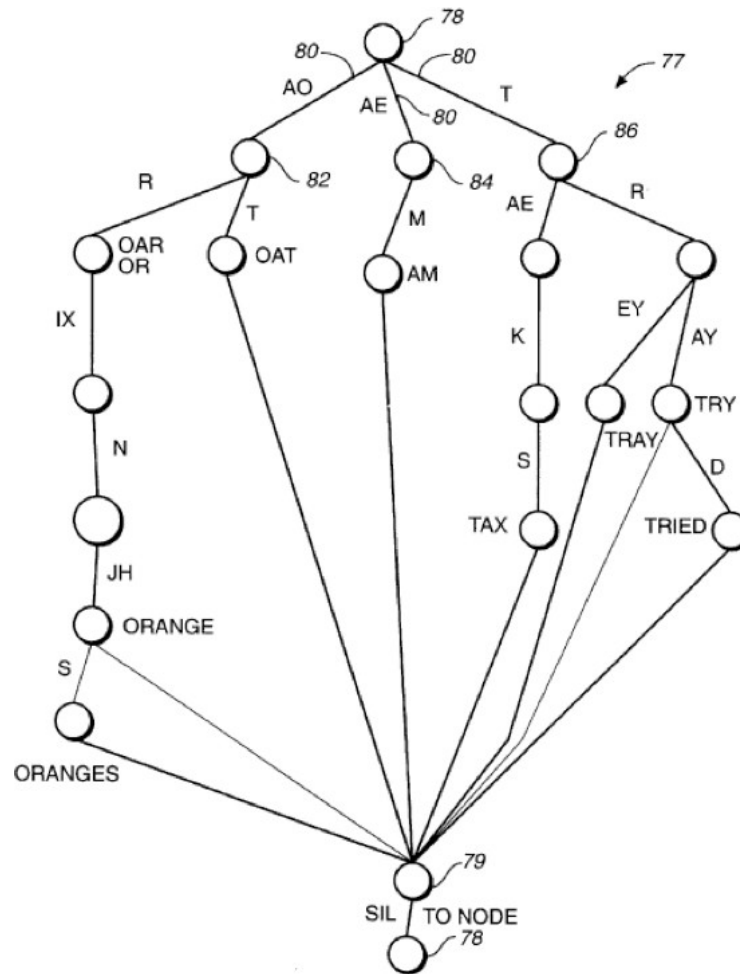
b) *Jiang* Teaches “the speech recognition circuit is configured to generate an initial score”

As discussed above, *Jiang* teaches computing *probabilities* to determine the most likely phoneme. *Jiang*, 7:30-45, 2:1-15, 1:19-33. Based on the accessed phonetic speech unit models used to determine the most likely phoneme, “a word which most likely represents the codewords received by feature extraction module” is determined using a lexical (or prefix) tree. *Jiang*, 7:49-54, 8:16-43; *Dec.*, 84 (*citing Ravishankar*, 33, which teaches “[o]rganizing the HMMs to be searched as a phonetic tree...is probably the most often cited improvement in search techniques

in use currently”). As discussed at *Jiang*, 8:5-43, a prefix tree is traversed to identify the word, where the possible words are constrained by the lexicon 70, and where the tree comprises a plurality of phoneme branches corresponding to the constrained set of words from the lexicon. *Jiang*, 8:5-15, 7:48-53 and 7:65–8:4 (discussing lexicon 70), 8:5-15 (discussing prefix tree), Fig. 3. As the tree is traversed, a **score** is determined for each node and associated phoneme branch to determine the likelihood the phoneme being examined is the phoneme indicated by the codeword. *Jiang*, 8:16-24; *Dec.*, 87 (noting *Jiang*’s similarity to *Ravishankar*, 33). Therefore, as the speech is input, sampled, and feature vectors extracted, the tree search engine traverses the prefix tree, computing a score based on the HMMs. *Dec.*, 87.

Jiang thus teaches the extracted feature vectors or codewords (i.e., the claimed “audio sample”) are provided to a tree search engine 74 configured to search a lexical tree (*Jiang*’s prefix tree) to determine the most likely string of phonemes spoken, ultimately resulting in a determination of a user’s utterance. *Jiang*, 7:29-54; *Dec.*, 87; *see also* Section VI.C.4. *Jiang*’s tree organization is shown below in Fig. 3, which illustrates an exemplary subset of words stemming from the initial phonemes (after silence node 78) of AO, AE, and T²:

² The discussion of *Jiang*’s lexical tree is presented with reference to Fig. 3, though some cited portions of *Jiang* refer to Figs. 4-6. In the mapping below, the trees



Jiang, Fig. 3, 8:5-43; *Dec.*, 87. *Jiang* explains that in searching for the most probable word uttered, “as tree 77 is traversed from the input node 78 to the output node 79, **a score is assigned to each node** connected to a phoneme branch then under consideration by the speech recognition system. **The score is indicative of a likelihood that the particular phoneme being examined is the actual phoneme**

illustrated in Figs. 3-6 and descriptions thereof are appropriately discussed as a single concept. Dec., 84.

indicated by the codeword received from the feature extraction module 66” in what *Jiang* characterizes as a “known manner.” *Jiang*, 8:16-24, 8:44-51.

Jiang provides a detailed example in which the disclosed circuit is recognizing the word “ORANGE”:

For example, if the word ORANGE were input into system 60 by the user, feature extraction module 66 would likely divide the word ORANGE into codewords indicative of the phonemes represented as follows: AO, R, IX, N, JH. **As the tree search engine traverses tree 77, it preferably computes a score, for each phoneme branch considered in tree 77, wherein the score represents the likelihood that the particular phoneme encoded by the codeword corresponds to the phoneme for the branch under consideration.** Thus, tree search engine 74 **computes a score for node 82** which indicates that the first codeword under consideration is highly likely to be represented by the AO phoneme corresponding to the branch under consideration. **Tree search engine 74 also preferably computes a score for each of the other nodes 84 and 86 in tree 77,** wherein the score is indicative of the likelihood that the codeword being analyzed is represented by the phonemes AE, and T. Under ideal circumstances, the score assigned to nodes 84 and 86 is lower than the score assigned to node 82.

Jiang, 8:25-43.

Per Mr. Schmandt, *Jiang* teaches computing a **score** at each node and associated phoneme branch (i.e., computing a score for node 82, along with computing scores for each of nodes 84, 86 in tree 77). *Dec.*, 87; *Jiang*, 8:29-37. Node

82 is determined to have the highest score, while branches stemming from nodes with sufficiently lower scores may be “pruned” from further searching, as will be discussed in further detail for Claim 1(e). *Jiang*, 8:44-64. As Mr. Schmandt explains, because *Jiang* teaches sampled audio used to generate scores as the tree is traversed, tree search engine 74 “generates an initial score” for an audio sample as claimed. *Dec.*, 87.

To the extent this “initial score” must specifically be the *first* score generated for the audio sample, *Jiang* teaches or at least renders obvious such. *Jiang* expressly describes calculating scores for each of the initial branches 82, 84, and 86 of tree 77 compared to the “first codeword under consideration.” *Jiang*, 8:35-43, Fig. 3.

6. ***Claim 1(e): “[wherein] the initial score is used to determine whether to continue processing to determine a final score via processing a larger amount of model data than that was processed to generate the initial score”***

a) **Prosecution History of the ’140 Patent Regarding Claim 1(e)**

During prosecution of the ’140 Patent, the Examiner issued a § 112(a), written description rejection, stating “the limitation, where the speech recognition circuit is configured to generate an initial score for an audio sample used to determine whether to continue processing to determine a final score via processing a larger amount of model data than that was processed to generate the initial score is **new matter**.” ’140

File History, 89. Applicant responded that the specification provided support for this limitation via its description of standard pruning threshold techniques, stating:

Here, one of ordinary skill would conclude that Applicant's application clearly supports the full scope of the claims and there is possession of the claimed invention. For example, **Figure 4 of Applicant's Specification illustrates a "process performed by a lexical tree processor to determine a temporary lexical tree score."** Applicant's As-filed Specification at 11 (emphasis added). *See also* FIG. 4; Applicant's As-filed Specification at 11-12 (describing the process for determining a temporary lexical tree score for a lexical tree in connection with Figure 4); FIG. 6 (illustrating steps S42 for temporary lexical tree score determination). **The temporary lexical tree score "can be used to prune out lexical trees from the processing which have a low score."** Applicant's As-filed Specification at 11 (emphasis added). **In other words, the process for processing the lexical tree branches having a temporary lexical tree score lower than a threshold would not continue to determine a "final score via processing a larger amount of model data than that was processed to generate the initial score,"** as originally recited in claim 10 (now canceled; the subject matter of which has been incorporated in amended independent claim 1). *See, e.g.*, Applicant's As-filed Specification at 13 (**"Pruning** is applied to the lexical tree by the lexical tree processor to delete paths in order to keep the breadth of the search manageable The **pruning** applied locally by the lexical tree processor can be purely on the basis of a local **threshold**").

'140 *File History*, 104 (Ex. 1002); *see also* '140 Patent, 7:46-8:44, 8:64–9:4, Figs. 4, 6; *Dec.*, 86. The Examiner found this support sufficient, issuing a Notice of Allowance. '140 *File History*, 120. Thus, Applicant contended that pruning branches lower than a threshold supports, and thus teaches, Claim 1(e).

b) *Jiang* Teaches Claim 1(e)

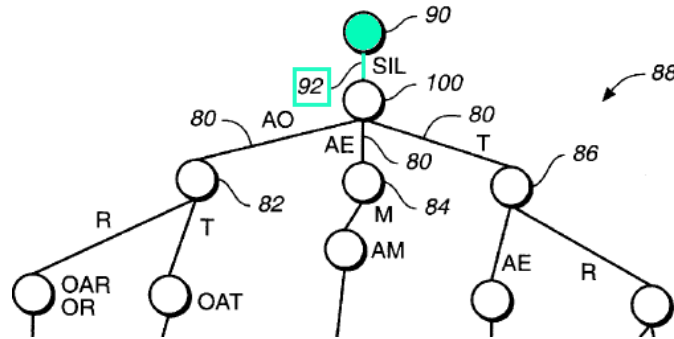
Jiang teaches or renders obvious Claim 1(e). In *Jiang*, the prefix tree comprises a plurality of nodes connected by phoneme branches, with each node and associated branch representing a phoneme, the plurality of nodes of the tree representing a word to be recognized, and the end node representing the leaf or recognized word. *Jiang*, 2:53-67, 4:8-12, 7:48–8:4 (discussing traversing the tree from root to leaf), 11:36-43 (discussing a training module for generating the HMMs and that the prefix tree is structured to provide a leaf associated with a word). A score is assigned to each node of the prefix tree as the tree is traversed. *Jiang*, 8:16-51. As discussed below, these initial scores are used to determine, via pruning, whether to continue processing branches. When the prefix tree search terminates at an end (“leaf”) node, a final score representing a final determined word is output as the recognized spoken utterance (such as “ORANGE” in the example described at 8:16-64).

Jiang teaches a “pruning technique” employed during its lexical tree search accomplished by

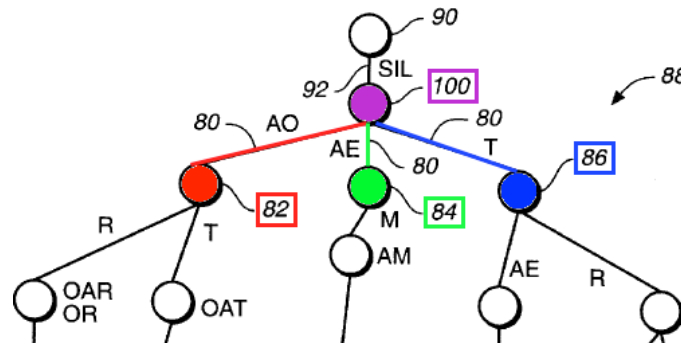
[A] pruning technique can be used. Pruning is accomplished by **comparing, at each node, the scores assigned to that node with the largest score on any of the other nodes** corresponding to the frame being considered. **If the score at a particular node is sufficiently low** compared to the maximum score for other corresponding nodes in tree 77, it is assumed that the branches leading to the node under consideration (the node with the lower score) will not likely be part of the completed most likely phoneme sequence (i.e., the phonemes in that sequence will not likely be part of the final word recognized by the system). Thus, **that branch is dropped (or pruned) from tree 77 and is no longer considered in further processing.**

Jiang, 8:52-64, *see also* 2:25-46, 10:51-11:15; *Dec.*, 89. *Jiang* expressly characterizes the value below which branches may be pruned as a “minimum threshold value.” *Jiang*, 2:39-46, 10:65–11:15 (discussing a “predetermined threshold level” employed for pruning), 11:16-24 (discussing setting threshold value for pruning based on computation savings versus error rate).

Figures 3-6 of *Jiang* help illustrate *Jiang’s* teachings further. Figure 3 illustrates a prior art lexical tree. *Jiang*, 8:5-15, Figs. 3-6. *Jiang* is generally directed to improving efficiency by including silence nodes. *Jiang*, 4:7-20. Thus, Fig. 4 illustrates the tree of Fig. 3 with added silence nodes. *Jiang*, 9:28-48; *Dec.* 84. In Fig. 4, the tree begins with node 90 connected to silence branch 92, highlighted below. *Jiang*, 9:28-37.



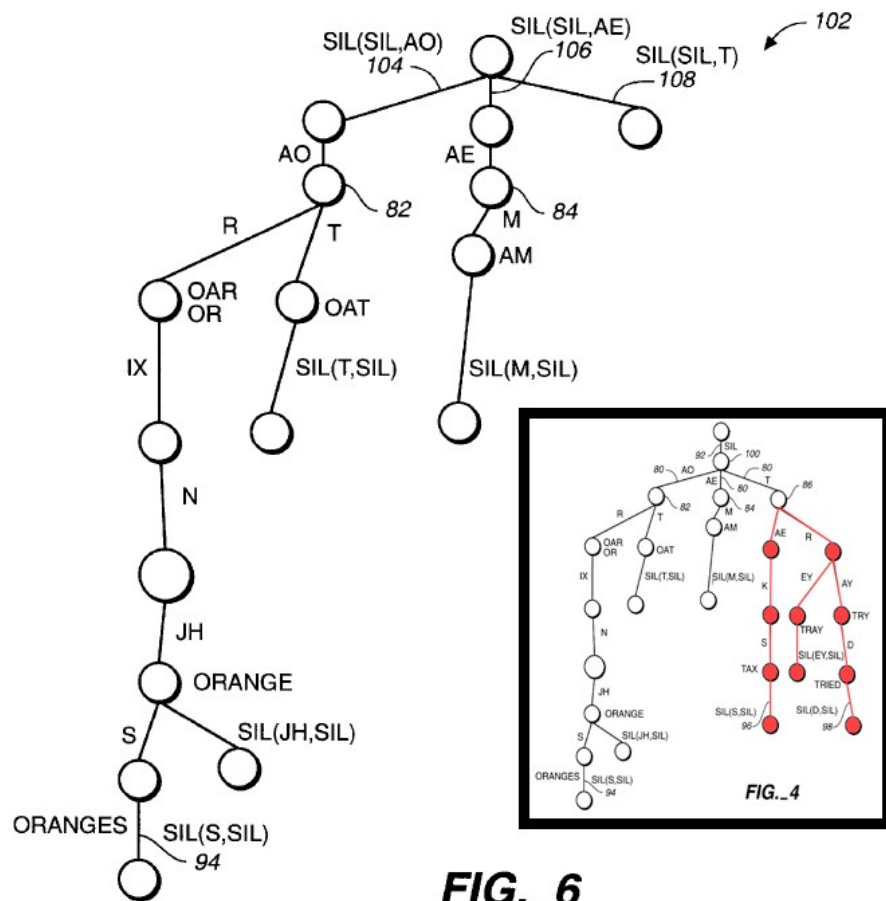
Jiang, Fig. 4 (excerpted and annotated). The top node 100 is connected to the *initial* branches for phonemes AO, AE, and T, which are respectively connected to nodes 82, 84, and 86.



Jiang, Fig. 4 (excerpted and annotated). As the tree is traversed, branches/nodes with lower scores are pruned:

If the score for the node being compared is less than the maximum score by the threshold level, all subsequent branches connected to that node are pruned from the tree, thereby drastically reducing the search space for any given search. FIG. 6 illustrates that the score assigned to the node for silence branch 108 is sufficiently low that the remainder of the branches of the tree have been pruned from the tree.

Jiang, 11:8-15; *Dec.*, 90 (discussing that Figs. 3-6 are related, where Fig. 6 is the Fig. 5 tree pruned to remove branches stemming from “T” (*Jiang*, 10:65-67); Fig. 5 is a standard prior art lexical tree including additional silence nodes (*Jiang*, 10:34-40, 9:43-47). The pruning is shown below in Fig. 6, where node 86—which had a lower score than node 82—and its subsequent branches are pruned from the tree:

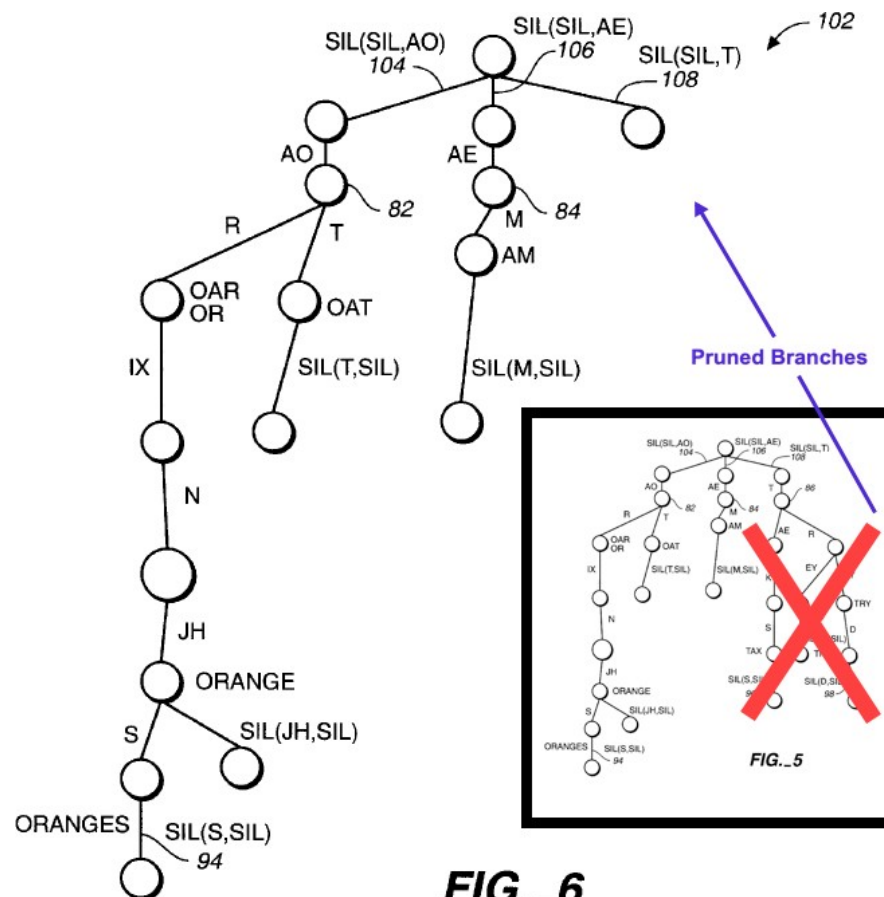


Jiang, Fig. 6 inset with Fig. 4 (annotated to illustrate pruned branches).

By comparing Fig. 4 to Fig. 6, a POSITA would have readily understood or found obvious *Jiang* teaches Claim 1(e). *Dec.*, 89-91. In Fig. 4, all of nodes 82, 84, and 86 with their subsequent branches are shown (ending in silence nodes 94, 96, 98). *Jiang* teaches that to obtain “computational savings” (*Jiang*, 11:21-24), a pruning technique is applied at the first level based on an empirically determined threshold. Any nodes with a score below the threshold are pruned: “all subsequent branches connected to that node are pruned from the tree.” *Jiang*, 11:8-15, 8:52-64. In Fig. 6, the first level are the silence nodes 104, 106, 108, and the pruning is based off the initial scores for 104, 106, 108. *Jiang*, 11:12-15; *compare with Jiang*, 11:16-17 (noting that the pruning decision need not be made at the first level).

In Fig. 6 of *Jiang*, the score for the “first level” of nodes 104, 106, 108 would have been understood by a POSITA to be an “initial score.” *Dec.*, 90; *Jiang* 10:53-62 (discussing the tree traversal “beginning” by assigning scores to nodes 104, 106, 108); 10:67–11:3 (discussing assigning scores to each of nodes 104, 106, 108). This **initial score** for nodes 104, 106, 108 (i.e., the nodes corresponding to possible first phonemes for the word “ORANGE”) is used to determined what branches to continue processing to determine a **final score for a final node** (i.e., the node corresponding to the word “ORANGE” or “ORANGES” in Fig 6). *Dec.*, 90; *Jiang*, 11:8-12, 8:52-64.

Because the initial score is used to determine which branches to prune (i.e., nodes and corresponding branches below the threshold level), then a final score is determined via processing a larger amount of model data. *Dec.*, 90-91. For example, for node 104, the branches associated with R, IX, N, and JH are maintained and used to determine a final score. *Dec.*, 90. In contrast, the branches stemming from node 108 are pruned because node 108 is below the threshold. *Jiang*, 11:12-15. Thus, the branches stemming from node 108, i.e., AE, R, K, EY, etc., are no longer processed to generate a final score. *Dec.*, 90.



Jiang, Fig. 6 inset with Fig. 5 (annotated to illustrate pruned branches).

Per Mr. Schmandt, *Jiang* thus teaches an initial score (e.g., the score for nodes 104, 106, 108) used to determine whether to continue processing to determine a final score, where the final score is determined via processing a larger amount of data (i.e., more nodes/branches, such as branches R, IX, N, etc. stemming from node 104) than was processed to generate the initial score (i.e., node/branch 104 only). *Dec.*, 90-91. As such, *Jiang* teaches claim 1(e), wherein branches with an initial score is used to determine whether to continue processing to determine a final score via processing a larger amount of model data than that was processed to generate the initial score.

C. Claim 2:

1. Claim 2: “...wherein the probability is an input to an evaluation of a state transition of a model of states”

Jiang teaches a plurality of processors computing probabilities using an HMM acoustic model. *See* Sections VI.C.3-VI.C.4. In using an HMM, probabilities are successively computed for a sequence of observed spectral states, “frame by frame.” *Jiang*, 1:19-25, 1:43-45, 2:1-7. As the probability for each frame is computed, it is successively input to an evaluation of a cumulative probability indicating that the sequence of observed states represents a particular corresponding subword unit, such as a phoneme. *Jiang*, 2:1-15, 1:41-48, 1:58-67; *Dec.*, 92. *Jiang* thus teaches “the

probability is an input to an evaluation of a state transition of a model of states[,]” as claimed and discussed further below. *Jiang*, 1:19-2:15; *Dec.*, 92.

In more detail, *Jiang* expressly teaches an HMM as including “a **plurality of states**, wherein a transition probability is defined for each **transition** from each state to every state, including transitions to the same state.” *Jiang*, 1:19-27. The prior art describes in detail the relationship between HMMs and states: “An HMM consists of a number of internal states; the model passes from an initial state to the final state as a step-by-step process generating an observable output at each step (**state transition**). For example, the states may correspond to phonemes in a word with the observable output corresponding to the presence or absence of a number of acoustic features.” *Schmandt*, 57-58; *Dec.*, 92. *Jiang* confirms this well-known understanding in the art, stating: “Consequently, HMMs have been employed to model observed sequences of speech spectra, where specific spectra are **probabilistically associated with a state in an HMM**. In other words, for a given observed sequence of speech spectra, there is a most likely sequence off states in a corresponding HMM.” *Jiang*, 1:43-48. *Jiang* further teaches:

In order to find the most likely sequence of states corresponding to a sequence of acoustic vectors, the Viterbi algorithm is employed. **The Viterbi algorithm performs a computation which starts at the first frame and proceeds one frame at a time**, in a time-synchronous manner. **A probability score is computed for each state** in the state

sequences (i.e., the HMMs) being considered. Therefore, **a cumulative probability score is successively computed** for each of the possible state sequences as the Viterbi algorithm analyzes the acoustic signal frame by frame. By the end of an utterance, the state sequence (or HMM or series of HMMs) having the highest probability score computed by the Viterbi algorithm provides the most likely state sequence for the entire utterance. **The most likely state sequence is then converted into a corresponding spoken subword unit**, word, or word sequence.

Jiang, 2:1-15; *Dec.*, 78-82, 92.

A POSITA would have recognized *Jiang*'s HMM acoustic model includes state transition data of its model of states. *Jiang*, 1:19-2:15; *Dec.*, 92. As quoted above, *Jiang* expressly teaches such in stating each state has a "transition probability." *Jiang*, 1:20-25. Further, *Jiang* specifies "the proper phoneme can be chosen...by examining the particular senones **calculated for each state of the HMMs** for each phoneme." *Jiang*, 7:38-46. Thus, because *Jiang* teaches a speech recognition process in which "calculated probabilities" (probabilities calculated frame by frame) are input to the evaluation of an HMM model including "transition probabilities for state transitions" to determine the most likely phoneme uttered, *Jiang* teaches "the probability is an input to an evaluation of a state transition of a model of states[,]" as claimed. *Jiang*, 1:19-27, 3:10-21 (expressly teaching "evaluating the probability" of phonemes), 7:29-46; *Dec.*, 92.

Additionally, *Jiang* explains that the HMM technique can be extended, “using models of how sub-word units are combined to form words.” *Jiang*, 1:49-57. Specifically, *Jiang* teaches performing a “tree search” using computed probabilities to determine the most likely word uttered. *See* Sections V.B.4-V.B.5 (*citing Jiang*, 7:29-54, 8:5-51). According to *Jiang*’s algorithm, a sequence of incoming states is successively cumulated by *Jiang*’s circuit to “determine a most likely phoneme represented by” the features extracted from the user’s utterance. *Jiang*, 7:38-46, 1:41-53, 2:1-15. *Jiang*’s circuit then searches a phoneme-based prefix tree to “determine a most likely state sequence for which the overall probability is maximum, given the transition probabilities between states and the observation probabilities.” *Jiang*, 1:19-57, 2:47-3:21, 7:12-15, 8:16-51, 2:1-15; *Dec.*, 93.

Thus, through both its description of (1) probabilities being calculated frame by frame to determine the most likely phonemes uttered; and (2) successively inputting probabilities of the likely phonemes uttered to search a lexical tree, *Jiang* teaches “the probability is an input to an evaluation of a state transition of a model of states[,]” as claimed. *Dec.*, 92-93.

D. Claim 3

1. Claim 3: “...wherein the model is a Hidden Markov Model”

Jiang teaches this limitation. *See* Section V.B.3 (*citing Jiang*, 1:34-57, 2:1-15, explaining that an HMM may be used at the subword level (phoneme calculation).

E. Claim 5

1. ***Claim 5: “...further comprising: a buffer for storing one or more feature vectors coupled to at least one of the plurality of processors of the one or more clusters of processors”***

The speech recognition circuit of *Jiang-Chen* teaches or at least renders obvious claim 5.

Jiang teaches speech input into a microphone is converted into a “sequence of digital signals.” *Jiang*, 6:46-7:10. The signals are provided to a feature extraction module 66 by A/D converter at a sample rate of approximately 16 kHz, which divides the digital signal into frames approximately 10 milliseconds in duration. *Id.* The frames are then encoded by the feature extraction module into a feature vector and encoded into codewords. *Id.* “Thus, feature extraction module 66 provides, at its output the feature vectors (or codewords) for each spoken utterance. The feature extraction module 66 preferably provides the feature vectors (or codewords) at a rate of one codeword approximately every 10 milliseconds.” *Jiang*, 7:5-10, *see also* 1:58-67, Fig. 2.

Jiang further teaches a “processing unit 21, system memory 22, and a system bus 23 that couples various system components including the system memory to the processing unit 21.” *Jiang*, 5:4-39 (pinpoint at 5:7-9), Fig. 1. The system bus 23 may be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus

architectures. *Id.* The system memory includes read only memory (ROM) 24 and a random access memory (RAM) 25. *Id.*

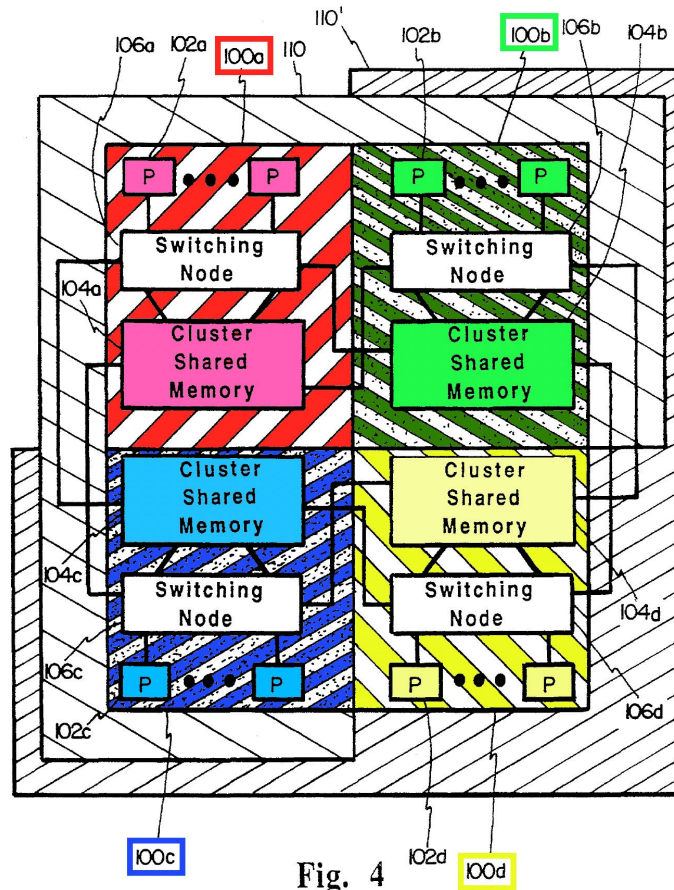
A POSITA would have recognized *Jiang*'s sequence of feature vectors would have been provided from the feature extraction module to the processing unit to the tree search engine implemented in processing unit 21 via the system memory 22. *Jiang*, 5:4-39, 6:39-42; *Dec.*, 94. Specifically, a POSITA would have recognized *Jiang* teaches "storing" the feature vectors, teaching or otherwise rendering obvious the claimed "buffer." *Id.* Further, for the combination described above in which *Jiang*'s processing unit 21 is modified pursuant to *Chen* to comprise one or more clusters of processors, a POSITA would have recognized *Jiang* in view of *Chen* teaches storing one or more feature vectors coupled to at least one of the plurality of processors of the one or more clusters of processors, rendering obvious the claimed "buffer." *See* Section VI.C.2; *Dec.*, 94.

F. Claim 7

1. ***Claim 7: "...the one or more clusters of processors comprises a first cluster of processors and a second cluster of processors; the first cluster comprises a first acoustic model memory; and the second cluster comprises a second acoustic model memory that is distinct and separate from the first acoustic model memory"***

Jiang in view of *Chen* teaches claim 7. Specifically, as discussed, *Jiang* in view of *Chen* teaches a first cluster of processors and second cluster of processors,

each of which comprises a cluster shared memory, as color-coded below in annotated Fig. 4.



Chen, Fig. 4 (annotated); *See* Sections V.B.2, V.B.3. In the proposed mapping, each cluster 100a-d would have acoustic model data stored in its respective cluster shared memory 104a-d. Thus, *Jiang* in view of *Chen* teaches a first cluster comprising a first acoustic model memory and a second cluster comprising a second acoustic model memory. *Id.*

Chen further teaches that some, but not all, the clusters are interconnected. *Chen*, 9:5-19. Specifically, *Chen* states the processors 102a in cluster 100a can directly access memories 104a, 104b, and 104c of clusters 100a, 100b, and 100c, respectively, but cannot directly or adjacently access the memory 104d of cluster 100d. *Chen*, 9:30-43; *Dec.*, 95. Similarly, the processors 102d in cluster 100d can directly access memory 104d, 104b, and 104c, but cannot directly or adjacently access the memory 104a. *Id.* Thus, a POSITA would have recognized that *Chen* teaches or at least renders obvious at least a first cluster (e.g., cluster 100a) comprising a first acoustic model memory (e.g., memory 104a) and a second cluster (e.g., cluster 100d) comprising a second acoustic model memory (e.g., memory 104d) that is “separate and distinct” from the first acoustic model memory (104a). *Id.* Further, as discussed, *Chen*’s clustered computing architecture is described as providing multiple processors in “**physically separable clusters**, each cluster having a common cluster shared memory that is symmetrically accessible by all of the processors in that cluster.” *Chen*, 5:9-17; *Dec.*, 95. Based on this description, to the extent the clusters illustrated in the embodiment of Fig. 4 are not sufficiently “separate and distinct,” a POSITA would have recognized *Chen* teaches embodiments in which clusters 100a-d are further separate and distinct from one another. *Id.*

G. Claim 8

1. **Claim 8:** “...the one or more processors comprises a first cluster of processors; and

the first cluster comprises a first acoustic model coupled to all processors of the first cluster.”

Jiang in view of *Chen* teaches this limitation. See Sections V.B.2-V.B.4. As discussed, *Jiang* teaches an HMM acoustic model stored in model memory 72. See Section V.B.3. *Jiang* modified pursuant to the clustered processor description of *Chen* teaches a first cluster of processors. See Section V.B.2. As further discussed, the circuit of *Jiang* in view of *Chen* teaches the HMM acoustic model is stored in the shared cluster memories 104a-d of each of the clusters 100a-d. See Section

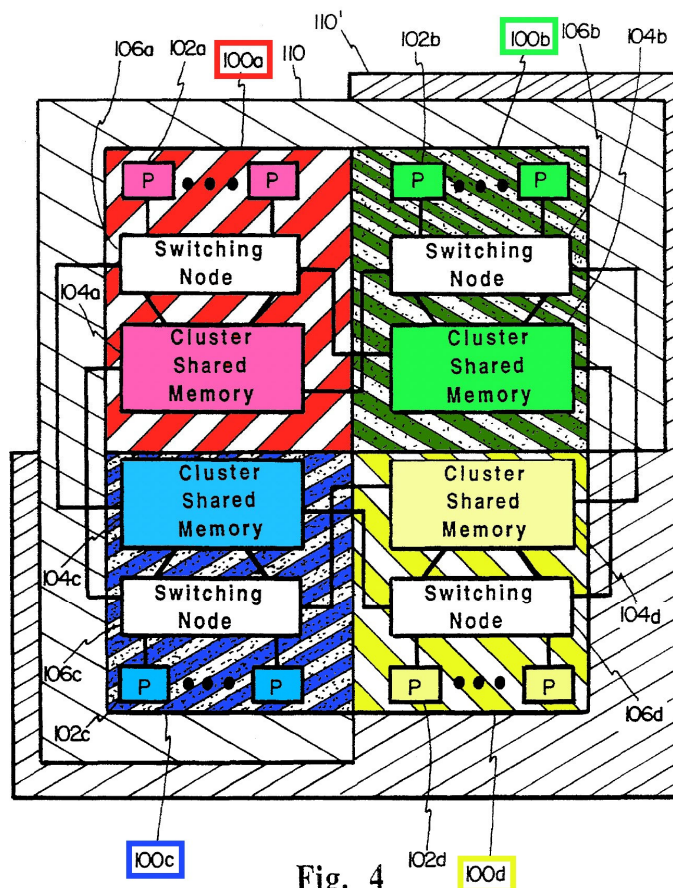


Fig. 4

V.B.4. Specifically, as seen below in Fig. 4 of Chen, the memories 104a-d within each cluster 100a-d are coupled to each of the respective processors 102a-d comprising that cluster.

Chen, Fig. 4 (annotated), 9:20-34 (“...the processors 102a in cluster 100a can directly access the cluster shared memory 104a of cluster 100a, and...the processors 102d in cluster 100d can directly access the cluster shared memory 104d of cluster 100d...”). Thus, *Jiang* in view of *Chen* teaches claim 8.

VI. GROUND 2: CLAIMS 1-3, 5, AND 7-8 ARE OBVIOUS OVER *JIANG* IN VIEW OF *CHEN* IN FURTHER VIEW OF *LUCKE*

A. Overview of Ground 2

As discussed, *Jiang* in view of *Chen* teaches or otherwise renders obvious claims 1-8. Further, *Lucke* teaches a speech recognition circuit that expressly teaches the claim language of “the initial score is used to determine whether to continue processing to determine a final score via processing a larger amount of model data than that was processed to generate the initial score” recited in claim 1, as detailed in the below mapping. *Lucke*, 0004-0005, 0081-0082.

To the extent Patent Owner argues *Jiang*’s pruning disclosure (focusing on *not* continuing processing when scores are *below* a threshold, as described by the ’140 Patent and File History) fails to teach claim 1, it would have been obvious to a POSITA to configure the circuit of *Jiang* in view of *Chen* to *continue* processing

using a larger amount of model data when scores are *above* a threshold, as taught by *Lucke*. *Lucke*, 0081-0082; *Dec.*, 96-98; *see also* Section VI.C.6.

B. Claim 1

1. Claim 1(Pre): “A speech recognition circuit comprising:”

To the extent the preamble is limiting, *Jiang* in view of *Chen* teaches a speech recognition circuit. *See* Section VI.C.1.

2. Claim 1(a): “one or more clusters of processors, each of the one or more clusters of processors comprising: a plurality of processors;”

Jiang in view of *Chen* teaches this limitation. *See* Section VI.C.2.

3. Claim 1(b): “and [one or more clusters of processors, each of the one or more clusters of processors comprising:] an acoustic model memory storing acoustic model data”

Jiang in view of *Chen* teaches this limitation. *See* Section VI.C.3.

4. Claim 1(c): “wherein each of the plurality of processors is configured to compute a probability using the acoustic model data in the acoustic model memory, wherein:”

Jiang in view of *Chen* teaches this limitation. *See* Section VI.C.4.

5. Claim 1(d): “the speech recognition circuit is configured to generate an initial score for an audio sample; and

Jiang in view of *Chen* teaches this limitation. *See* Section VI.C.5.

6. Claim 1(e): “the initial score is used to determine whether to continue processing to determine a final score via processing a

larger amount of model data than that was processed to generate the initial score

Jiang in view of *Chen* teaches this limitation. See Section VI.C.6. Further, *Lucke* expressly teaches the circuit continues processing branches whose initial score is above a threshold.

Specifically, *Lucke* teaches a speech recognition circuit (*Lucke*, 0002-0005, Figs. 1,7) employing an HMM “acoustic model” (*Lucke*, 0008, 0025, 0079) that searches a lexical tree (*Lucke*, 0017-0023, Fig. 4) for an uttered word. As part of its tree searching algorithm, *Lucke*’s circuit calculates “acoustic scores [] using a series of features on the basis of acoustic models which are connected along a series of branches...extending from the root node” of the word network tree to the “end node” representing a recognized word. *Lucke*, 0025, 0070. *Lucke* expressly refers to its circuit as calculating a “final score.” *Lucke*, 0012-0016, 0029, 0085.

Like *Jiang*, *Lucke* teaches a pruning process for its lexical tree search in which branches with an acoustic score lower than a threshold are “discarded” to reduce processing requirements. *Lucke*, 0027-0029, 0087; *Dec.*, 96; see also Section VI.C.6. However, *Lucke* goes on to expressly state that branches representing words having an acoustic score “greater than a predetermined threshold value” are selected to remain in consideration and **“the processing continues.”** *Lucke*, 0027-0029, 0081-0082; *Dec.*, 96. Thereafter, *Lucke*’s circuit continues performing calculations of hypothesized words (and, subsequently through iterations, word sequences) until

its matching section determines “from among such word sequences, a word sequence having the greatest final score...as the speech recognition result.” *Lucke*, 0029, 0085; *Dec.*, 96. That is, *Lucke* expressly teaches a speech recognition circuit in which an initial score is used to determine whether to “continue processing” to determine a final score via processing a larger amount of model data than that was processed to generate the initial score. *Id.*

A POSITA, having read the disclosure of *Lucke*, would have recognized the circuit taught by *Jiang* in view of *Chen*, which is already expressly configured to *cease* processing particular branches of a lexical tree based on a *low* initial score, would have appropriately been configured to *continue* processing for branches of the lexical tree having a sufficiently *high* initial score. *Lucke*, 0027-0029, 0081-0082; *Dec.*, 97; *See* Section VI.C.6. Such a functionality (continuing to process sufficiently highly-scoring branches) would have been the common-sense, expected algorithm for lexical tree-based speech recognition systems of the time performing pruning. *Dec.*, 97. Any modification necessary to modify the system of *Jiang-Chen* to continue processing high-scoring branches would have only required minor software revisions, well within the skill of a POSITA at the time of the ’140 Patent. *Id.* A POSITA would have recognized processing would have needed to continue for branches with a sufficiently high score for the system to determine any final

speech recognition result, leading to a reasonable expectation of success for the combination of *Jiang*, *Chen*, and *Lucke*. *Id.*

C. Claim 2

1. ***Claim 2: “...the probability is an input to an evaluation of a state transition of a model of states”***

Jiang in view of *Chen* and *Lucke* teaches this limitation for the reasons detailed above for *Jiang-Chen*. See Section VI.C.1.

D. Claim 3

1. ***Claim 3: “...the model is a Hidden Markov Model”***

Jiang in view of *Chen* and *Lucke* teaches this limitation for the reasons detailed above for *Jiang-Chen*. See Section VI.D.1.

E. Claim 5

1. ***Claim 5: “...a buffer for storing one or more feature vectors coupled to at least one of the plurality of processors of the one or more clusters of processors”***

Jiang teaches this limitation. See Section VI.F.1. Further, *Lucke* explicitly teaches a “buffer [that] temporarily stores the features supplied from the feature extraction section.” *Lucke*, 0076-0077. *Lucke*’s matching section calculates scores using “the series of the necessary features stored in the buffer.” *Id.*

To the extent it is argued that *Jiang* in view of *Chen* does not teach the claimed “buffer” for storing its feature vectors, a POSITA would have been motivated to combine *Lucke*’s buffer into the system as a well-known, operable method of storing

the feature vectors. *Dec.*, 98. Doing so would have provided flexibility in the scheduling of the processors of *Jiang-Chen*, allowing each coupled processor to access the feature vector as the particular processor becomes available for a next operation. *Id.* Like *Jiang*, *Lucke*'s circuit extracts feature vectors via frames derived from sound input to a microphone and supplied to an A/D converter, leading a POSITA to a reasonable expectation of success for the combination. *Compare Jiang* 6:46-7:10 *with Chen*, 0005-0006; *Dec.*, 98.

F. Claim 7

1. ***Claim 7: "...the one or more clusters of processors comprises a first cluster of processors and a second cluster of processors; the first cluster comprises a first acoustic model memory; and the second cluster comprises a second acoustic model memory that is distinct and separate from the first acoustic model memory"***

Jiang in view of *Chen* and *Lucke* teaches this limitation for the reasons detailed above for *Jiang-Chen*. See Section VI.F.1.

G. Claim 8

1. ***Claim 8: "...the one or more processors comprises a first cluster of processors; and the first cluster comprises a first acoustic model coupled to all processors of the first cluster."***

Jiang in view of *Chen* and *Lucke* teaches this limitation for the reasons detailed above for *Jiang-Chen*. See Section VI.G.1.

**VII. GROUND 3: CLAIM 4 IS OBVIOUS OVER *JIANG* IN VIEW OF *CHEN*
IN FURTHER VIEW OF *ROBINSON***

A. Claim 4: “...wherein the probability is computed from a Gaussian mixture model and one or more feature vectors”

As discussed, *Jiang* in view of *Chen* teaches claim 1. *Robinson* teaches a similar speech recognition circuit comparing sampled speech to an HMM acoustic model to compute probabilities of matches. *Robinson*, 1:5-9, 8:28-67, 9:52-57; *Dec.*, 99. *Robinson*’s phoneme probability calculator relates the “acoustic observations” of speech to “linguistic states such as phonemes” to compute the “**probability** of seeing an acoustic **vector**, $O(t)$ at time (t) given the state $S(t)$.” *Robinson*, 8:56–9:16; *Dec.*, 99 (explaining that the “acoustic vector” is a “feature vector,” as claimed). *Robinson*’s circuit specifically performs these probability computations using a Gaussian mixture model: “Several parametric models may be used to approximate this function. One such model is a mixture of Gaussians” *Robinson*, 9:4-5. Therefore, *Robinson* teaches Claim 4.

A POSITA, having read *Robinson*, would have been motivated to perform the feature vector probability computations already taught by *Jiang* using the Gaussian mixture model technique taught by *Robinson* to improve the similar circuit of *Jiang-Chen* in the same way. *Dec.*, 99 (citing *Ravishankar*, 28-29). Gaussian mixture models were one of “[s]everal parametric models [] used to approximate” the probability computation known at the time of the ’140 Patent, and thus would have

been obvious for a POSITA to try. *Id.* Using the Gaussian mixture model would have only required programming hardware already present in the circuit of *Jiang-Chen* that would have been within the skill of a POSITA without undue experimentation, leading to a reasonable likelihood of success for the combination. *Id.*

VIII. GROUND 4: CLAIM 4 IS OBVIOUS OVER *JIANG* IN VIEW OF *CHEN* IN FURTHER VIEW OF *LUCKE* IN FURTHER VIEW OF *ROBINSON*

A. Claim 4

Jiang in view of *Chen*, *Lucke*, and *Robinson* teaches this limitation for the reasons detailed above for *Jiang-Chen-Robinson*. See Section VIII.A.

IX. GROUND 5: CLAIM 6 IS OBVIOUS OVER *JIANG* IN VIEW OF *CHEN* IN FURTHER VIEW OF *WRENCH*

A. Claim 6: “...a search controller coupled to the at least one of the plurality of processors of the one or more clusters of processors, the search controller capable of controlling the at least one of the plurality of processors of the one or more clusters of processors to initiate speech recognition processing in at least a subset of processors of the one or more clusters of processors”

Jiang teaches performing a tree “search” to determine a most likely phoneme represented by features extracted from a user’s utterance. See Section VI.C.1, *citing Jiang*, 7:34-43. *Chen* describes a shared control mechanism 122 coupled to at least one of its plurality of processors that is utilized by a common parallel user interface to coordinate and communicate among processes executing in parallel in the clustered processing computer system. *Chen*, 6:4-22, 18:43-63, 19:14-57.

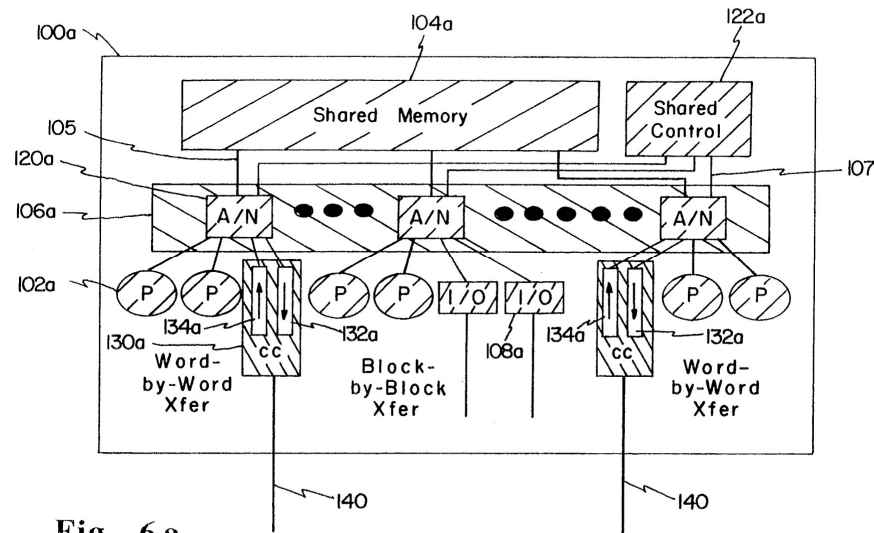


Fig. 6 a

Chen, Fig. 6a, depicting shared control mechanism 122a. However, neither *Chen* nor *Jiang* expressly teaches initiating speech recognition processing in at least a subset of processors via a search controller.

Wrench, a similar multiprocessor speech recognition circuit, teaches this limitation via its “master processor” that acts as an executive to its plurality of speech recognition processors. *Wrench*, 4:40-58, 5:6-24, 6:38-7:32, Fig. 2; *Dec.*, 100. The master processor’s software includes an “initialization” program module that “performs the various tasks **initializing** the [speech recognition subsystem] parameters.” *Wrench*, 6:40-50, 2:44-58; *Dec.*, 100. A POSITA would have recognized, therefore, that *Wrench*’s master processor teaches a controller coupled to a plurality of processors that “initiates speech recognition processing in at least a subset of processors,” as claimed. *Dec.*, 100.

A POSITA would have been motivated to incorporate *Wrench's* master processor as part of the shared control mechanism taught by *Chen* to initiate the search process taught by *Jiang* to initialize speech recognition in the plurality of processors. *Dec.*, 100. Adding *Wrench's* master processor initialization technique would have ensured smooth, cohesive operation of the plurality of processors, improving the functionality of the overall circuit. *Id.* This modification to *Jiang-Chen* would have only required providing *Wrench's* software modules to hardware already present in the circuit of *Jiang-Chen*. *Id.* Modifying software stored in memory would have been within the skill of a POSITA without undue experimentation, leading to a reasonable likelihood of success. *Id.*

X. GROUND 6: CLAIM 6 IS OBVIOUS OVER *JIANG* IN VIEW OF *CHEN* IN FURTHER VIEW OF *LUCKE* IN FURTHER VIEW OF *WRENCH*

A. Claim 6

Jiang in view of *Chen*, *Lucke*, and *Wrench* teaches this limitation for the reasons detailed above for *Jiang-Chen-Wrench*. See Section IX.A.

XI. THE BOARD'S DISCRETION UNDER 35 U.S.C. § 314(a) AND § 325(d)

§ 314(a). None of the factors articulated in *General Plastic Industrial Co., Ltd. v. Canon Kabushiki Kaisha*, IPR2016-01357, Paper 19 at 16 (PTAB Sept. 6, 2017) favor discretionary denial. Regarding Factors 1–5, Petitioners have filed no prior petition regarding the '140 Patent. Regarding Factors 6–7, joinder will neither consume additional Board resources nor affect the timing of a final written decision,

because Petitioners agree to assume an understudy role behind Apple (with the same case schedule and arguments) and not actively participate in proceedings, unless Apple's participation terminates.

The considerations in *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019 (PTAB, Mar. 20, 2021) also disfavor discretionary denial. Trial in the co-pending litigation, scheduled for April 22, 2024 (Ex. 1044), will begin approximately at the same time a Final Written Decision is due, approximately May 1, 2024. (*Fintiv* Factor 2). The court only recently issued its Claim Construction Order on June 16, 2023, and dispositive motions are not due until January 26, 2024. Therefore, the parties and the court are at the early stages of the invalidity analysis. (*Fintiv* Factor 3). The Petition also presents a strong showing of unpatentability; indeed, the Board has already found Apple's Petition, that presented grounds of unpatentability identical to those presented here, has shown that there is a reasonable likelihood it would prevail with respect to at least one of the claims challenged in the Petition. (IPR2023-00037, Paper 10; *Fintiv* Factor 6). These factors favor not exercising discretionary denial under § 314(a).

§ 325(d). Petitioners seek to join already instituted IPR2023-00037, based on the same grounds as applied to the same claims. Petitioners agree to assume an understudy role and will only actively participate in proceedings if Apple's participation terminates.

XII. CONCLUSION

For the forgoing reasons, Petitioners respectfully request *inter partes* review of the Challenged Claims of the '140 Patent.

Dated: July 12, 2023

Respectfully submitted,

/J. David Hadden/
J. David Hadden
Reg. No. 40,629

Attorneys for Petitioners
Amazon Web Services, Inc. and
Amazon.com Services LLC

XIII. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8(A)(1)

A. Real Party-In-Interest

Amazon.com Services LLC and Amazon Web Services, Inc. are the real parties-in-interest. 37 C.F.R. § 42.8(b)(1).

B. Related Matters

The '140 Patent is currently the subject of *Apple Inc. v. Zentian Ltd.*, Inter Partes Review No. IPR2023-00037 (P.T.A.B.) (instituted Jun. 12, 2023). This Petition is substantively identical to the petition submitted in IPR2023-00037 and Petitioners have concurrently filed a motion to join that proceeding.

The '140 Patent is also asserted in the following pending patent infringement lawsuits filed by Zentian Ltd:

- Zentian Ltd v. Apple Inc., No. 6:22-cv-00122 (W.D. Tex. Feb. 2, 2022);
- Zentian Ltd v. Amazon.com, Inc., No. 6:22-cv-00123 (W.D. Tex. Feb. 2, 2022).

C. Lead and Back-Up Counsel

Petitioners provide the following designation and service information for lead and back-up counsel. 37 C.F.R. § 42.8(b)(3) and (b)(4).

LEAD COUNSEL	BACK-UP COUNSEL
J. David Hadden (Reg. No. 40,629) dhadden@fenwick.com Fenwick & West LLP 801 California Street Mountain View, CA 94041	Saina Shamilov (Reg. No. 48,266) sshamilov@fenwick.com Fenwick & West LLP 801 California Street Mountain View, CA 94041

Petition for *Inter Partes* Review of
U.S. Patent No. 10,971,140

Telephone: (650) 988-8500 Facsimile: (650) 938-5200	Telephone: (650) 988-8500 Facsimile: (650) 938-5200 Dargaye Churnet (Reg. No. 71,288) dchurnet@fenwick.com Fenwick & West LLP 555 California Street, 12th Floor San Francisco, CA 94104 Telephone: (415) 875-2300 Facsimile: (415) 281-1350
--	---

D. Service Information

Service of any documents may be made to the postal mailing address above.

Petitioners consent to electronic service at Amazon-Zentian-IPR@fenwick.com. A Power of Attorney is being filed concurrently with the designation of counsel in accordance with 37 C.F.R. § 42.10(b).

CLAIM LISTING APPENDIX
U.S. Patent No. 10,971,140 for Claims 1-8

Claim Designation	Claim Language
Claim 1(Pre)	A speech recognition circuit comprising:
Claim 1(a)	one or more clusters of processors, each of the one or more clusters of processors comprising: a plurality of processors;
Claim 1(b)	and an acoustic model memory storing acoustic model data,
Claim 1(c)	wherein each of the plurality of processors is configured to compute a probability using the acoustic model data in the acoustic model memory,
Claim 1(d)	wherein: the speech recognition circuit is configured to generate an initial score for an audio sample;
Claim 1(e)	and the initial score is used to determine whether to continue processing to determine a final score via processing a larger amount of model data than that was processed to generate the initial score.
Claim 2	The speech recognition circuit of claim 1, wherein the probability is an input to an evaluation of a state transition of a model of states.
Claim 3	The speech recognition circuit of claim 2, wherein the model is a Hidden Markov Model.
Claim 4	The speech recognition circuit of claim 2, wherein the probability is computed from a Gaussian mixture model and one or more feature vectors.
Claim 5(Pre)	The speech recognition circuit of claim 1, further comprising:
Claim 5(a)	a buffer for storing one or more feature vectors coupled to at least one of the plurality of processors of the one or more clusters of processors.

Claim Designation	Claim Language
Claim 6(Pre)	The speech recognition circuit of claim 5, further comprising:
Claim 6(a)	a search controller coupled to the at least one of the plurality of processors of the one or more clusters of processors, the search controller capable of controlling the at least one of the plurality of processors of the one or more clusters of processors to initiate speech recognition processing in at least a subset of processors of the one or more clusters of processors.
Claim 7(Pre)	The speech recognition circuit of claim 1, wherein:
Claim 7(a)	the one or more clusters of processors comprises a first cluster of processors and a second cluster of processors;
Claim 7(b)	the first cluster comprises a first acoustic model memory; and
Claim 7(c)	the second cluster comprises a second acoustic model memory that is distinct and separate from the first acoustic model memory.
Claim 8(Pre)	The speech recognition circuit of claim 1, wherein:
Claim 8(a)	the one or more clusters of processors comprises a first cluster of processors; and
Claim 8(b)	the first cluster comprises a first acoustic model coupled to all processors of the first cluster.

CERTIFICATION OF WORD COUNT

The undersigned certifies pursuant to 37 C.F.R. § 42.24 that the foregoing Petition for *Inter Partes* Review, excluding any table of contents, mandatory notices under 37 C.F.R. § 42.8, certificates of service or word count, or appendix of exhibits, contains 12,431 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: July 12, 2023

/J. David Hadden/
J. David Hadden
Reg. No. 40,629

Attorneys for Petitioners
Amazon Web Services, Inc. and
Amazon.com Services LLC

**CERTIFICATE OF SERVICE ON PATENT OWNER
UNDER 37 C.F.R. § 42.105**

I hereby certify, pursuant to 37 C.F.R. §§ 42.6(e) and 42.105, that a complete copy of this Petition for *Inter Partes* Review including all exhibits are being served via Federal Express on July 12, 2023, upon Patent Owner by serving the correspondence address of record with the USPTO as follows:

Finnegan, Henderson, Farabow, Garrett & Dunner LLP
901 New York Avenue, NW
Washington DC 20001-4413

Further, a courtesy copy of this Petition for *Inter Partes* Review was sent via Federal Express to Patent Owner's litigation counsel:

John M. Hughes
Katherine L.I Hacker
Bartlit Beck LLP
1801 Wewatta Street, Suite 1200
Denver, CO 80202

Katherine E. Rhoades
Nevin M. Gewertz
Rebecca T. Horwitz
Bartlit Beck LLP
54 West Hubbard Street, Suite 300
Chicago, IL 60654

John Francis Luman , III
Daniels & Tredennick PLLC
6363 Woodway, Suite 700
Houston, TX 77057

/J. David Hadden/
J. David Hadden
Reg. No. 40,629

Attorneys for Petitioners
Amazon Web Services, Inc. and
Amazon.com Services LLC