
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. 7,587,319

Issued: September 8, 2009

Inventor: Mark Catchpole

Title: SPEECH RECOGNITION CIRCUIT USING PARALLEL PROCESSORS

Inter Partes Review No. IPR2023-01192

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 7,587,319**

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 '319 PATENT	1
A. Patent Description	1
B. Summary of Unpatentability	2
C. Priority Date	5
D. Level of Skill of a POSITA.....	5
III. REQUIREMENTS FOR IPR UNDER 37 C.F.R. §42.104.....	6
A. Grounds for Standing (37 C.F.R. §42.104(a))	6
B. Identification of Challenge (37 C.F.R. §42.104(b)).....	6
C. Claim Construction (37 C.F.R. §42.104(b)(3)).....	7
IV. SHOWING OF ANALOGOUS PRIOR ART	7
A. <i>Thelen</i>	7
B. <i>Bailey</i>	8
C. <i>Chen</i>	8
D. <i>Huang</i>	8
E. <i>Jiang</i>	9
F. <i>Katz</i>	9
V. GROUND 1: CLAIMS 46, 50, 54, 64, AND 67 ARE OBVIOUS OVER <i>THELEN</i> IN VIEW OF <i>BAILEY</i> IN FURTHER VIEW OF <i>CHEN</i>	10
A. Claim 46	10

TABLE OF CONTENTS (Continued)

	Page
1. Claim 46(Pre): “A speech recognition circuit comprising:”	10
a) <i>Thelen’s</i> Teachings	10
b) Obviousness of Speech Recognition “circuit”	11
2. Claim 46(a): “an input buffer receiving processed speech parameters”	12
a) The ’319 Patent’s Description of “Processed Speech Parameters”	12
b) “processed speech parameters”	13
c) “an input buffer for receiving [processed speech parameters]”	14
(1) <i>Thelen’s</i> Teachings	14
(2) <i>Bailey’s</i> Teachings	17
d) Motivation to Combine <i>Bailey</i> with <i>Thelen</i> for Claim 46(b)	21
3. Claim 46(b): “a plurality of lexical memories containing in combination complete lexical data for word recognition, each lexical memory containing part of said complete lexical data”	23
a) <i>Thelen</i> Teaches a Memory Containing “in combination complete lexical data for word recognition” and “part of said complete lexical data”	24
(1) Overview of <i>Thelen</i>	24
(2) “lexical data”	26
(3) “in combination complete lexical data”	27

TABLE OF CONTENTS (Continued)

	Page
(4) “for word recognition”	28
(5) “part of said complete lexical data”	28
(6) Lexical Memory	30
b) <i>Bailey</i> Teaches a “plurality of lexical memories”	31
4. Claim 46(c): “a plurality of processors connected in parallel to said input buffer for processing the speech parameters in parallel”	33
5. Claim 46(d): “said processors being arranged in groups of processors, each group of processors being connected to a [] memory”	38
a) <i>Chen</i> Teaches Claim 46(d) Except for a “lexical” Memory.....	38
b) Motivation to Combine <i>Chen</i> with <i>Thelen-Bailey</i> for Claims 46(b) and 46(d).....	40
6. Claim 46(e): “a control processor controlling each processor to process said speech parameters using partial lexical data read from a respective said lexical memory”	46
a) <i>Thelen’s</i> Teachings	46
b) <i>Bailey’s</i> Teachings.....	48
c) Motivation to Modify <i>Thelen-Bailey-Chen</i> to Include <i>Bailey’s</i> CPU for Claim 46(d)	49
7. Claim 46(f): “a results memory storing the results of the processing of the speech parameters from said processors”	51
a) <i>Bailey</i> Teaches Claim 46(f)	51

TABLE OF CONTENTS
(Continued)

	Page
b) Motivation to Combine <i>Bailey</i> with <i>Thelen</i> for Claim 46(e).....	52
B. Claim 50	54
1. Claim 50: “...wherein each processor compares the processed speech parameters with the lexical data in a corresponding lexical memory to identify words as a word recognition event and sends information identifying the identified words to said results memory as the processing results.”	54
C. Claim 54	55
1. Claim 54: “...wherein said processors process said speech parameters by comparing said speech parameters with said lexical data or data derived from lexical data in respective said lexical memories.”	55
D. Claim 64	57
1. Claim 64: “...wherein said input buffer is arranged to receive said speech parameters as feature vectors.”	57
E. Claim 67	57
1. Claim 67: “...wherein said lexical memories store said lexical data as Hidden Markov Models, and each processor performs the Viterbi search algorithm using a respective part of said lexical data”	57
VI. GROUND 2: CLAIMS 47-49 ARE OBVIOUS OVER <i>THELEN</i> IN VIEW OF <i>BAILEY</i> AND <i>CHEN</i> AND IN FURTHER VIEW OF <i>HUANG</i>	58
A. Claim 47	58

TABLE OF CONTENTS
(Continued)

	Page
1. Claim 47: "...wherein each lexical memory stores lexical data as at least one lexical tree data structure, each lexical tree data structure comprises a model of words having common prefix components, and an initial component of each lexical tree data structure is unique"	58
B. Claim 48	62
1. Claim 48: "...wherein each lexical memory stores said lexical tree data structure as an n phone model of words, where n is an integer, and said components comprise n phones."	62
C. Claim 49	64
1. Claim 49: "...wherein each lexical memory stores said lexical tree data structure as a mono phone model of words, and said lexical tree processors use said mono phone models to generate context dependant [sic] phone models of words dynamically for use in processing the speech parameters."	64
VII. VII. GROUND 3: CLAIMS 53 AND 58-59 ARE OBVIOUS OVER <i>THELEN</i> IN VIEW OF <i>BAILEY</i> AND <i>CHEN</i> IN FURTHER VIEW OF <i>JIANG</i>	67
A. Claim 53	67
1. Claim 53: "...wherein each processor determines and outputs scores for words in the processing results at word recognition events."	67
B. Claim 58	70
1. Claim 58: "...wherein said processors determine and output scores in the processing results during the processing of said speech parameters"	70

TABLE OF CONTENTS
(Continued)

	Page
C. Claim 59	72
1. Claim 59: “...wherein said processors prune the processing of the speech parameters in dependence upon the determined scores”	72
VIII. GROUND 4: CLAIM 51 IS OBVIOUS OVER <i>THELEN</i> IN VIEW OF <i>BAILEY</i> AND <i>CHEN</i> IN FURTHER VIEW OF <i>KATZ</i>	72
A. Claim 51	72
1. Claim 51: “...including language model processor providing a language model output for modifying the processing results at a word recognition event by a said processor”	72
IX. THE BOARD’S DISCRETION UNDER 35 U.S.C. §314(a) and §325(d)	75
X. CONCLUSION	76
XI. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8(A)(1)	77
A. Real Party-In-Interest	77
B. Related Matters	77
C. Lead and Back-Up Counsel	77
D. Service Information	78
CLAIM LISTING APPENDIX	81

TABLE OF AUTHORITIES

	Page(s)
CASES	
<i>Apple Inc. v. Fintiv, Inc.</i> , IPR2020-00019 (PTAB, Mar. 20, 2021)	77
<i>General Plastic Industrial Co., Ltd. v. Canon Kabushiki Kaisha</i> , IPR2016-01357, Paper 19 (PTAB Sept. 6, 2017).....	77
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303 (Fed. Cir. 2005) (<i>en banc</i>)	7
STATUTES AND RULES	
35 U.S.C. §102	7, 8, 9
35 U.S.C. §315	1, 6
OTHER AUTHORITIES	
37 C.F.R. §42.100(b)	7
37 C.F.R. §42.104	6, 7
37 C.F.R. §42.122	6
83 Fed. Reg. 197 (Oct. 11, 2018).....	7

EXHIBIT LIST (37 C.F.R. § 42.63(E))

Exhibit	Description
1001	U.S. Patent No. 7,587,319 (“’319 Patent”)
1002	Prosecution History for the 7,587,319 Patent (“’319 File History”)
1003	Declaration of Christopher Schmandt
1004	U.S. Patent No. 6,374,219 to Jiang (“ <i>Jiang</i> ”)
1005	U.S. Patent No. 5,428,803 to Chen et al. (“ <i>Chen</i> ”)
1006	Hsiao-Wuen Hon, <i>A survey of hardware architectures designed for speech recognition</i> , Carnegie Mellon University, 1991 (“ <i>Hon</i> ”)
1007	Ph.D. Thesis of Mosur Ravishankar (“ <i>Ravishankar</i> ”)
1008	[RESERVED]
1009	U.S. Patent No. 5,983,180 to Robinson (“ <i>Robinson</i> ”)
1010	U.S. Patent No. 5,036,539 to Wrench Jr., et al. (“ <i>Wrench</i> ”)
1011	Frederick Jelinek, <i>Statistical Methods for Speech Recognition</i> , The MIT Press, 1997 (“ <i>Jelinek</i> ”)
1012	Christopher Schmandt, <i>Voice Communication with Computers</i> , Van Nostrand Reinhold, 1994 (“ <i>Schmandt</i> ”)
1013	Lawrence Rabiner and Biing-Hwang Juang, <i>Fundamentals of Speech Recognition</i> , Prentice Hall PTR, 1993 (“ <i>Rabiner</i> ”)
1014	Richard Klevans and Robert Rodman, <i>Voice Recognition</i> , Artech House, 1997 (“ <i>Klevans</i> ”)
1015	U.S. Patent No. 7,120,582 to Young et al. (“ <i>Young</i> ”)
1016	John Holmes and Wendy Holmes, <i>Speech Synthesis and Recognition</i> , 2nd Edition, Taylor & Francis, 2001 (“ <i>Holmes</i> ”)

Petition for *Inter Partes* Review of
U.S. Patent No. 7,587,319

Exhibit	Description
1017	U.S. Patent No. 4,926,488 to Nadas et al. (“ <i>Nadas</i> ”)
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 Stublely et al. (“ <i>Stublely</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: ImplicationsofNetworkandMultiprocessorComputer 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	U.S. Patent No. 6,526,380 to Thelen et al. (“ <i>Thelen</i> ”)
1031	U.S. Patent No. 5,459,798 to Bailey et al. (“ <i>Bailey</i> ”)
1032	Zentian, Ltd. v Apple Inc. WDTX Waco (6:22-cv-122) Scheduling Order “ <i>Scheduling Order</i> ”

Petition for *Inter Partes* Review of
U.S. Patent No. 7,587,319

Exhibit	Description
1033	Declaration of June Ann Munford (“ <i>Munford</i> ”)
1034	U.S. Patent No. 6,076,056 to Huang et al. (“ <i>Huang</i> ”)
1035	U.S. Patent No. 4,831,550 to Katz (“ <i>Katz</i> ”)
1036	U.S. Patent No. 4,567,606 to Vensko et al. (“ <i>Vensko</i> ”)
1037	U.S. Patent No. 4,907,278 to Cecinati et al. (“ <i>Cecinati</i> ”)
1038	B. I. (Raj) Pawate and Peter D. Robinson, “ <i>Implementation of an HMM-Based, Speaker-Independent Speech Recognition System on the TMS320C2x and TMS320C5x</i> ”, 1996 (“ <i>Pawate</i> ”)
1039	U.S. Patent No. 4,386,416 to Giltner et al. (“ <i>Giltner</i> ”)
1040	U.S. Patent No. 4,571,700 to Emry, Jr. et al. (“ <i>Emry</i> ”)
1041	U.S. Patent No. 4,646,261 to Ng (“ <i>Ng</i> ”)
1042	U.S. Patent No. 5,150,399 to Yasuda (“ <i>Yasuda</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

I. INTRODUCTION

Petitioners Amazon.com Services LLC and Amazon Web Services, Inc. (“Petitioners”) request *Inter Partes* Review of Claims 46-51, 53-54, 58-59, 64, and 67 (“Challenged Claims”) of USPN 7,587,319, assigned to Zentian Limited upon information and belief. ’319 Patent (EX1001). This Petition raises grounds of unpatentability identical to those in already-instituted IPR2023-00033. 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). Accordingly, IPR of the Challenged Claims should be instituted.

II. SUMMARY OF THE ’319 PATENT

A. Patent Description

The ’319 Patent claims a speech recognition circuit utilizing a plurality of processors connected in parallel for processing speech parameters in parallel, wherein the processors are arranged in groups of processors, and each group of processors is connected to a lexical memory. ’319 Patent, 1:4-6, 4:53-5:13, Fig. 2. The claimed circuit comprises an input buffer connected in parallel to the plurality of processors controlled by a control processor. *Id.*, 3:8-16, 3:29-43.

B. Summary of Unpatentability

Claim 46 of the '319 Patent generally recites five elements: an input buffer, a plurality of lexical memories, a plurality of grouped processors, a control processor, and a results memory. *Thelen* teaches a speech recognition circuit including a plurality of processors controlled by a model selector 360 and controller 350, operable to recognize processed incoming speech parameters using a plurality of recognition models, wherein each of the “recognition models” is targeted to a “specific part of the huge vocabulary” of the system. *Thelen*, 1:53-2:16, 7:30-49, FIG. 3. Each of *Thelen*’s recognition models is described as “having [its] own processing resources,” such as separate processors or processing units in a parallel operating processor. *Thelen*, 2:30-33, 7:17-23.

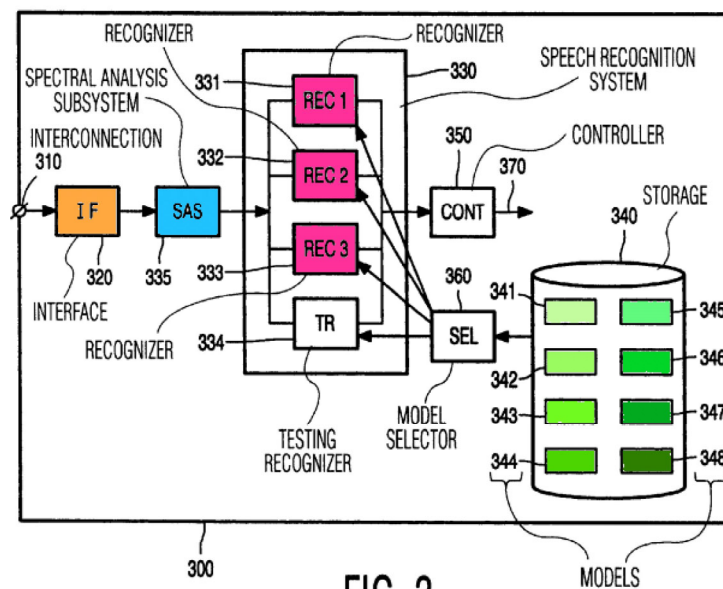


FIG. 3

Thelen, FIG. 3 (annotated).

Thelen does not include express teachings of several of the claimed hardware limitations in Claim 46, such as an input buffer and a results memory. However, such hardware components were extremely well-known for speech recognition systems, as discussed in the detailed mapping. In related art, *Bailey* expressly teaches certain claimed hardware components.

Bailey teaches a similar speech recognition circuit including a plurality of pattern recognition engines (PREs) 525a-c. In a first embodiment of *Bailey*, FIG. 3A, each PRE is connected to a private memory 615, and in a second embodiment, FIG. 3B, a plurality of PREs is connected to a shared private memory 615:

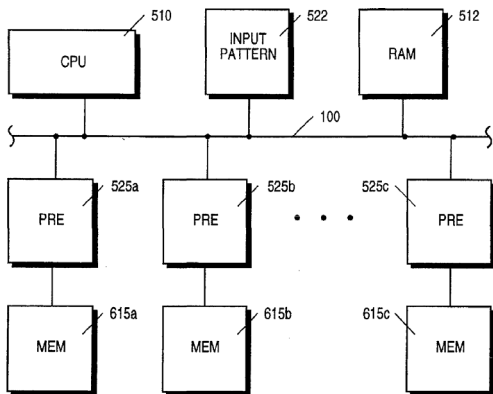


FIG. 3A

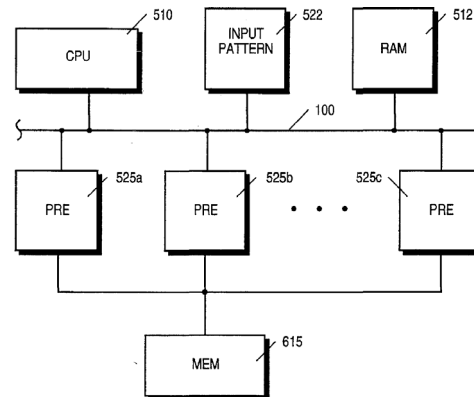


FIG. 3B

Bailey, 10:18–11:37, FIGs. 3A-3B.

Although *Thelen* and *Bailey* each teaches a plurality of processors processing speech parameters, and *Bailey* teaches both (i) a plurality of processors each connected to a lexical memory; and (ii) a plurality of processors connected to a single

lexical memory, *Chen* provides further teachings detailing groups of processors and connected memory. *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. *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 group.

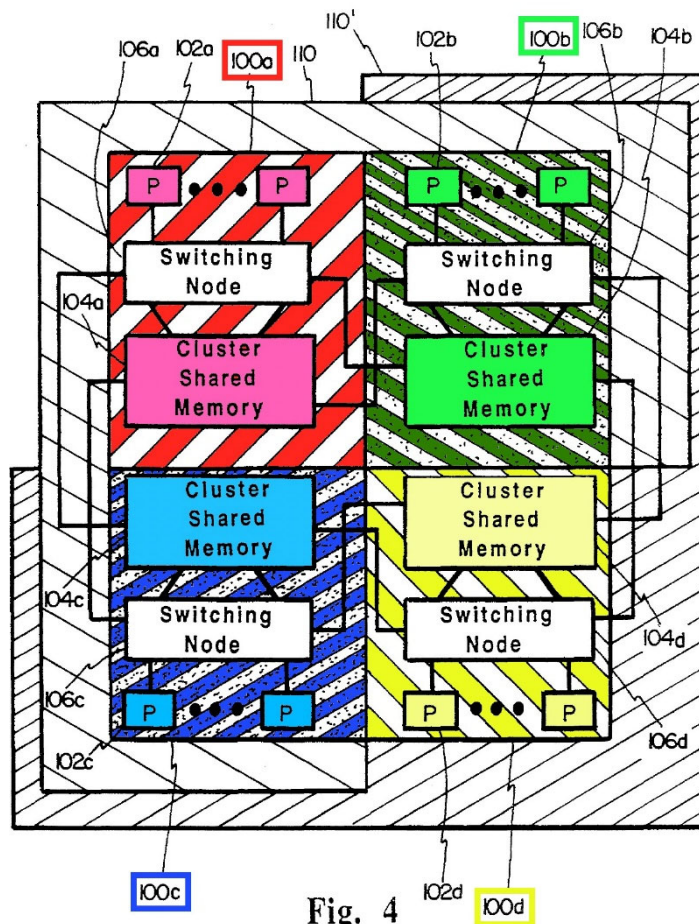


Fig. 4

Chen, FIG. 4 (annotated), 9:5-43.

The speech recognition circuit of *Thelen-Bailey* modified to utilize *Chen's* architecture of a plurality of grouped processors renders obvious every limitation of the Challenged Claims, as detailed below.

C. Priority Date

The '319 Patent claims priority to a GB application filed February 4, 2002. '319 Patent, (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 '319 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 Submitted herewith is the Declaration of Les Atlas, Ph.D., who has reviewed, agrees with, and adopts the opinions in the Schmandt Declaration (IPR2023-00033, EX1003). EX1043, ¶¶ 25-26. For consistency with Apple's Petition, citations to Dr. Atlas's opinion are to paragraph numbers of EX1003 containing those opinions as expressed by Mr. Schmandt.

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

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

Petitioners certify the '319 Patent is available for IPR and Petitioners are not barred or estopped from requesting IPR of the Challenged Claims. Because Petitioners file an accompanying motion for joinder within one month of institution in IPR2023-00033, this petition is timely under 35 U.S.C. §315(c) and 37 C.F.R. §42.122.

B. Identification of Challenges (37 C.F.R. §42.104(b))

In view of the prior art and evidence presented, the Challenged Claims are unpatentable and should be cancelled. 37 C.F.R. §42.104(b)(1).

Proposed Grounds of Unpatentability	
<u>Ground 1:</u>	Claims 46, 50, 54, 64, and 67 are rendered obvious over <i>Thelen, Bailey, and Chen</i>
<u>Ground 2:</u>	Claims 47-49 are rendered obvious over <i>Thelen, Bailey, Chen, and Huang</i>
<u>Ground 3:</u>	Claims 53 and 58-59 are rendered obvious over <i>Thelen, Bailey, Chen, and Jiang</i>
<u>Ground 4:</u>	Claim 51 is rendered obvious over <i>Thelen, Bailey, Chen, and Katz</i>

Sections V-VIII 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-VIII. 37 C.F.R. §42.104(b)(5). Exhibits 1001–1044 are also attached.

C. Claim Construction (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 '319 Patent are indefinite or additional terms need construction.

IV. SHOWING OF ANALOGOUS PRIOR ART

A. *Thelen*

Thelen was neither cited nor considered during prosecution of the '319 Patent. U.S. Patent No. 6,526,380 to Thelen et al. ("*Thelen*") filed August 9, 1999, issued February 25, 2003, and thus qualifies as prior art under 35 U.S.C. §102(e) Pre-AIA. *Thelen* describes a speech recognition circuit under the control of a controller that uses lexical data for speech recognition divided into parts. *Compare Thelen*, 1:53-2:16, 7:30-58 *with '319 Patent*, 3:39-43, 10:59-11:5; *Dec.*, 62. Thus, *Thelen* is analogous art.

B. *Bailey*

Bailey was neither cited nor considered during prosecution of the '319 Patent. U.S. Patent No. 5,459,798 to Bailey et al. ("*Bailey*") issued October 17, 1995, and thus qualifies as prior art under 35 U.S.C. §102(b) Pre-AIA. *Bailey* describes a speech recognition circuit using multiple processors connected in parallel to an input buffer and controlled by a control processor. *Compare Bailey*, 9:30-39, 10:59-11:37, *with '319 Patent*, 4:53-5:4, 5:25-31; *Dec.*, 63. Thus, *Bailey* is analogous art.

C. *Chen*

Chen was neither cited nor considered during prosecution of the '319 Patent. U.S. Patent No. 5,428,803 to Chen ("*Chen*") published June 27, 1995 and thus qualifies as prior art under 35 U.S.C. §102(b) Pre-AIA. *Chen* describes a parallel processing computing platform utilizing clustered processors. *Compare Chen*, 5:9-17, 9:5-43, Fig. 4 *with '319 Patent*, 3:13-24, 5:6-36, Fig. 2; *Dec.*, 64. Thus, *Chen* is analogous art.

D. *Huang*

Huang was neither cited nor considered during prosecution of the '319 Patent. U.S. Patent No. 6,076,056 to Huang et al. ("*Huang*") filed September 19, 1997, issued June 13, 2000, and thus qualifies as prior art under 35 U.S.C. §102(b) Pre-AIA. *Huang* describes a speech recognition circuit utilizing an HMM acoustic

model and lexical trees. *Compare Huang*, 1:5-39, 2:32-52 with '319 Patent, 1:30-44, 4:53-5:20, 6:41-48; *Dec.*, 65. Thus, *Huang* is analogous art.

E. *Jiang*

Jiang was neither cited nor considered during prosecution of the '319 Patent. U.S. Patent No. 6,374,219 to Jiang ("*Jiang*") filed February 20, 1998, issued April 16, 2002, and thus qualifies as prior art under 35 U.S.C. §102(e) Pre-AIA. *Jiang* describes a speech recognition circuit utilizing an HMM acoustic model and lexical trees. *Compare Jiang*, 6:17-28, 7:29-54, 9:28-42 with '319 Patent, 1:30-43, 4:53-5:20, 6:41-48; *Dec.*, 66. Thus, *Jiang* is analogous art.

F. *Katz*

Katz was neither cited nor considered during prosecution of the '319 Patent. U.S. Patent No. 4,831,550 to Katz ("*Katz*") issued May 16, 1989, and thus qualifies as prior art under 35 U.S.C. §102(b) Pre-AIA. *Katz* describes a speech recognition circuit utilizing a language model processor. *Compare Katz*, 6:33-41, 2:34-44, Figs. 2-3, with '319 Patent, 5:21-46, 7:64-8:9, 8:52-9:6, Fig. 2; *Dec.*, 67. Thus, *Katz* is analogous art.

**V. GROUND 1: CLAIMS 46, 50, 54, 64, AND 67 ARE OBVIOUS OVER
THELEN IN VIEW OF *BAILEY* IN FURTHER VIEW OF *CHEN***

A. Claim 46

1. Claim 46(Pre): “A speech recognition circuit comprising:”

a) *Thelen*’s Teachings

To the extent the preamble is limiting, *Thelen* teaches or at least renders obvious a speech recognition circuit, as claimed. *Thelen*, 1:5-12, 6:46-66, Abstract, Figs. 1, 3; *Dec.*, 69. *Thelen* teaches a speech recognition apparatus including a microphone, A/D converter, and processors.

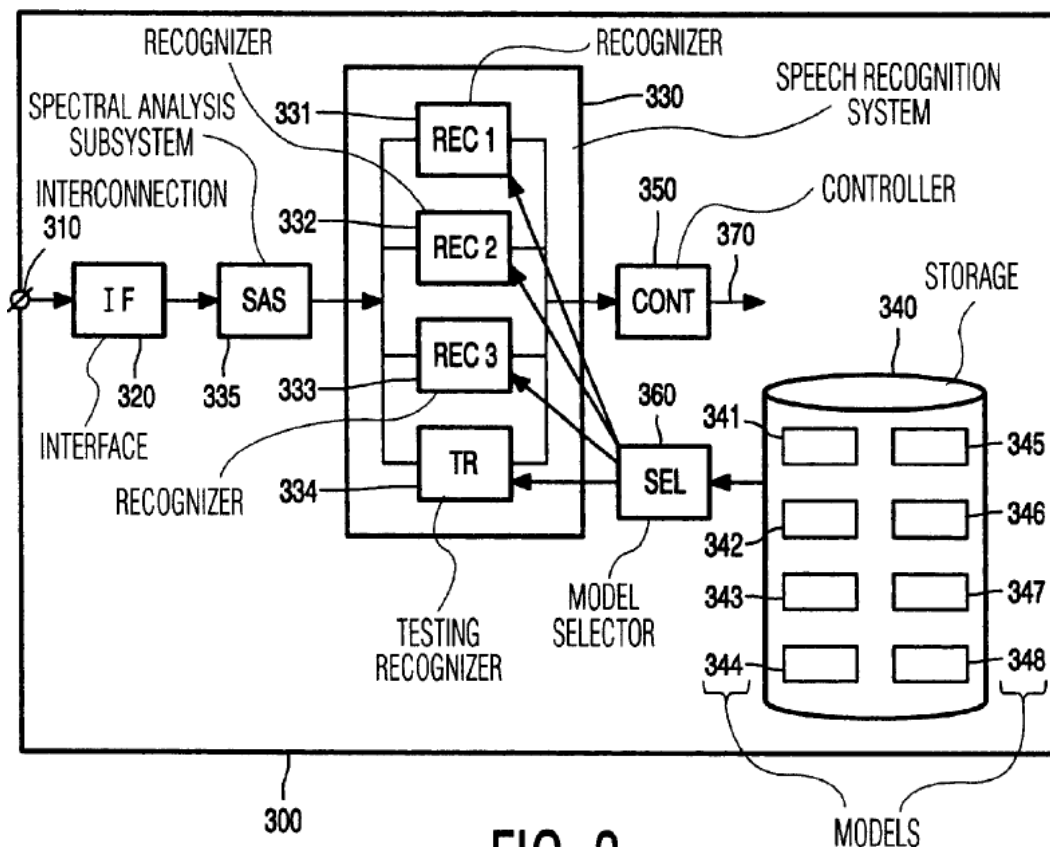


FIG. 3

Thelen, 2:30-37, 6:46-66, 7:20-26, Fig. 3. *Thelen*'s speech recognition system may be implemented on a typical PC. *Thelen*, 6:46-47. As discussed in the following section, *Thelen*'s speech recognition system renders obvious a speech recognition *circuit*.

b) Obviousness of Speech Recognition “circuit”

To the extent Patent Owner argues recited elements of the Challenged Claims (including Claim 46(Pre)) must be embodied as a “circuit,” it would have been obvious to a POSITA to implement any of the computerized techniques or physical architectures described in the cited references using one or more “circuits.” *Dec.*, 68. 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.*, 68. 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.*, 68. 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.*, 68.

A POSITA would have understood *Thelen's* speech recognition system implemented on a computer teaches or at least renders obvious a “speech recognition circuit.” *Dec.*, 68-69. Given it was generally obvious to configure computerized functions in circuitry, and further given that *Thelen* expressly teaches its invention may be implemented on a “PC,” it would have been obvious to a POSITA to implement *Thelen's* speech recognition functionality as circuitry. *Dec.*, 69.

2. *Claim 46(a): “an input buffer receiving processed speech parameters”*

a) **The '319 Patent's Description of “Processed Speech Parameters”**

The '319 Patent describes “processed speech parameters” as “encompass[ing] any type of parameterization of speech data as an input to the circuit,” at least including feature vectors. *'319 Patent*, 1:7-11, 4:48-52, 11:34-38, 4:12-13, 4:32-47. In one example, it describes incoming speech may be sampled via an “analogue-to-digital [A/D] converter 3” then processed into normalized frames to produce feature vectors. *Id.*, 4:32-37. The '319 Patent expressly characterizes this description as “purely given for illustration,” stating “the present invention encompasses **any means by which speech and data can be parameterized** to a suitable form for input to the search process.” *Id.*, 4:48-52.²

² All emphases added unless otherwise noted.

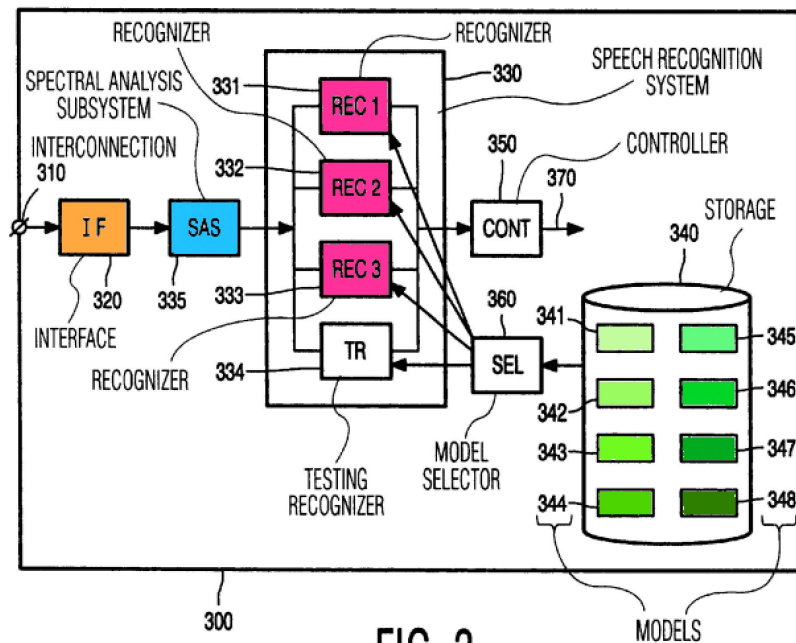
b) “processed speech parameters”

Thelen teaches “processed speech parameters,” such as observation vectors (also referred to in *Thelen* as feature vectors). *Thelen* teaches a “spectral analysis subsystem 110” in which a speech input signal is “spectrally and/or temporally analyzed to calculate a representative vector of features.” *Thelen*, 4:29-33. An A/D converter samples analog speech to digital samples suitable for further “processing.” *Thelen*, 4:34-36, 6:55-66. The samples are blocked into frames to produce a “representative vector of features (observation vector, OV),” which *Thelen* further identifies as a “feature vector[.]” *Thelen*, 4:34-43, 4:66–5:2 (discussing estimating a probability of an observation vector for a given word string). Because *Thelen* teaches generating observation (feature) vectors, *Thelen* teaches “processed speech parameters.”

Thelen further teaches the spectral analysis subsystem 110 is used in the embodiment of the standalone speech recognition system 300 of Fig. 3:

For a typical recognition as shown in FIG. 1, the individual recognizers can share the model-independent **spectral analysis subsystem 110 of FIG. 1 as shown in FIG. 3 under number 335.**

Thelen, 6:63–7:6. Therefore, the discussion of SAS 110 at *Thelen*, 4:29-36 is applicable to SAS 335 of the FIG. 3 embodiment:



Thelen, FIG. 3 (annotated).

c) “an input buffer for receiving [processed speech parameters]”

(1) *Thelen’s Teachings*

As in *Thelen’s* FIG. 3 below, speech is first input to interface 320 and then transmitted to Spectral Analysis Subsystem 335 where the speech is converted to observation (feature) vectors comprising processed speech parameters, as discussed in Section V.A.2.b. *Thelen*, 6:53-66. Processed speech parameters are then provided in parallel to a plurality of speech recognizers 331-333, described in further detail below with respect to Claim 46(c). *Thelen*, 7:6-7, 7:17-29, FIG. 3.

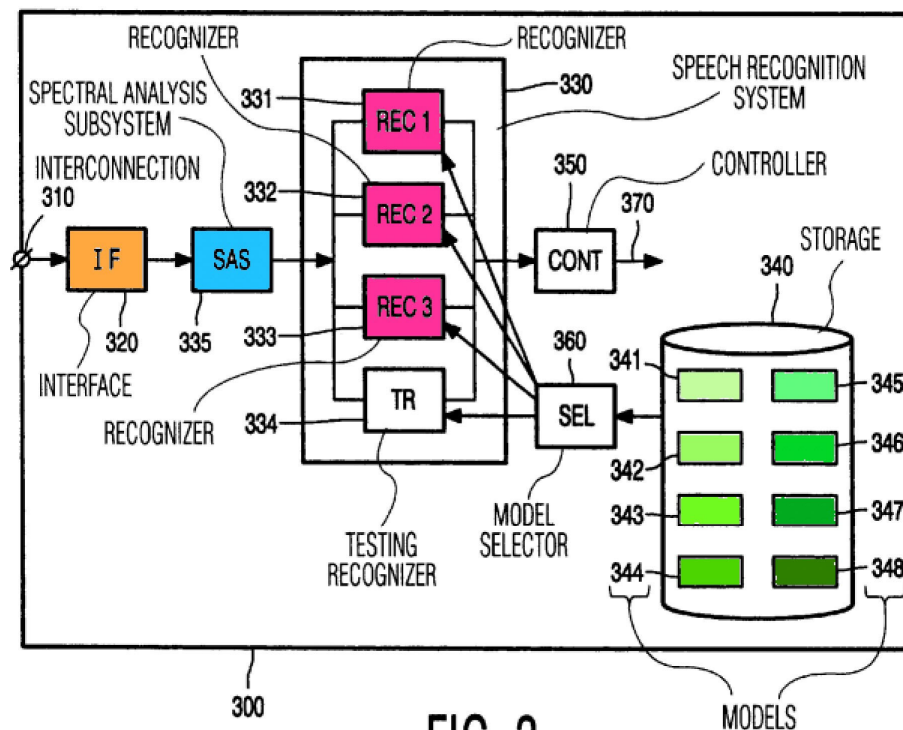


FIG. 3

Thelen, Fig. 3 (annotated).

Thelen illustrates the observation vectors OV (i.e., the claimed “processed speech parameters”) transmitted from the spectral analysis subsystem (identified in FIG. 1 as [RN 110](#) and in FIG. 3 as [RN 335](#)):

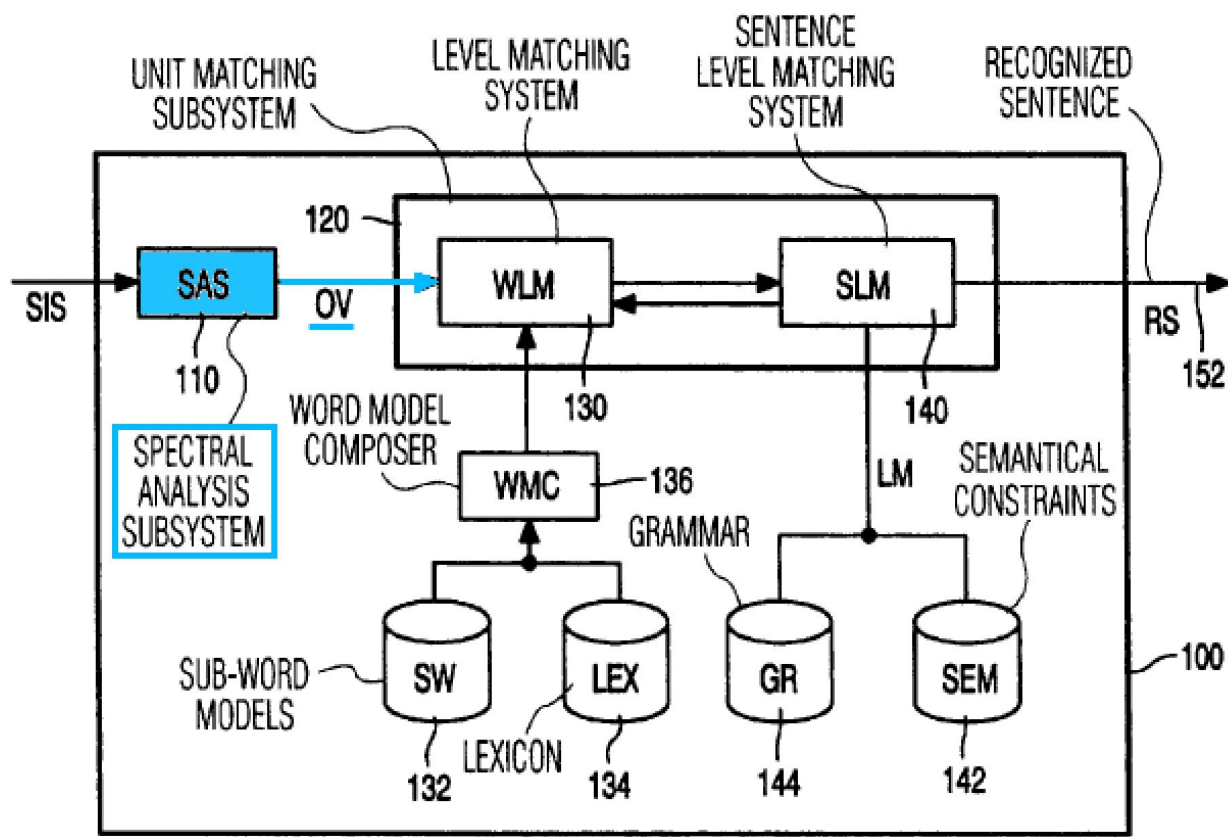


FIG. 1

Thelen, FIG. 1 (annotated).

It would have been obvious and a POSITA would have been motivated to modify *Thelen* to include an input buffer for receiving the observation vectors generated from the SAS (110 of FIG. 1, 335 of FIG. 3, henceforth referred to as SAS 110/335), as further discussed below regarding *Bailey. Dec.*, 70-75.

(2) **Bailey's Teachings**

Thelen does not expressly recite an “input buffer” receiving the processed speech parameters (generated OV's). In related art, *Bailey* teaches such, namely RAM 512.

Bailey teaches systems and methods for pattern recognition in handwriting and voice recognition systems. *Bailey*, Abstract, 10:59-65. A standard computer system 500 may be “optimized for pattern recognition by a specialized multiprocessor.” *Bailey*, 9:5-8. The system 500 includes a signal generation device 522 including an input device, microphone, and analog to digital converter (“digitizer”) “for capturing sounds into analog signal form and transforming those sounds into digital data which the general purpose computer system may process.” *Bailey*, 9:30-35. Figure 2 illustrates system 500 including CPU 510, RAM 512, and a pattern recognition engine 525:

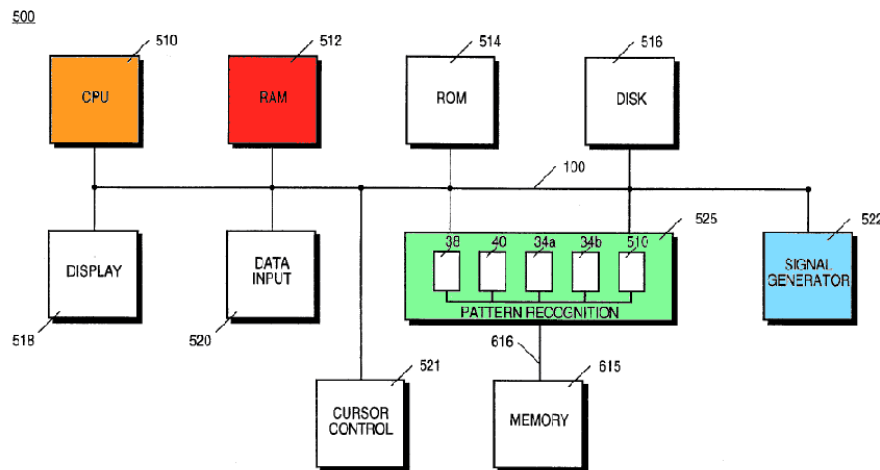


FIG. 2

Bailey, 9:1-39, FIG. 2 (annotated), 14:24-29 (referring to memory unit 512 as CPU memory).

Bailey, like *Thelen*, teaches pre-processing speech input to extract features and normalize components. Specifically, *Bailey* teaches “recognition procedures include three major processing steps: 1) **a preprocessing or feature extraction and normalization step**; 2) a search and match step; and 3) a post processing and final decision step.” *Compare Bailey*, 12:12-15, *with Thelen*, 4:20-33, 5:2-53, 7:50-58. *Bailey* recognizes the search and match step is the most “processing intensive” and therefore, delegates the search and match step to the pattern recognition engine 525 while the CPU 510 performs the “higher level” tasks such as pre-processing and post-processing. *Bailey*, 11:42–12:25, 12:20-21 (“pre-processing and post-processing steps are accomplished by the CPU 510”). Thus, speech is preprocessed by CPU 510, and the preprocessed speech is sent to the PRE 525 for pattern matching. *Bailey*, 11:15-37. Also similar to *Thelen*, *Bailey* teaches performing pattern recognition using Hidden Markov Models (HMMs), where the “unknown pattern is represented by an ordered set of states.” *Compare Bailey*, 7:51–8:36, pinpoint at 7:59-60, FIG. 1B, *with Thelen*, 5:29-45.

Bailey teaches performing pattern recognition on unknown sounds stored in RAM:

The signal generator 522 also includes an [sic] mechanism for **isolating unknown sound segments from the sampled digital data and storing those unknown sound segments into a memory unit such as RAM 512**. These sound segments are one type of unknown pattern that the present invention will process in order to recognize.

Bailey, 9:35-41. *Bailey* further teaches the “unknown input **pattern** may be **temporarily stored** in memory array 512.” *Bailey*, 11:16-17.

A POSITA would have understood or found obvious RAM 512 provides an “input buffer for receiving processed speech parameters,” as claimed, because RAM 512 temporarily stores the unknown input pattern to be compared to known reference patterns. *Dec.*, 72. A POSITA would have recognized *Bailey*’s reference to storing in RAM 512 an “unknown input pattern” (11:16-17) suggests storing preprocessed speech, such as the feature-extracted speech, in RAM. *Dec.*, 72. Thus, the temporarily stored unknown input pattern includes, at the least, the preprocessed speech parameters. *Dec.*, 72.

In more detail, *Bailey* teaches CPU preprocessing speech to extract features, such that preprocessed speech is input to PRE 525. *Bailey*, 12:12-21, 12:34-36 (discussing FIG. 4, block 402, with respect to pattern recognition of handwriting, the CPU preprocessing coordinates and sending a normalized set of features to the PR engine 525). *Bailey* further teaches the “CPU may input up to n separate unknown patterns from block 522 during a single recognition processing cycle, [where] a

separate unknown would be sent to each PR engine over bus 100.” *Bailey*, 11:23-26.

Yet further, *Bailey* specifically teaches temporarily storing the unknown input pattern in RAM 512. *Bailey*, 11:16-17.

Bailey also explains that an unknown input pattern comprises individual HMM states compared to a reference pattern. *Bailey*, 1:45-51 (referring to an “input pattern also composed of a multitude of points or states” and comparing the input pattern to the reference pattern point-by-point or state-by-state). Figure 1B of *Bailey* illustrates an HMM of reference states compared to states of an input unknown pattern. *Bailey*, 4:35-39, FIG. 1B. As discussed above, *Bailey* teaches feature extraction and performing pattern recognition using HMMs, including representing the unknown pattern by a plurality of states. *Bailey*, 7:59-60, 12:12-15. The points or states of the unknown input pattern are determined via feature extraction. *Dec.*, 73 (*citing Bailey*, 49:51-66).

The collective teachings of *Bailey* would have indicated to a POSITA that the unknown input patterns include processed speech parameters (such as feature vectors used for HMM pattern matching) that are temporarily stored in RAM. *Dec.*, 72. Therefore, *Bailey* teaches memory (RAM) for receiving processed speech parameters. *Id.*

Additionally, a POSITA would have recognized or found obvious that RAM 512 comprises an input buffer. *Dec.*, 73-74. A POSITA would have known that

RAM was a well-known hardware component for providing a “buffer” prior to the ’319 Patent. Mr. Schmandt provides background references establishing that RAM was widely used in the art for providing input buffers, explaining it was known in the art to employ buffers temporarily storing processed speech parameters, such as feature vectors, prior to pattern matching. *Dec.*, 74. A buffer was well-known to provide temporary storage in memory for passing data between computing components, modules, or stages, similar to the description of *Bailey* of passing data from its input pattern block 522 (11:11-17) to the plurality of pattern recognition (PR) engines 525a-c via temporary storage in RAM 512. *Dec.*, 70. Thus, a POSITA would have understood *Bailey*’s unknown input pattern comprising feature vectors processed from input speech and temporarily stored in RAM 512 teaches the claimed “input buffer for receiving processed speech parameters.” *Dec.*, 72-74.

d) Motivation to Combine *Bailey* with *Thelen* for Claim 46(b)

A POSITA would have found it obvious and been motivated to configure the speech recognition circuit taught by *Thelen* to specifically store *Thelen*’s feature vectors in an input buffer, as taught by *Bailey*. *Dec.*, 70. Buffers were well-known, inexpensive, and easy-to-implement solutions to improve system efficiency. *Dec.*, 74 (*citing* numerous background references teaching utilizing buffers for temporary storage of data). A POSITA would have recognized input buffers were also well-known and widely used prior to the ’319 Patent as a simple, operable structure for

temporarily storing generated data, such as feature vectors generated by *Thelen's* spectral analysis subsystem 110/335, prior to input to a speech recognizer, such as *Thelen's* plurality of speech recognizers, leading to a reasonable expectation of success for the combination. *Dec.*, 70, 74-75 (*citing Pawate*). A POSITA would have appreciated an input buffer would have improved the efficiency of *Thelen's* circuit by allowing the processed feature vectors (a form of unknown speech patterns, in *Bailey's* description) to be stored for subsequent recognition processing (further discussed below) as soon as recognition resources are available, decoupling the rate of feature vector processing from the rate of recognition, further motivating the combination. *Dec.*, 75. The necessary modifications to *Thelen* to include an input buffer communicatively coupled between *Thelen's* SAS 110/335 and speech recognition system 330 to receive and store *Thelen's* observation vectors (OVs) would have only required RAM (a form of digital "storage," already described by *Thelen* and ubiquitous prior to the '319 Patent) and software changes well within the skill of a POSITA at the time. *Id.*

For example, *Thelen* teaches consecutive samples of digitized speech input are blocked or grouped into overlapping frames, where a feature vector is determined for each frame of speech. *Thelen*, 4:30-43. The frames are then transmitted to the plurality of recognizers, such that the recognizers receive the processed speech parameters. *Dec.*, 75; FIGs. 1 and 3 (collectively illustrating OVs transmitted from

SAS to recognizers). A POSITA would have recognized the use of an input buffer, such as that disclosed by *Bailey*, would have enabled the SAS 110/335 of the modified *Thelen* circuit to sample and block/group frames at its maximum operational rate, loading them into the input buffer for the pattern recognition (PR) engines 525a-c to utilize at *their* maximum operational rate. *Dec.*, 75. Without implementing an input buffer, the combined system would have been bottlenecked, operating at the lesser of the two maximum operational rates. *Id.* Thus, a POSITA would have been motivated to combine the input buffer for increased efficiency and speed of the circuit. *Id.*

3. Claim 46(b): “a plurality of lexical memories containing in combination complete lexical data for word recognition, each lexical memory containing part of said complete lexical data”

As mapped in detail below, *Thelen* teaches a plurality of recognition models that collectively comprise complete lexical data for huge vocabulary recognition. *Thelen*, Abstract. Each of the recognition models is targeted to a “specific part of the huge vocabulary,” where each recognition model comprises “large vocabulary models for specific contexts,” such as health, entertainment, computer, business, etc. *Thelen*, Abstract, 1:66–2:29, 7:39. Each recognition model comprises a lexicon 134 for the “specific part of the huge vocabulary,” such that the collective lexicons for the collective recognition models 341-348 comprises “in combination complete lexical data,” as claimed. *Thelen*, 7:37-40.

Thelen teaches a “storage 340” for storing the recognition models (including the lexicons) but does not teach a *plurality* of lexical memories. *Thelen*, 2:30-33, 7:17-23, FIG. 3. In related art, *Bailey* expressly teaches a plurality of lexical memories.

a) ***Thelen* Teaches a Memory Containing “in combination complete lexical data for word recognition” and “part of said complete lexical data”**

Thelen teaches storing “in combination complete lexical data for word recognition,” as claimed, namely the collective lexicons 134 associated with the plurality of recognition models 341-348. *Thelen* also teaches storing “part of said complete lexical data,” namely each lexicon 134 of each recognition model, where the lexicons (which are part of the recognition model) are “different” from each other and associated with a particular context. *Thelen*, 7:30-34, 1:56-65.

(1) **Overview of *Thelen***

Thelen teaches a speech recognition circuit intended for use with a “huge” vocabulary. *Thelen*, Abstract, 1:5-12, 1:53-55, 6:19-39. *Thelen* teaches a plurality of speech recognizers *each* associated with *a recognition model*, where each recognition model is associated with a “large” vocabulary, which is *a part* of a huge vocabulary for speech recognition:

It is an object of the invention to provide a recognition system which is better capable of dealing with huge vocabularies. [¶] To achieve the object, the system is characterized in that the system comprises a

plurality of N large vocabulary speech recognizers, each being associated with a respective, **different large vocabulary **recognition model**; each of the recognition models being targeted to a specific part of the huge vocabulary....**

Thelen, 1:53–2:3, 7:30-35 (“According to the invention, each of the large vocabulary speech recognizers is associated with a respective, different large vocabulary recognition model, where each of the recognition models is targeted to a specific part of the huge vocabulary.”). The plurality of recognition models 341-348 (collectively comprising the mapped “complete lexical data for word recognition”) are stored in storage 340:

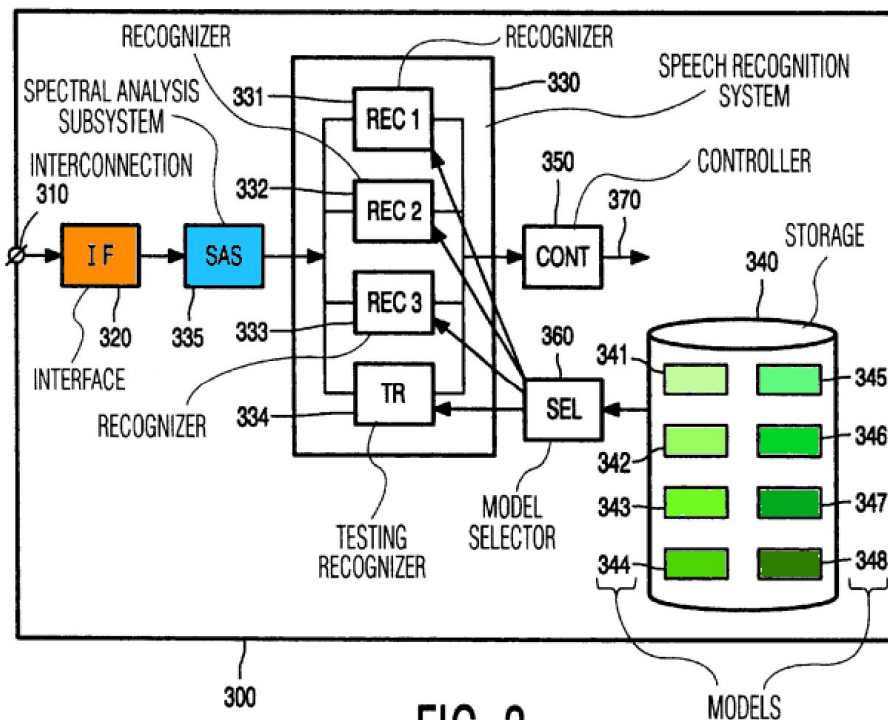


FIG. 3

Thelen, 7:30-40, FIG. 3 (annotated).

Each recognition model comprises a word model and a language model, and each word model 136 comprises **lexicon 134** and sub-word models 132 (such as phones, diphones, or syllables). *Thelen*, 7:30-49.

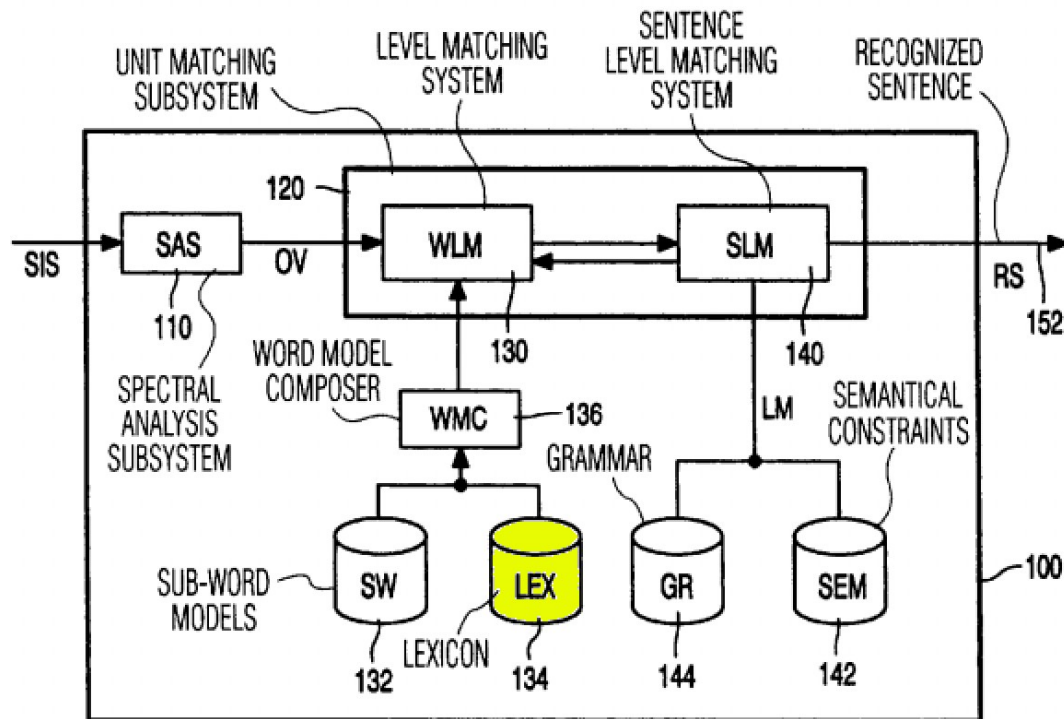


FIG. 1

Thelen, 5:14-22, FIG. 1 (annotated), 6:12-39 (discussing word size of huge vocabulary and noting a lexicon can have up to 60,000 words).

(2) “lexical data”

Each lexicon 134 for each recognition model comprises “lexical data,” as claimed. *Thelen* limits the possible sequences of sub-word units that may be matched to only sequences stored in lexicon 134. *Thelen*, 5:24-45 (explaining the word model

is a Hidden Markov Model (HMM)), 5:49-52 (“If sub-word units are used, **constraints** can be placed on the matching **by using the lexicon 134 to limit** the possible sequence of sub-word units to sequences in the lexicon 134.”). Thus, a POSITA would have understood each lexicon 134 of each recognition model comprises “lexical data,” as claimed, because the lexicon for the speech recognition system is for a huge vocabulary, and the speech recognition system is **constrained** to only identify words stored in the lexicon for the particular recognition model. *Dec.*, 76.

(3) **“in combination complete lexical data”**

A POSITA would also have understood the collection of lexicons 134 for the plurality of recognition models 341-348 comprises “in combination complete lexical data,” as claimed. *Dec.*, 76-77. Each recognizer is associated with a different recognition model, “where each of the recognition models is targeted to a **specific part** of the huge vocabulary.” *Thelen*, 7:30-34. *Thelen* teaches recognition models can be directed to various contexts, where overlap among the recognition models “may and will normally exist between the various recognition models.” *Thelen*, 7:40-42, 8:54-59. *Thelen* also teaches the system includes a “basic recognition model, covering the commonly used vocabulary.” *Thelen*, 8:1-5.

(4) **“for word recognition”**

Thelen’s complete lexical data comprising the collective lexicons 134 is “for word recognition,” as claimed, because *Thelen* expressly teaches recognizing “a sequence of spoken words.” *Thelen*, Abstract, 7:17-23 (discussing recognizing speech), 7:50-52 (discussing recognizing a word sequence).

(5) **“part of said complete lexical data”**

Thelen teaches “part of said complete lexical data,” as claimed, namely each lexicon 134 associated with a specific recognition model. As an initial matter, Petitioners note Claim 46(b) merely recites a “part of said complete lexical data” without any specific requirements for what entails a “part” of lexical data, other than that the “parts” contain, in combination, complete lexical data for word recognition. *Thelen* explains, “[b]y using several recognizers **each with a specific recognition model targeted at a part of the huge vocabulary**, the task of building a recognition model for a huge vocabulary is broken down into the manageable task of building large vocabulary models for specific contexts.” *Thelen*, 1:66–2:3. Example of “contexts” that may be recognized by a particular recognition model include health, entertainment, computers, arts, business, education, government, science, news, and/or travel. *Thelen*, 2:3-5. *Thelen* continues, stating:

[E]ach of those contexts will normally **overlap in vocabulary**, for instance in the **general words** of the language. The contexts will **differ in statistics of those common words as well in the jargon**

specific for those contexts. By using several of those models to recognize the input, a wider range of utterances can be recognized using properly trained models. A further advantage of using several models is that this allows a better discrimination during the recognition. If one huge vocabulary was used, certain utterances would only be recognized in one specific meaning (and spelling).

Thelen, 2:5-16, *see also* 7:30-49 (stating that overlap between recognition models will usually occur in **part** of the vocabulary and that each recognizer may be exclusively associated with a recognition model in a “fixed one-to-one relationship”); *Dec.*, 77.

A POSITA would have recognized each of *Thelen*’s plurality of recognition models and associated lexicons would have each included at least some words that were not present in other recognition models, such as “jargon specific” words for that recognition model’s context. *Id.* (opining that *Thelen*’s description that the overlap of words among recognition models would be “general words of the language” further indicates the non-overlapping words are specific words to the context for the recognition model). Because each lexicon for each recognition model is different, at least in part, from the other lexicons for the other recognition models, each lexicon 134 comprises “part of said complete lexical data,” as claimed. *Dec.*, 76-77.

Because (1) *Thelen* is directed to “huge vocabulary speech recognition” (*Thelen*, 1:5-12), (2) each of the recognition models is “targeted at a part of the huge vocabulary” (1:66–2:12), (3) the recognition models (and included lexicons) collectively allow for recognizing the huge vocabulary according to “a wider range of utterances” (*id.*), and (4) at least one of the recognition models comprises basic words “covering the commonly used vocabulary” (8:1-5), *Thelen* teaches or otherwise renders obvious the collective lexicons 134 of the plurality of recognition models comprises “in combination complete lexical data,” as claimed, and that each lexicon is a “part of said complete lexical data.” *Dec.*, 76-77; *Thelen*, FIG. 3 (RN 341-348), 7:49.

(6) Lexical Memory

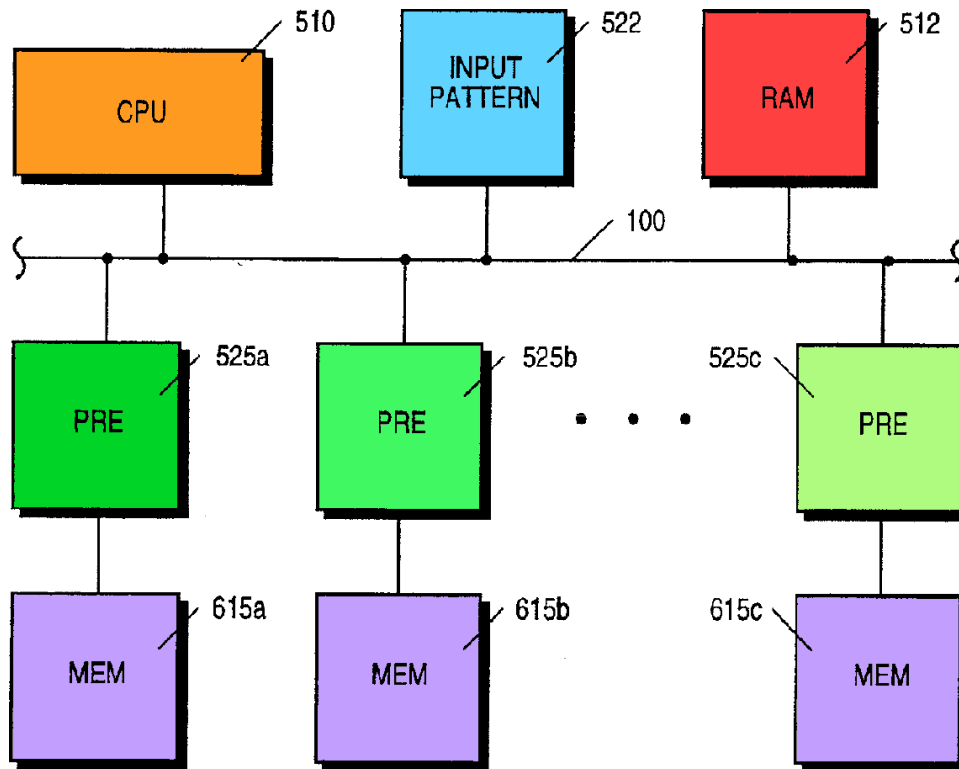
Thelen also teaches a memory containing the complete lexical data. *Thelen* expressly states its recognition models, which are formed of lexicon 134 and sub-word models 132, are “loaded from a storage 340.” *Thelen*, 7:30-34, 5:22-24, FIG. 3; *Dec.*, 78 (opining that because the recognition models are loaded from storage, the recognition models are stored in storage 340). Thus, a POSITA would have recognized *Thelen* teaches storage 340 “contains” in combination “complete lexical data.” *Dec.*, 77 (opining lexicon 134 within storage 340 contains “lexical tree data,” which is not specifically required to be organized as a “lexical tree” until

dependent Claim 47). Section V.A.3.b, below, maps *Bailey* for teaching the claimed “plurality of lexical memories.”

b) *Bailey* Teaches a “plurality of lexical memories”

Thelen teaches storing the recognition models, with each model comprising a lexicon, though does not expressly teach that each of its recognition models and associated lexicons (the mapped “parts” of said complete lexical data) are stored in a plurality of lexical memories. *Thelen* does suggest, however, that “[p]referably, the recognizers operate in parallel in the sense that the user does not experience a significant delay in the recognition. This may be achieved using separate recognition engines **each having own processing resources.**” *Thelen*, 2:30-33. *Thelen* envisions each recognizer (including its processing resources, such as processors and memory) may be associated with an exclusive recognition model in a “fixed one-to-one relationship.” *Thelen*, 7:42-49; *Dec.*, 78.

In related art, *Bailey* teaches a “high speed speech recognition” circuit “having the processing power and efficiency to realize real-time accurate pattern recognition” using a plurality of pattern recognition engines 525 in parallel, “splitting the processing work.” *Bailey*, Abstract, 1:16-20, 2:62-66, 10:18-44–11:8, Fig. 3A. *Bailey* teaches an embodiment, illustrated below in Figure 3A, in which “each PR engine 525a-525c [is] interfaced with its own private memory 615a-615c.”



Bailey, 11:1-34 (pinpoint at 11:1-3), 10:18-11:37, 13:25-51, Fig. 3A (annotated); *Dec.*, 79. Each **pattern recognizer** compares, in parallel, frames of unknown speech data to “its own library of prototype patterns” stored in its **private memory**. *Bailey*, 11:9-29, 10:30-44. Thus, *Bailey* teaches a **plurality of memories (615a-c)** each associated with a **processing resource (PREs 525a-c)**, wherein each of the **memories** stores a “library of reference patterns” for recognizing speech. *Bailey*, 10:18-11:44, 11:9-29, 13:25-51, Fig. 3A; *Dec.*, 79.

Bailey also teaches, for the FIG. 3B embodiment, that each of the PREs 525a-c may be interfaced with a single memory 615. *Bailey*, 11:1-4, FIG. 3B. In the proposed modification to *Thelen* employing groups of processors as mapped for

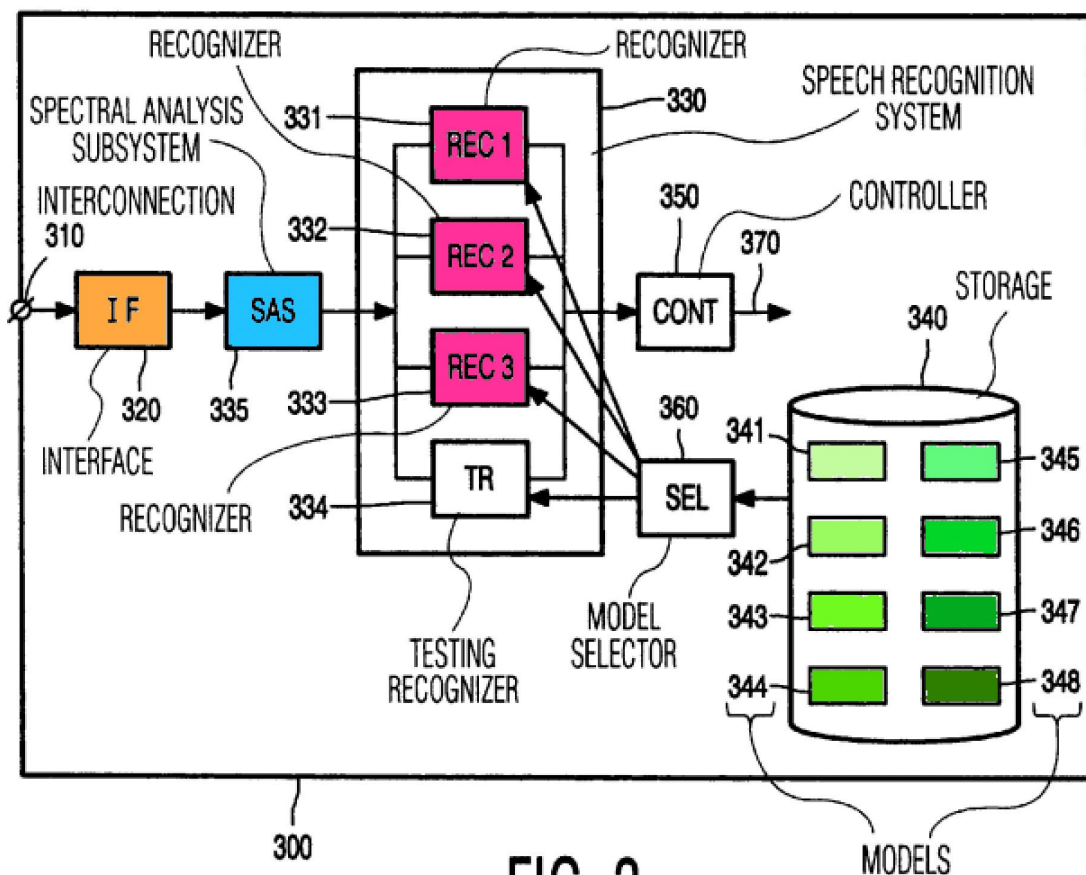
Claim 46(d), it would have been obvious to group a plurality of processors together and connected to a lexical memory, as illustrated in *Bailey*, FIG. 3B and discussed further below for Claim 46(d).

A motivation to combine *Bailey's* plurality of lexical memories with *Thelen* is provided in Section V.A.5.b.

4. Claim 46(c): “a plurality of processors connected in parallel to said input buffer for processing the speech parameters in parallel”

The combination of *Thelen-Bailey* teaches or renders obvious Claim 46(c). Both *Thelen* and *Bailey* teach a plurality of processors connected in parallel, and *Thelen-Bailey* teaches a plurality of processors connected in parallel to the input buffer.

In more detail, *Thelen* teaches a plurality of recognizers 331-333 (also referred to as “recognition engines”) connected in parallel, where each recognizer includes a processor:



Thelen, FIG. 3 (annotated). As shown in FIG. 3, speech recognition system 330 includes a plurality of speech recognizers REC1-REC3 identified by RN 331-333 and connected in parallel. *Thelen*, FIG. 3, 7:6-9, 2:38-40 (stating “using parallel speech recognition engines is known”), Title (“Speech Recognition System Having **Parallel Large Vocabulary Recognition Engines**”).

Thelen teaches each recognizer operating in parallel includes a separate processor:

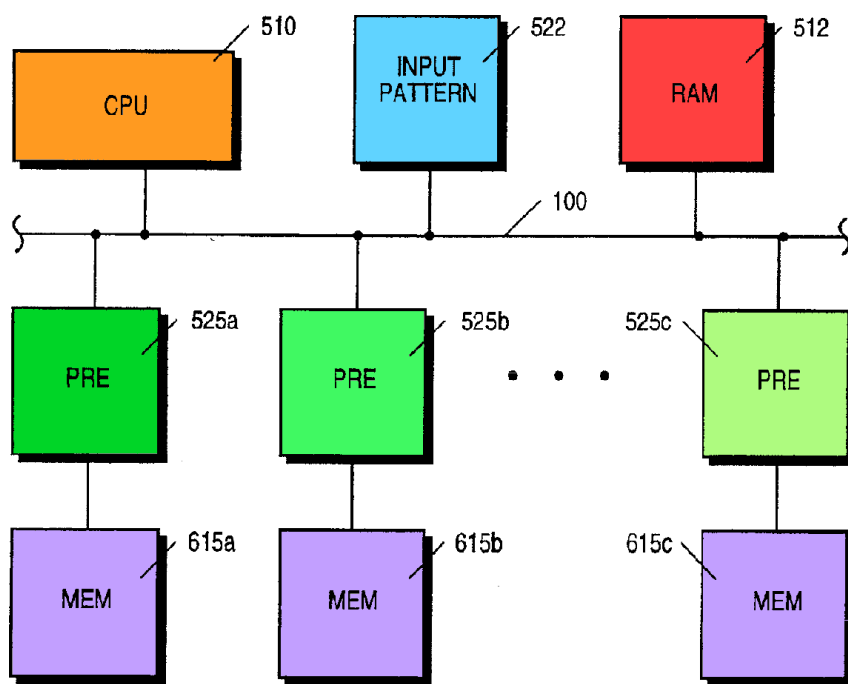
The recognizers operate in ‘parallel’, in the sense that they independently recognize the same speech input at almost the same moment. This can be realized by using separate resources for each of the recognizers, such as **a separate processor** or processing unit in a ‘parallel’ operating processor, such as a **VLIW processor**.

Thelen, 7:17-23. Because each recognizer has separate resources, including a separate processor such as a VLIW (very long instruction word) processor, *Thelen* teaches “a plurality of processors connected in parallel,” as claimed. *Dec.*, 81.

Each of *Thelen*’s processors “process[es] the speech parameters in parallel,” as claimed. As quoted above, the recognizers (which utilize a plurality of processors) operate in parallel in that input speech is recognized at almost the same moment. *Thelen*, 7:17-23. *Thelen* teaches “[p]referably, the recognition is ‘real-time’ in the sense that no significant delay occurs in the recognition of a word, after the word has been received in the system.” *Thelen*, 7:26-29. *Thelen* further teaches the “individual recognizers can share the model-independent spectral analysis subsystem 110,” where FIGs. 2-3 collectively illustrate the observation vectors OV (i.e., the claimed “speech parameters”) being transmitted to the speech recognizers 331-333. *Thelen*, 7:2-7, FIGs. 2-3. Because *Thelen* teaches operating on the input speech “in parallel,” such that each recognizer independently recognizes speech, and further because *Thelen* teaches inputting observation vectors (i.e., speech

parameters) to the recognizers, a POSITA would have understood or found obvious *Thelen* teaches the plurality of processors “process[] the speech parameters in parallel,” as claimed. *Thelen*, 7:17-23. *Dec.*, 81.

Similarly, *Bailey* teaches a plurality of pattern recognition engines 525a-c connected in parallel:



Bailey, 11:23-29, FIG. 3A (annotated). As shown in FIG. 3A, each of the PR engines 525a-c is connected in parallel to RAM 512 (comprising the input buffer, as mapped above for Claim 46(a)) via communication bus 100. *Dec.*, 81. As discussed regarding Claim 46(a), *Bailey*’s RAM 512 comprises an input buffer connected to a plurality of pattern recognition (PR) engines 525a-c for comparison to a plurality of known prototype patterns. *Bailey*, 3:60-4:1, 9:36-41, 11:9-29, Figs. 2, 3A; *see also* Section

V.A.2. Each of the PREs comprises a processor performing processing via execution block 417. *Bailey*, 13:25-51, 14:15-29; *Dec.*, 81. Per the mapping for Claim 46(a), the modified *Thelen* speech recognition circuit includes *Bailey*'s RAM comprising an input buffer for receiving the processed speech parameters (e.g., observation (feature) vectors as taught by *Thelen*) from *Thelen*'s SAS 110/335. See Section V.A.2.c.

Thus, in the modified circuit of *Thelen-Bailey*, *Thelen*'s plurality of recognizers (with each recognizer including its own processor) is connected in parallel to *Bailey*'s RAM 512 (which includes the claimed input buffer), and where the plurality of *Thelen* recognizers process the speech parameters (OVs) in parallel. *Dec.*, 81.

Additionally or alternatively, *Bailey*'s plurality of PREs (which is each a programmable multiprocessor) is arranged in parallel and connected to *Bailey*'s RAM 512.

A POSITA would have recognized *Bailey*'s memory architecture would have combined naturally with the plurality of parallel "recognition engines" taught by *Thelen* and furthered *Thelen*'s expressly-stated goal of operating in "real-time." *Thelen*, 2:30-37, 7:26-29; *Dec.*, 82. A POSITA would have recognized that the necessary modifications to *Thelen* would have only required arrangement of well-known memory components and relocation of lexical data already present in

Thelen, leading a POSITA to a reasonable expectation of success for the combination. *Dec.*, 82. Further, a POSITA would have recognized the combination of *Thelen* and *Bailey* would have constituted combining prior art elements according to known methods without undue experimentation to yield the predictable result of a plurality of speech recognizers operating in parallel using lexical data stored in a plurality of lexical memories. *Id.*

5. ***Claim 46(d): “said processors being arranged in groups of processors, each group of processors being connected to a [] memory”***

a) ***Chen Teaches Claim 46(d) Except for a “lexical” Memory***

Chen teaches a plurality of clusters 100, where each cluster of the plurality of clusters groups a plurality of processors 102 with a cluster shared memory 104. As seen below in FIG. 4, clusters 100a-d comprise groups of processors 102a-d that are symmetrically connected to a cluster shared memory 104a-d via connection nodes 106a-d:

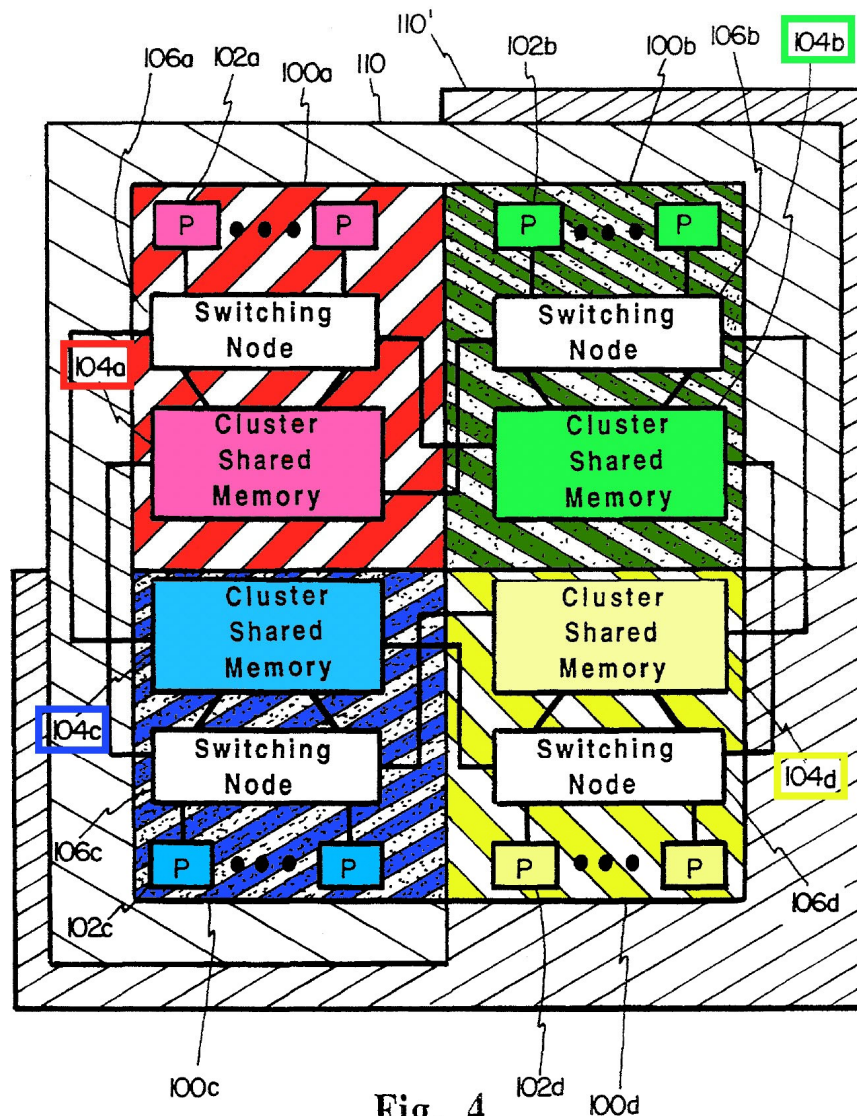


Fig. 4

Chen, FIG. 4 (annotated), 9:5-19; *Dec.*, 83. The resulting computing architecture is described by *Chen* 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. Thus, *Chen* teaches the claimed “processors being arranged in groups of processors, each group of processors being connected to a memory.” *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.*, 83.

Chen does not specifically teach a “lexical” memory. In the proposed modification to the *Thelen-Bailey* circuit, each group of processors (embodied in recognizers) includes a memory, where the memory is a *lexical* memory as already taught by *Thelen-Bailey*.

b) Motivation to Combine *Chen* with *Thelen-Bailey* for Claims 46(b) and 46(d)

The proposed modification to *Thelen* is grouping *Thelen's* recognizers 331-333 in groups of recognizers (where each recognizer has a processor as already taught by *Thelen*) and connecting each group of recognizers (and associated processors) to a memory, all of which is taught by *Chen* and mapped for Claim 46(d). *Bailey* teaches a plurality of lexical memories 615a-c (*see Bailey*, FIG. 3A) and connecting a group of recognition engines with a lexical memory (FIG. 3B). Thus, in the modified *Thelen*, each group of recognizers is connected to a lexical memory, such that the plurality of groups of recognizers is respectively connected to the plurality of lexical memories comprising, in combination, complete lexical data.

A POSITA would have been motivated to and found obvious modifying *Thelen* to include grouping recognizers together for the purpose of associating the generic and specific contexts (e.g., all sports contexts) together in a group for efficient word recognition. *Thelen* teaches multiple recognition models, where each

model is associated with a context, and there may be generic contexts (e.g., sports) and subdivisions of generic contexts (e.g., football, baseball, etc.). *Thelen*, 8:6-8, 8:54-56 (discussing “recognition models are hierarchically arranged from models with a more generic context to models with a more specific context”), 8:54–9:32, pinpoint at 9:24-26 (discussing specific sports models such as football, baseball, etc.). *Thelen* discusses how the speech recognition system utilizes multiple recognizers, a plurality of which may be active at a given time, each active recognizer independently operating on the same speech input at almost the same moment. *Thelen*, 7:17-58. *Thelen* also teaches that a “basic recognition model, covering the commonly used vocabulary, is always active.” *Thelen*, 8:1-3. Finally, *Thelen* teaches that the “output of the recognizers is directed to controller 350 to make the final selection of a recognized word sequence,” accounting for the context to which each active recognizer is directed.

A POSITA would have been motivated and found obvious to group a plurality of *Thelen*’s recognition processors, as taught by *Chen*, according to the particular context of the assigned active recognition model, so as to facilitate more efficient processing and retrieval of stored lexical data within each recognition model. *Dec.*, 84-85, 80. Additionally, a POSITA would have found it obvious to store all similar contexts on a lexical memory connected to the group of recognizers directed to those contexts. *Id.*

For example, *Thelen* already teaches recognition models associated with a particular context and ranging from generic to specific contexts for a particular topic, such as sports, as discussed above. *Thelen* also teaches that each recognizer is associated with one recognition model at a time. *Thelen*, 7:30-34. It would have been obvious to group a plurality of recognizers together to access the same memory storing recognition models for similar contexts, such as “sports,” to facilitate more efficient control and processing. Indeed, *Thelen* already teaches grouping the recognition models according to the hierarchical association. *Thelen*, 8:54-56. For example, FIG. 4 illustrates generic recognition models 410-440, and subdivisions of the generic models 422, 432, 424.

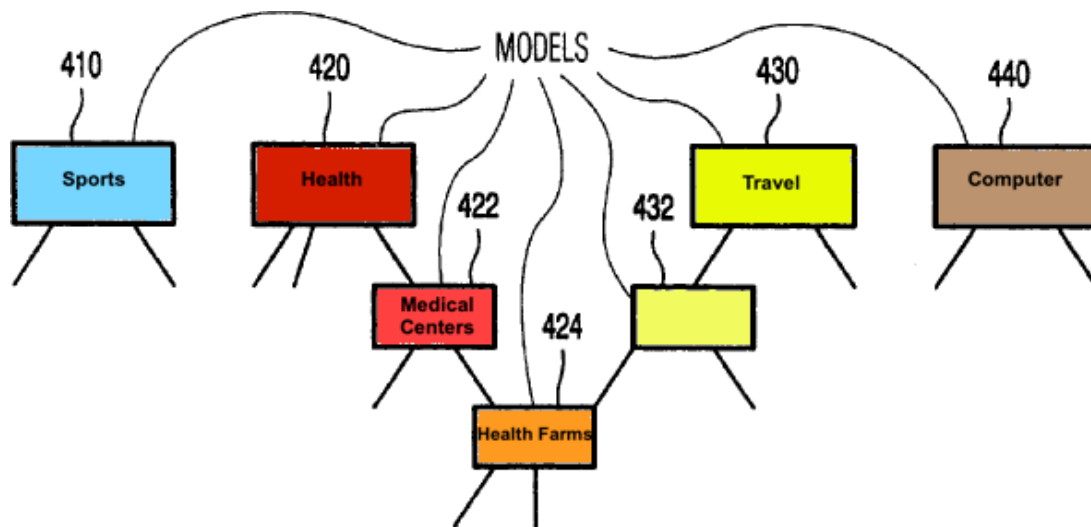


FIG. 4

Thelen, FIG. 4 (annotated to include text and color). In the modified *Thelen-Bailey* circuit and as an exemplary use case only, a first group of processors is associated

with the “sports” context, such that the lexical memory (e.g., memory 615 taught in *Bailey*) stores the recognition models for the sports contexts. A second group of processors relates to the “health” context, where, as taught by *Thelen*, one of the subdivisions (model 422) is “hospitals/medical centers.” *Thelen*, 8:62-66. An even further subdivision (model 424) may be “health farms.” *Thelen*, 8:65–9:1. *Thelen* notes how the “health farms” context may also be related to travel below model 432, stemming from a “travel” context. *Thelen*, 9:1-11.

In the modified *Thelen* system, each grouping of recognizers (and associated processors) stores the recognition models associated with a particular context (e.g., all sports-related recognition models) in cluster shared memory (per *Chen*), where the cluster shared memory is a lexical memory (per *Bailey*). The collective groups of recognizers then allow for speech recognition of the entire huge vocabulary because each group of recognizers and connected lexical memory contains, in combination, complete lexical data. That is, group 1 related to the context of “sports” stores all sports-related recognition models in a first lexical memory; group 2 related to the context of “health” stores all health-related recognition models in a second lexical memory; and the grouping architecture continues until all recognition models are stored in a context-related grouping.

The modification is obvious at least because *Thelen* already teaches grouping recognition models according to context. *Dec.*, 84. *Chen* expressly teaches its

“unified parallel processing architecture” creates “a high performance parallel processing computer system.” *Chen*, Abstract. It would have been a common sense extension of *Thelen*’s grouped, hierarchical recognition models to similarly group the recognizers (per *Chen*), where each recognizer is associated with a recognition model (per *Thelen*), and each group of recognizers (and associated processors) is connected to a lexical memory, per *Bailey*. *Dec.*, 84. Such a grouping of the recognizers (and therefore, the recognition processors) would have yielded the predictable result of consolidating speech recognition for a particular context within a particular group of recognition processors. *Dec.*, 84. The group of processors would then share the same lexical memory 615 (taught by *Bailey*, FIG. 3B), such that the same lexical memory stores all context-related recognition models. Grouping context-related recognizers together with a shared memory storing all context-related recognition models would have more efficiently allowed for sharing of lexical data common to each of the recognition models. *Dec.*, 84. Because all context-related recognition models are stored in the same lexical memory, the system is configured to transfer recognition models to/from the lexical memory for processing more efficiently. *Dec.*, 84.

Doing so would have yielded the predicable result of reducing the time required to recognize a user’s speech, desirably speeding the system’s response. *Dec.*, 85. *Thelen* expressly states a goal of recognizing speech in “‘real time’ in the

sense that no significant delay occurs” between a word being received in the system and recognizing the word. *Thelen*, 7:20-26. Further, a POSITA would have recognized both *Chen* and *Thelen-Bailey* contemplate a hardware architecture using multiple, parallel processors; therefore, the more efficient processing architecture of *Chen* would have been simply substituted for *Thelen-Bailey*’s recognizers to achieve a compatible improvement in processing speed and power, leading to a reasonable expectation of success for the combination. *Dec.*, 85. 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.*, 85 (*citing Hon*, 19-20).

There would have been a reasonable expectation of success in the modification to *Thelen*’s system. Modifying *Thelen* to store all context-related recognition models in one of the plurality of memories 615a-c taught by *Bailey* would have allowed for “high speed” efficient speech recognition by “splitting the processing work” (*Bailey*, 11:4-8, 10:62-65, 51:21-26) for various contexts amongst the speech recognizers (*Thelen*, 7:2-9). A POSITA would have recognized *Bailey*’s memory architecture would have combined naturally with the plurality of parallel “recognition engines” taught by *Thelen* and furthered *Thelen*’s expressly-stated goal of operating in “real-time.” *Thelen*, 2:30-37, 7:26-29; *Dec.*, 85. A POSITA would have recognized that the necessary modifications to *Thelen* would have only

required arrangement of well-known memory components and relocation of lexical data already present in *Thelen*, leading a POSITA to a reasonable expectation of success for the combination. *Dec.*, 85. Further, a POSITA would have recognized the combination of *Thelen-Bailey-Chen* would have constituted combining prior art elements according to known methods without undue experimentation to yield the predictable result of a plurality of speech recognizers operating in parallel using lexical data stored in a plurality of lexical memories. *Id.*

6. ***Claim 46(e): “a control processor controlling each processor to process said speech parameters using partial lexical data read from a respective said lexical memory”***

Both *Thelen* and *Bailey* teach a “control processor,” as claimed.

a) ***Thelen’s Teachings***

Thelen teaches a “control processor,” as claimed, namely model selector 360 in combination with controller 350.

Regarding model selector 360, *Thelen* teaches:

A model selector 360 is used to select for at least one of the speech recognizers the associated recognition model from the M models in dependence on a recognition context. The model selector 360 may select a model for each of the active recognizers.

Thelen, 7:63–8:1, FIG. 3. *Thelen* explains that the “model selector 360 **controls** at least one test recognizer” *Thelen*, 8:36-41, 9:8-10.

Per Mr. Schmandt, because model selector 360 controls a test recognizer (and thereby the recognizer's processor) and selects recognizers from among all the recognizers, the model selector controls processing of the speech parameters by each recognizer to use partial lexical data, i.e., the recognition model associated with the recognizer. *Dec.*, 86. The partial lexical data is read from a respective lexical memory, as discussed for Claim 46(b). Further, a POSITA would have readily understood the model selector 360 includes a "processor" because the model selector performs functionality associated with instructing the recognizers and associated processors. *Dec.*, 86.

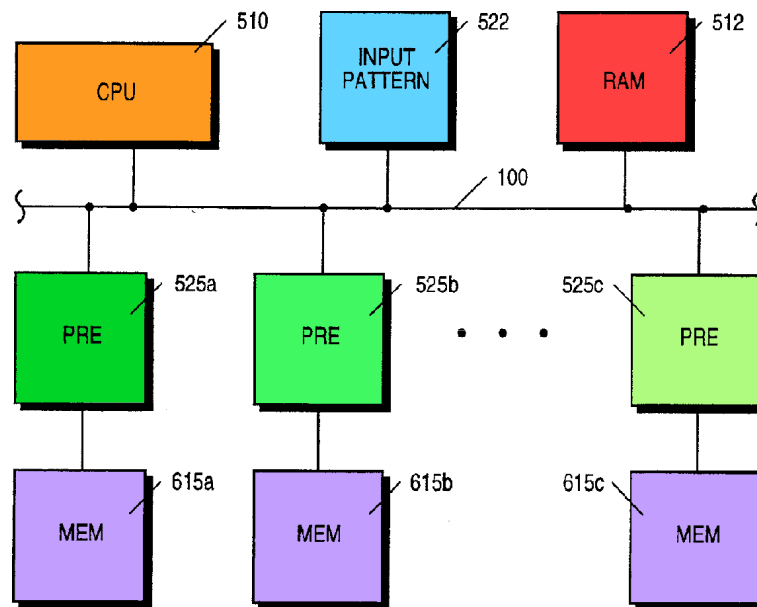
Thelen also teaches controller 350 "make[s] the final selection of a recognized word sequence" and is "responsible for directing the speech input to the recognizers." *Thelen*, 7:50-61, FIG. 3, 10:20-24 (Claim 10) ("the system comprises a controller operative to direct the input pattern to a plurality of the speech recognizers"). A POSITA would have understood controller 350 comprises a processor because the controller performs functionality associated with instructing the recognizers, e.g., directing the speech input to the recognizers. *Dec.*, 86.

Because *Thelen* teaches controller 350 directs the speech input or input patterns (i.e., the speech parameters) to the recognizers, and because the recognizers use partial lexical data (e.g., the recognition model associated with each recognizer) read from a respective lexical memory (e.g., lexical memory 615 for the group of

processors in the modified *Thelen* system), the controller 350 in combination with model selector 360 teaches or renders obvious Claim 46(e). *Dec.*, 86. The collective controller and model selector control each recognizer (and associated processor) direct speech to the recognizer and to select the recognizer (and its associated recognition model comprising partial lexical data) for processing the speech parameters. *Dec.*, 86.

b) *Bailey's Teachings*

Bailey also teaches a control processor, namely CPU 510. *Bailey* teaches CPU 510, illustrated below in Fig. 3A, “is responsible for splitting the processing work from among the separate PR engines,” by sending, during each recognition cycle, unknown patterns of speech to each PR engine 525 for comparison against the library of patterns stored in the lexical memories 615a-c.



Bailey, Fig. 3A (annotated), 11:4-8, 9:5-39, 11:23-34, 11:59-64 (teaching the “CPU is performing higher level scheduling, synchronization and result determination steps of the HMM and DTW pattern recognition procedures that are not processing intensive while the PR engine 525 performs the bulk of the lower level processing intensive aspects of these procedures”), 12:12-25 (teaching division of steps between CPU and PRE), 13:5-23 (discussing the CPU controlling updating of the PRE’s memory unit 615), FIGs. 3B, 2; *Dec.*, 86-87. Because the CPU 510 of *Bailey* is responsible for splitting the processing work among the PREs and for performing scheduling, synchronization, and result determination steps, a POSITA would have understood or found obvious the *Bailey* CPU 510 controls each PRE (and associated processor) to process the received speech parameters. *Dec.*, 87.

c) Motivation to Modify *Thelen-Bailey-Chen* to Include *Bailey*’s CPU for Claim 46(d)

A POSITA would have found it obvious and been motivated to modify the *Thelen-Bailey-Chen* system to include *Bailey*’s CPU 510 for controlling the plurality of *Thelen* recognizers 331-333 (and associated processors) to process the speech parameters, e.g., *Thelen*’s input unknown pattern. *Dec.*, 86-87. A POSITA would have recognized the modification would have constituted combining prior art elements according to known methods without undue experimentation to yield the predictable result of a control processor (*Thelen*’s model selector 360 and controller

350, embodied as *Bailey's* CPU 510) controlling each processor (the processors of *Thelen's* recognizers grouped in clusters per *Chen*) to process speech parameters using partial lexical data (*Thelen's* recognition models 341-348) read from a respective said lexical memory (*Bailey's* memories 615a-c modified pursuant to *Chen* to each be shared cluster memories 104a-d of the groups of clustered processors). *Dec.*, 86-88. A POSITA would have recognized that controlling a plurality of processors via a single control (or “master”) processor was well-known for decades prior to the '319 Patent, leading to a reasonable expectation of success for the combination. *Dec.*, 86 (citing *Wrench*, *Vensko*, and *Cecinatti*). A POSITA would have been further motivated to modify *Thelen* by *Bailey's* express teaching that its system, employing a CPU control processor, may be used to increase the accuracy and performance regarding pattern recognition applications, providing “real-time responses of at most 0.1 seconds between unknown pattern input and pattern identification, thus providing a system that is fully user interactive.” *Bailey*, 12:2-11; *Dec.*, 88. A POSITA would have been motivated to employ *Thelen's* control processor using *Bailey's* CPU 510 as an operable, efficient method of administering the speech processing functions of each of *Chen's* clusters of processors and offload the processing necessary for choosing a final determined recognized word. *Id.*

7. ***Claim 46(f): “a results memory storing the results of the processing of the speech parameters from said processors”***

a) ***Bailey Teaches Claim 46(f)***

As discussed, the modified speech recognition circuit of *Thelen-Bailey-Chen* provides a plurality of parallel recognizers to compare speech parameters to recognition models stored in lexical memories under the control of a control processor, operable to select a final recognized word based on the results of the comparisons. *Bailey* teaches a “results memory,” as claimed, within RAM 512.

Thelen itself teaches that “the output of the recognizers is directed to a controller 350 to make the final selection of a recognized word sequence.” *Thelen*, 7:50-58. Per Mr. Schmandt, *Thelen* strongly suggests results memory because controller 350 would not be able to select among multiple outputs unless they were stored in a memory for processing. *Dec.*, 89.

Regardless, *Bailey* teaches control processor (*Thelen*’s control processor 350 and model selector 360 implemented using *Bailey*’s CPU 510) may receive results of the comparisons via one of two schemes: (1) via a report generated at the end of each prototype reference comparison, or (2) at the end of the total library. *Bailey*, 11:42-59, 12:41-61, 11:23-34, Fig. 4; *Dec.*, 89. Using either scheme, the CPU 510 “receives an interrupt indicating a match complete” and **“generates a transfer of the results from the PR engine to the system memory 512.”** *Id.* Then, the CPU 510 post-processes all of the results generated by the processors of the PR engine

and formulates a final matched prototype word pattern based on the prototype candidates selected by the PR engine. *Id.* Thus, *Bailey* teaches a results memory (in RAM 512) storing the results of the processing of the speech parameters from each of the processors, as claimed. *Id.*

b) Motivation to Combine *Bailey* with *Thelen* for Claim 46(e)

A POSITA would have recognized the combination of *Thelen* and *Bailey* would have constituted combining prior art elements according to known methods without undue experimentation to yield the predictable result of a control processor (*Thelen*'s model selector 360 and controller 350, embodied as *Bailey*'s CPU 510) receiving results from each recognition processor via storage in a results memory (implemented in *Bailey*'s RAM 512). *Dec.*, 89. A POSITA would have recognized that storing results in memory was well-known prior to the '319 Patent, leading to a reasonable expectation of success for the combination. *Dec.*, 90 (*citing Robinson and Cecinati*). A POSITA would have been further motivated to modify *Thelen* by *Bailey*'s express teaching that its system, employing a results memory, may be used to increase the accuracy and performance regarding pattern recognition applications, providing "real-time responses of at most 0.1 seconds between unknown pattern input and pattern identification, thus providing a system that is fully user interactive." *Bailey*, 12:2-11; *Dec.*, 90. A POSITA would have been motivated to report outcomes of the comparisons by the processors of the circuit's pattern

recognition engine for choosing a final determined recognized word. *Id.* A POSITA would have recognized that the results of the comparisons would naturally have to be reported to the control processor for it to determine the recognized word, and that storing the results in memory would have been the most expected, straightforward manner of configuring the speech recognition circuit. *Id.*

Though *Bailey* describes each of the input buffer and results memory as being implemented using RAM 512, a POSITA would have recognized that the input buffer and results memory of *Bailey* would have been implemented in logically and/or physically distinct portions of RAM 512. *Dec.*, 91. A POSITA would have recognized that RAM 512 as described and illustrated by *Bailey* is intended to conceptually describe the RAM available to the speech recognition circuit, which may physically be provided in multiple portions and/or locations within the circuit. *Id.*

Further, to the extent it is argued that the input buffer of Claim 46(a) and the results memory of Claim 46(f) must be implemented in distinct memories, a POSITA would have found it obvious to provide separate RAM for each of the input buffer and results memory to cleanly distinguish between the incoming speech parameters and the results of the comparisons of the plurality of processors to the recognition models. *Dec.*, 91. This modification would have only required including an additional element of hardware already present in *Bailey* (RAM 512) and

potential reprogramming of the circuit to store results in a different physical piece of hardware, skills available to a POSITA without undue experimentation. *Id.*

B. Claim 50

1. ***Claim 50: “...wherein each processor compares the processed speech parameters with the lexical data in a corresponding lexical memory to identify words as a word recognition event and sends information identifying the identified words to said results memory as the processing results.”***

The speech recognition circuit of *Thelen-Bailey-Chen* teaches this limitation. In the modified *Thelen* circuit mapped for Claim 46, **each** processor of *Thelen*’s recognizers 331-331 compares the received unknown input pattern (the processed speech parameters) with the recognition model (the lexical data) stored in corresponding memory (memory 615 of *Bailey* organized in groups of processors with a lexical memory, per mapped Claim 46(d)). *See Claim 46*. Words, including word sequences, are identified. *Thelen*, Abstract, 7:50-52.

Bailey further teaches when a pattern comparison being performed by the processors is complete, akin to *Thelen*’s word recognition using the recognition models, the *Bailey* PR engine 525 generates an interrupt to CPU 510, which in turn “generates a transfer of the results from the PR engine to the system memory 512” (the mapped “results memory”). *Bailey*, 12:41-61, pinpoint at 12:46-48 (“Results are signaled to the CPU by the PR engine via an interrupt generated by the PR engine ...”), 11:42-56 (discussing PRE performing the pattern comparison and generating

an interrupt indicating PRE processing complete), 12:53-55 (teaching generating an interrupt to transfer results to memory 512), 15:41-45 (similar teaching).

A POSITA would have recognized the “interrupts” being generated upon comparison completion teach or at least render obvious the claimed “word recognition event.” *Bailey*, 11:42-59, 12:41-61; *Dec.*, 92-93. In the modified *Thelen* system, as each recognizer completes its comparison using the recognition model to recognize a word, an interrupt event is signaled to the *Thelen* control processor, operating pursuant to *Bailey*’s CPU 510, to initiate the results transfer to the results memory for final word selection by the controller. *Thelen*, 7:50-58 (“The output of the recognizers is directed to a controller 350 to make the final selection of a recognized word sequence.”); *Dec.*, 93.

A POSITA would have been motivated and found it obvious to modify *Thelen*’s speech recognition circuit to generate word recognition events, such as via an interrupt, to assist in processing by informing the control processor that the speech recognition is complete. *Bailey* teaches interrupts are “typically used to indicate completion” of a word recognition event, thus indicating that utilizing interrupts would have been a well-known modification to *Thelen*’s circuit with a reasonable expectation of success. *Bailey*, 15:43-45; *Dec.*, 93.

C. Claim 54

- 1. Claim 54: “...wherein said processors process said speech parameters by comparing said speech parameters with said*

lexical data or data derived from lexical data in respective said lexical memories.”

The circuit of *Thelen-Bailey-Chen* teaches Claim 54. *Thelen*’s recognizers 331-333 include recognition processors, per Claim 46(c). *Thelen* further teaches “matching” observation vectors from the input speech (i.e., the mapped “speech parameters”) against a speech recognition unit, such as a phone, modeled by an HMM. *Thelen*, 5:2-45. *Thelen* teaches the speech recognition units are part of the word model, which includes the lexicon 134 and sub-word models 132. *Thelen*, 5:14-22. *Thelen* also teaches a “word level matching system...matches the observation vectors against all sequences of speech recognition units and provides the likelihoods of a match between the vector and a sequence.” *Thelen*, 5:46-49. Therefore, *Thelen* teaches the recognizers (including the processors of the recognizers) comparing (i.e., *matching*) speech parameters (i.e., the observations vectors from the input speech) with lexical data (i.e., the lexicon that includes all sequences of speech recognition units, per *Thelen*, 5:49-52)). As mapped for Claim 46(a), the lexicons are included in the recognition models stored in the lexical memories of the modified *Thelen*.

D. Claim 64

1. ***Claim 64: “...wherein said input buffer is arranged to receive said speech parameters as feature vectors.”***

The circuit of *Thelen-Bailey-Chen* teaches Claim 64. In the combined system, *Thelen*’s speech parameters, which are received by an input buffer in RAM 512 taught by *Bailey*, are feature vectors. See Section V.A.2, citing *Thelen*, 4:12-37 and *Bailey*, 9:1-39, Figs. 2, 3A. A detailed motivation for combining the input buffer of *Bailey* to receive the feature vectors of *Thelen* is provided above for Claim 46(a) and incorporated herein by reference. See Section V.A.2.d.

E. Claim 67

1. ***Claim 67: “...wherein said lexical memories store said lexical data as Hidden Markov Models, and each processor performs the Viterbi search algorithm using a respective part of said lexical data”***

The circuit of *Thelen-Bailey-Chen* teaches Claim 67. *Thelen* teaches each recognition model is characterized by a Hidden Markov Model (HMM) generated from training data. *Thelen*, 5:29-42. Further, *Bailey* explains the Viterbi score procedure is used on HMMs to “compute the highest probability that a given set of observation points matches a reference pattern.” *Bailey*, 7:51–8:1, see also 8:2-36 detailing a Viterbi search algorithm being applied as illustrated in Fig. 1B; *Dec.*, 94. Further, a POSITA would have recognized that “the Viterbi algorithm is a commonly used method” of classifying unknown patterns of input acoustic data

using HMMs, expecting that the Viterbi search explicitly described by *Bailey* would have been similarly used in *Thelen's* speech recognition process, though not expressly stated in *Thelen. Dec.*, 94 (*citing Schmandt*, 59-60.)

A POSITA would have configured the combined speech recognition circuit to perform a Viterbi search algorithm using each of the lexical tree processors to maximize the benefit gained and maintain consistency between the lexical tree processors. *Dec.*, 95. Additionally, a POSITA would have recognized that each of the processors in the modified system would have only utilized the lexical data stored in its respective lexical memory, and thus the circuit teaches processors performing a Viterbi search algorithm “using a respective part of said lexical data.” *Id.*; *Thelen*, 5:29-31 (teaching each word model, encompassed in each recognition model serving as “part of said lexical data” per Claim 46(b)’s mapping, is based on an HMM).

A motivation for combining the speech recognition circuits of *Thelen* and *Bailey* is provided above for Claim 46(d) and is incorporated herein by reference. *See* Section V.A.5.b.

VI. GROUND 2: CLAIMS 47-49 ARE OBVIOUS OVER *THELEN* IN VIEW OF *BAILEY* AND *CHEN* AND IN FURTHER VIEW OF *HUANG*

A. Claim 47

1. *Claim 47: “...wherein each lexical memory stores lexical data as at least one lexical tree data structure, each lexical tree data*

structure comprises a model of words having common prefix components, and an initial component of each lexical tree data structure is unique”

As discussed above, the circuit of *Thelen-Bailey-Chen* comprises a plurality of lexical memories, each containing recognition models corresponding to particular contexts. The *Thelen-Bailey-Chen* circuit does not, however, expressly teach structuring its lexical data in a lexical tree data structure.

In related art, *Huang* teaches a “conventional technique for further reducing the magnitude of computation required for speech recognition” using a prefix tree that “represents the lexicon of the speech recognition system as a tree structure wherein all of the words likely to be encountered by the system are represented in the tree structure.” *Huang*, 2:31-37. Thus, *Huang*’s “lexicon of all words” organized in a “prefix tree” teaches lexical data stored as at least one “lexical tree data structure” comprising a “model of words.” *Huang*, 2:31-58; *Dec.*, 96.

Huang describes its lexical tree as follows:

In such a prefix tree, **each subword unit (such as a phoneme) is typically represented by a branch** which is associated with a particular acoustic model (such as an HMM). The phoneme branches are connected, at nodes, to subsequent phoneme branches. **All words in the lexicon which share the same first phoneme share the same first branch.** All words which have the same first and second phonemes share the same first and second branches. By contrast, words which have a common first phoneme, but which have different second

phonemes, share the same first branch in the prefix tree but have second branches which diverge at the first node in the prefix tree, and so on. The tree structure continues in such a fashion such that all words likely to be encountered by the system are represented by the end nodes of the tree (i.e., the leaves on the tree).

Huang, 2:38-52. Based on this description, a POSITA would have recognized *Huang* teaches a lexical tree data structure comprises a plurality of branches, each branch having an initial prefix component which is *common* amongst the words in that branch (corresponding to a particular phoneme) and *distinct* from each of the initial prefix components of the other branches. *Dec.* 96. This is the well-known, typical, and expected structure of lexical trees. *Id.* Though Claim 47 refers to each of these distinct, branching data structures as a plurality of “lexical data tree structures,” a POSITA would have recognized *Huang*’s “branches” teach the same data structure. *Id.* A POSITA would have understood each branch of a lexical tree, when separated, itself represents a lexical tree, branching from a common initial node to terminating “leaves” representing words. *Id.* Thus, *Huang* teaches or at least renders obvious the lexical tree data structures recited by Claim 47. *Huang*, 2:31-52, 7:21-27, 10:16-33, 12:43-46, *see also* 5:62-6:17, 8:8-9, and Figs. 1-2 (teaching prefix tree storage); *Dec.*, 96.

A POSITA would have been motivated by *Huang*’s express statement of “reducing the magnitude of computation required for speech recognition” to modify

Thelen's HMM lexicon 134 into a lexical tree data structure, requiring only reprogramming of data stored in memory, performable by a POSITA without undue experimentation. *Dec.*, 97. Storing lexical data in a lexical tree data structure representing lexical content mapped to sub-word units (such as phonemes) was well-known to improve computational efficiency in an HMM speech recognition circuit. *Dec.*, 96-97 (*citing Ravishankar*, 68, stating “the single largest source of computational efficiency” in searching HMMs is “organizing the HMMs to be searched as a phonetic tree, instead of the flat structure”). Both *Thelen* and *Bailey* teach multiple processing units (*Thelen's* “recognizers” and *Bailey's* “PREs”) running in parallel to achieve better recognition results without increasing overall processing time between utterance and recognition. *Thelen*, 2;30-37, 7:17-29; *Bailey*, 11:9-34, FIG. 3A. Because the lexical tree structure comprises a plurality of branches, a POSITA would have appreciated that a lexical tree structure would have been an effective data representation of lexical content for processing distributed across multiple processors (as recited in Claim 46), as different branches may simply be assigned to the different processors. *Dec.*, 97.

As such, a POSITA would have had a reasonable expectation of success and been further motivated to organize *Thelen's* HMM lexicon 134 as a lexical tree data structure for increased computational efficiency. Reorganizing the lexicon 134 as the recited lexical tree data structure would have constituted using a known

technique (of lexical data storage organized as a “trees”) to improve similar devices (the speech recognition circuits of *Thelen-Bailey-Chen* and *Huang*) in the same way.

Dec., 97.

B. Claim 48

1. ***Claim 48: “...wherein each lexical memory stores said lexical tree data structure as an n phone model of words, where n is an integer, and said components comprise n phones.”***

As discussed regarding Claim 47, the *Thelen-Bailey-Chen-Huang* circuit comprises a plurality of lexical memories storing lexical tree data structures. Further, *Huang* teaches building a monophone model of words during a training phase, which are represented in the lexical tree data structures stored in lexical memory and subsequently used for speech recognition. *Huang*, 7:15-54, 8:16-20, 11:64-12:46, *see also* 6:62–7:18, FIGs. 1-2 (teaching monophone storage); *Dec.*, 98. A POSITA would have recognized monophones are n phone models of words where $n=1$, an integer. *Dec.*, 98.

Specifically, *Huang* describes modeling each phoneme in the data set of recognizable words (based on incoming training data) as monophone models which “are used in the recognition scheme to determine a most likely phoneme for an input speech utterance.” *Huang*, 7:45-54. Later, during speech recognition, the monophone models are used to determine target phonemes of spoken words to be searched in the prefix tree (after further steps regarding dynamically-generated

context dependent phone models, discussed below regarding Claim 49). *Huang*, 11:64–12:46; *Dec.*, 98.

Further, the “components” of *Huang*’s lexical tree data structure (i.e., the common prefix components of Claim 46, mapped to the initial phonemes of each branch of *Huang*’s lexical tree) are described as not exceeding the total number (40-50) of phonemes used in the English language. *Huang*, 2:53-61, 8:24-25; *Dec.*, 99. A POSITA would have recognized that these 40-50 phonemes described by *Huang* as the initial phonemes (the mapped “components”) are specifically monophones because there are approximately 50 monophones in the English language, but an exponentially larger number of n phones as n becomes larger (e.g., $50^3=125,000$ triphones). *Huang*, 2:53-61, 7:45-54 (teaching modeling “each phoneme found in the training data set as a monophone model”), 11:64-66; *Dec.*, 99, *citing* ’319 Patent, 11:18-20 (“If there are 50 mono phones, there are potentially 50^n lexical tree data structures. For tri-phones, this comprises 50^3 (i.e. 125,000).”). Thus, *Huang* teaches or at least renders obvious “said components comprise n phones” where $n=1$ for monophones. *Dec.*, 99.

A POSITA would have been motivated to store the lexical tree data structure as a monophone (an “ n phone”) model of words to minimize the amount of memory required for storing the trees. *Dec.*, 100. The modification would have required only reprogramming of data stored in memory, performable by a POSITA without undue

experimentation. *Dec.*, 100. Mr. Schmandt explains lexical tree data structures comprising monophones were well-known prior to the '319 Patent. *Dec.*, 99-100 (*citing Ravishankar*). As such, a POSITA would have had a reasonable expectation of success for using monophones in the lexical tree. *Dec.*, 100. Storing lexical tree data structures comprising monophones would have constituted using a known technique (storing lexical tree data as monophones) to improve similar devices (the speech recognition circuits of *Thelen-Bailey-Chen* and *Huang*) in the same way. *Id.* Specifically, for the motivations discussed, a POSITA would have stored lexical tree data structure as an n phone model of words, where n is an integer, and said components comprise n phones for each of the lexical memories, to maximize the benefit gained and maintain consistency between the lexical tree data structures. *Id.*

C. Claim 49

1. ***Claim 49: "...wherein each lexical memory stores said lexical tree data structure as a mono phone model of words, and said lexical tree processors use said mono phone models to generate context dependant [sic] phone models of words dynamically for use in processing the speech parameters."***

The '319 Patent describes a process of dynamically generating context dependent phone models by using mono phone models to determine a plurality of triphones associated with its lexical tree data structure. '319 Patent, 6:40–7:35, 9:7-16; *Dec.*, 101. While the '319 Patent's description specifically discusses using mono phone models to determine *next* lexical trees to process, Claim 49 merely

requires the lexical tree processors generally use mono phone models to generate context dependent phone models (including triphone models), which are then used for processing speech parameters.

As discussed regarding Claim 48, the circuit of *Thelen-Bailey-Chen-Huang* comprises a plurality of lexical memories storing tree data structures as a monophone model of words. Further, *Huang* teaches generating context dependent triphone models of words dynamically for use in processing speech parameters, teaching Claim 49. *Huang*, 7:45–8:19, 12:32-46, Fig. 9; *Dec.*, 101.

Huang teaches that its circuit compares incoming speech to its monophone model to obtain likely phonemes. *Huang*, 11:64–12:8, 7:45-54. After further intermediate processing regarding senone states, *Huang* then details how the circuit dynamically generates context dependent phone models:

Search engine 80 then accesses the information stored in memory 74 and retrieves a likely triphone which has been mapped to the likely senone sequence determined. This is indicated by block 172. [¶] Search engine 80 then determines whether all of the likely phonemes have been processed. If not, search engine 80 repeats the process and arrives at a likely senone sequence for each likely phoneme (and thus arrives at N likely triphones to be associated with the target phoneme) based upon the probability scores determined during comparison. This is indicated by blocks 174 and 176. [¶] Once the N likely triphones have been identified, search engine 80 accesses the prefix tree in memory 76.

Huang, 12:32-46, 7:45–8:20; *Dec.*, 102. Notably, *Huang* expressly characterizes a triphone as “a phoneme with corresponding left and right **context**.” *Huang*, 7:64-66. Thus, a POSITA would have recognized that using monophone models to generate triphone models, as taught by *Huang*, teaches using a mono phone model to generate context dependent phone models for words, as claimed. *Dec.*, 102.

A POSITA would have been motivated to configure the lexical tree processors of the *Thelen-Bailey-Chen-Huang* speech recognition circuit to use the monophone model of words (as discussed above regarding Claim 48) in each lexical memory to dynamically generate context dependent triphone models for use in speech-parameter processing to minimize the computational load with the benefit of increased accuracy provided by accounting for the preceding phone as taught by *Huang*. *Dec.*, 103. The modification would have required only reprogramming of the functionality of the speech recognition circuit, performable by a POSITA without undue experimentation and requiring no additional hardware. *Dec.*, 103. Mr. Schmandt explains context dependent models based on triphones were well-known prior to the '319 Patent to increase the accuracy of speech recognition circuits, leading a POSITA to a reasonable expectation of success for the modification. *Dec.*, 103-104 (*citing Ravishankar*). Generating context dependent models pursuant to *Huang* would have constituted using a known technique (using monophone models to determine triphone models) to improve similar devices (the

speech recognition circuits of *Thelen-Bailey-Chen* and *Huang*) in the same way. *Dec.*, 103-104. Specifically, for the motivations discussed, a POSITA would have modified the speech recognition circuit to generate context dependent triphone models using each of the lexical tree processors to maximize the benefit gained and maintain consistency between the lexical tree processors. *Id.*

**VII. VII. GROUND 3: CLAIMS 53 AND 58-59 ARE OBVIOUS OVER
THELENIN VIEW OF BAILEY AND CHENIN FURTHER VIEW OF
JIANG**

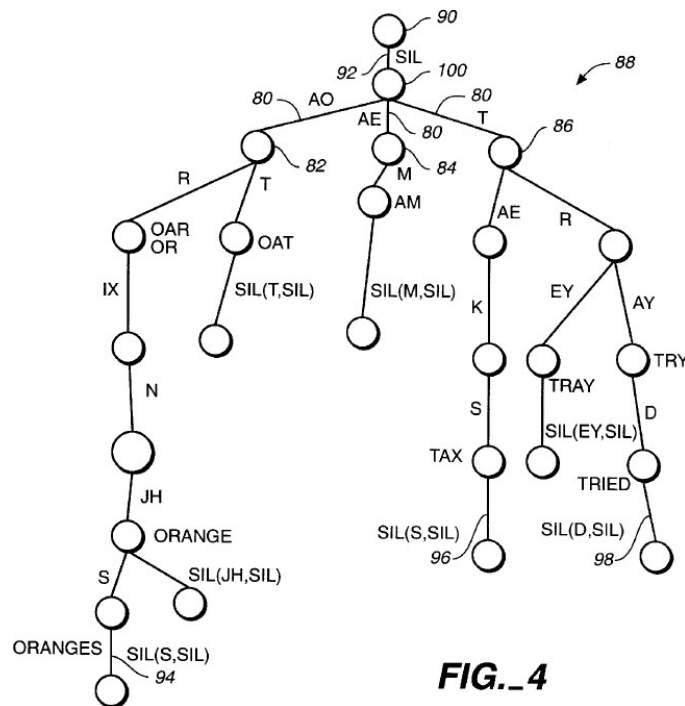
A. Claim 53

1. ***Claim 53: "...wherein each processor determines and outputs scores for words in the processing results at word recognition events."***

The circuit of *Thelen-Bailey-Chen* comprises a plurality of processors that each determines and outputs scores for words in processing results. *See* Sections V.A.3-V.A.7. Each of the plurality of processors recognizes speech according to recognition models. *Id.* *Thelen* states that word scores from each recognizer are provided to a controller 350 for "final selection" of a recognized word or sequence. *Thelen*, 7:50-58. *Bailey* explains speech recognition procedures include "three major processing steps: 1) a preprocessing or feature extraction and normalization step; 2) a search and match step; and 3) a post processing and final decision step." *Bailey*, 12:12-15. In this final stage, the CPU 510 acting as controller 350 (part of the mapped "control processor," *see* Section V.A.6) post-processes the results generated

by the PR engine and formulates a final match prototype pattern based on the candidates selected. *Bailey*, 12:53-61. However, the combined *Thelen-Bailey-Chen* circuit does not expressly characterize this final determination process as a “word recognition event.”

Jiang teaches a similar speech recognition circuit comprising a CPU 21, “which may include one or more processors[.]” *Jiang*, 4:8-9, 6:39-42. *Jiang* teaches a lexical tree (or “prefix tree”) associating phonetic models, such as those modeled by HMMs, with branches of the prefix tree. *Jiang*, 2:47-59, 7:34-38; *Dec.*, 105. In *Jiang*, the prefix tree comprises a plurality of nodes connected by phoneme branches, with end nodes of each branch (i.e., leaves on the tree) representing final phonemes of words to be recognized. *Jiang*, 2:53-67, 4:8-12, 7:48–8:4. Scores are assigned to each node of the prefix tree as the tree is traversed. *Jiang*, 8:16-51. *Jiang*’s speech recognition circuit computes “probability scores” for each state of an acoustic model stored in memory, such as the Hidden Markov Model, 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. When the prefix tree search terminates at an end (“leaf”) node, a final word is output as the recognized spoken utterance (such as “orange,” in the example described at 8:16-64); *Dec.*, 105. Thus, like *Thelen-Bailey-Chen*, *Jiang* teaches a speech recognition circuit that determines and outputs scores for words in the processing results. *Id.*

Further, *Jiang* teaches scores are output at word recognition events, referring to these events as “word boundaries.” *Jiang* 7:48-64, 8:65–9:3, 9:28-42; *Dec.*, 105. When a boundary between words is detected, as indicated by silence, the lexical tree is searched to determine the word mostly likely represented by the utterance. *Id.* After the tree has been traversed, a leaf on the tree (representing a word) is reached, after which a silence branch is enforced, indicating the boundary of the recognized

word (a “word recognition event”). *Id.* At this point, the score for the recognized words for the combined *Thelen-Bailey-Chen-Jiang* circuit would be output to the control processor. *Id.*

A POSITA would have been motivated to configure the circuit of *Thelen-Bailey-Chen* to determine and output scores for words at word recognition events as in *Jiang* as an operable, efficient method of determining a final recognized word, as already taught by *Thelen* and *Bailey*. A POSITA would have recognized outputting scores at word recognition events would have provided the control processor with data necessary to evaluate word hypotheses at an appropriate step in the process (i.e., at the end of words), leading to a reasonable expectation of success for the modification. *Dec.*, 106. Specifically, a POSITA would have configured the combined speech recognition circuit to determine and output scores for words in the processing results at word recognition events using each of the clusters of lexical tree processors to maintain consistency between the pattern recognition engines (PRs 525 of *Bailey*), allowing the scores to be evaluated uniformly. *Dec.*, 106.

B. Claim 58

1. Claim 58: “...wherein said processors determine and output scores in the processing results during the processing of said speech parameters”

As discussed, *Jiang* teaches a speech recognition circuit comprising processors that determine scores at word recognition events. *See* Section VII.A.1.

Further, *Jiang* teaches determining and outputting scores for a plurality of nodes in a lexical tree during processing of speech parameters to determine which branches should be “dropped (or pruned)...and no longer considered in further processing.”

Jiang, 8:52-64. Specifically, *Jiang* teaches performing a “pruning technique” 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, 2:25-46, 10:51–11:15; *Dec.*, 107. In *Jiang*, scores are determined “during the processing of speech parameters,” to obtain “computational savings” afforded by no longer processing pruned branches, making the system “more efficient” by “reducing the search space more quickly.” *Jiang*, 10:51-64, 11:8-24; *Dec.*, 107. A POSITA would have been motivated to configure the *Thelen-Bailey-Chen* circuit to determine and output scores during processing to enable the computational savings and improvement in efficiency afforded by *Jiang*’s pruning. *Dec.*, 107. Doing so would have constituted using a known technique (determining

pruning scores) to improve similar devices (the circuits of *Thelen-Bailey-Chen* and *Jiang*) in the same way, leading to a reasonable expectation of success for the combination. *Dec.*, 107.

C. Claim 59

1. ***Claim 59: “...wherein said processors prune the processing of the speech parameters in dependence upon the determined scores”***

The circuit of *Thelen-Bailey-Chen-Jiang* comprises processors that prune the processing of speech parameters in dependence upon on the scores determined processing. *See* Section VII.B.1 (Claim 58), *citing Jiang*, 8:52-64.

VIII. GROUND 4: CLAIM 51 IS OBVIOUS OVER *THELEN* IN VIEW OF *BAILEY* AND *CHEN* IN FURTHER VIEW OF *KATZ*

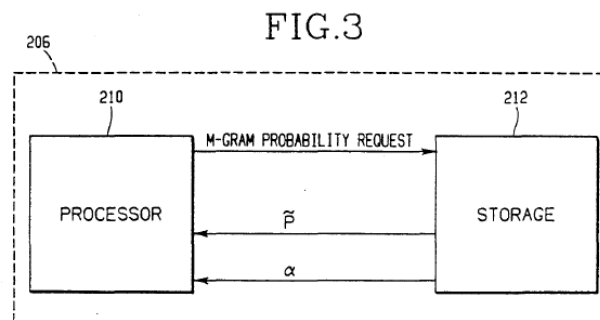
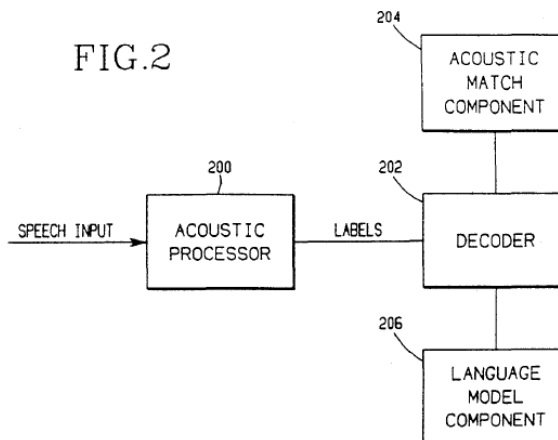
A. Claim 51

1. ***Claim 51: “...including language model processor providing a language model output for modifying the processing results at a word recognition event by a said processor”***

Thelen teaches improving results of speech recognition at the sentence level by ensuring “word sequences are proper sequences according to the language model.” *Thelen*, 4:20-25, 5:54–6:10. *Thelen* teaches its circuit “combin[es] the results of the acoustic model” (i.e., the processing results of word recognition) with the language model to produce a “recognized sequence,” and thus teaches “providing a language model output for modifying the processing results at a word recognition event,” as claimed. *Thelen*, 5:54–6:10; *Dec.*, 108 (explaining a POSITA would have

recognized the language model would have been applied after a word is recognized, thus “at a word recognition event”). *Thelen’s* sentence-level language model improvement to its speech recognition is deployed *in addition to* the word-level recognition models discussed regarding Claim 46, rather than as in replacement or alternative to word-level recognition models. *Thelen*, 7:30-43; *Dec.*, 108-109. *Thelen* does not expressly state the language model output modifying the processing results is performed by a “language model processor.”

In related art, *Katz* teaches a speech recognition circuit comprising a dedicated language model processor. *Katz*, 6:33-41, 2:34-44, FIGs. 2-3, 5:61–6:32, 6:42–7:46. Illustrated below in FIGs. 2-3, a “language model component 206” comprising a processor 210 (the mapped “language model processor”) and storage 212 operable to determine the probability that a selected vocabulary word is the next word in a string, given that the previous words in the string are known based on a sample text of phrases (such as three word combinations known as “tri-grams”).



Katz, 6:33-41, 1:44-54, 2:34-44, FIGs. 2-3; *Dec.*, 108-109.

A POSITA would have found it obvious and been motivated to utilize *Katz*'s language model processor as an operable structure for providing a language model output for modifying the processing results at a word recognition event already taught by *Thelen*. *Dec.*, 109. A POSITA would have recognized each of *Thelen* and *Katz* describes using tri-grams, leading to a reasonable expectation of success for the combination without undue experimentation. *Thelen*, 5:54-6:12; *Katz*, 1:44-54; *Dec.*, 109. A POSITA would have recognized offloading the language model processing to a language model processor would have conserved computational resources, further motivating the combination. *Dec.*, 109.

IX. 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 ’319 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 (EX1044), will begin approximately at the same time a Final Written Decision is due, approximately May 1, 2024. (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. (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. (IPR2023-00033, Paper 10; Factor 6).

§325(d). Petitioners seek to join already-instituted IPR2023-00033, 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.

X. CONCLUSION

For the forgoing reasons, Petitioners respectfully request *inter partes* review of the Challenged Claims of the '319 Patent.

Respectfully submitted,

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

XI. 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 '319 Patent is currently the subject of *Apple Inc. v. Zentian Ltd.*, Inter Partes Review No. IPR2023-00033 (P.T.A.B.) (instituted Jun. 12, 2023). This Petition is substantively identical to the petition submitted in IPR2023-00033 and Petitioners have concurrently filed a motion to join that proceeding.

The '319 Patent is also asserted in the following pending patent infringement lawsuits filed by Zentian Ltd:

- *Zentian Ltd v. Amazon.com, Inc.*, No. 6:22-cv-00123 (W.D. Tex. Feb. 2, 2022);
- *Zentian Ltd v. Apple Inc.*, No. 6:22-cv-00122 (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	Saina Shamilov (Reg. No. 48,266) sshamilov@fenwick.com Fenwick & West LLP 801 California Street

Petition for *Inter Partes* Review of
U.S. Patent No. 7,587,319

Mountain View, CA 94041 Telephone: (650) 988-8500 Facsimile: (650) 938-5200	Mountain View, CA 94041 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. 7,587,319 FOR CLAIMS 46-51, 53-54, 58-59, 64, 67-68

Claim Designation	Claim Language
Claim 46(Pre)	A speech recognition circuit comprising:
Claim 46(a)	an input buffer receiving processed speech parameters;
Claim 46(b)	a plurality of lexical memories containing in combination complete lexical data for word recognition, each lexical memory containing part of said complete lexical data;
Claim 46(c)	a plurality of processors connected in parallel to said input buffer for processing the speech parameters in parallel,
Claim 46(d)	said processors being arranged in groups of processors, each group of processors being connected to a lexical memory;
Claim 46(e)	a control processor controlling each processor to process said speech parameters using partial lexical data read from a respective said lexical memory; and
Claim 46(f)	a results memory storing the results of the processing of the speech parameters from said processors.
Claim 47	A speech recognition circuit according to claim 46, wherein each lexical memory stores lexical data as at least one lexical tree data structure, each lexical tree data structure comprises a model of words having common prefix components, and an initial component of each lexical tree data structure is unique.
Claim 48	A speech recognition circuit according to claim 47, wherein each lexical memory stores said lexical tree data structure as an n phone model of words, where n is an integer, and said components comprise n phones.

Claim Designation	Claim Language
Claim 49	A speech recognition circuit according to claim 48, wherein each lexical memory stores said lexical tree data structure as a mono phone model of words, and said lexical tree processors use said mono phone models to generate context dependant phone models of words dynamically for use in processing the speech parameters.
Claim 50	A speech recognition circuit according to claim 46, wherein each processor compares the processed speech parameters with the lexical data in a corresponding lexical memory to identify words as a word recognition event and sends information identifying the identified words to said results memory as the processing results.
Claim 51	A speech recognition circuit according to claim 50, a including language model processor providing a language model output for modifying the processing results at a word recognition event by a said processor.
Claim 53	A speech recognition circuit according to claim 51, wherein each processor determines and outputs scores for words in the processing results at word recognition events.
Claim 54	A speech recognition circuit according to claim 46, wherein said processors process said speech parameters by comparing said speech parameters with said lexical data or data derived from lexical data in respective said lexical memories.
Claim 58	A speech recognition circuit according to claim 46, wherein said processors determine and output scores in the processing results during the processing of said speech parameters.
Claim 59	A speech recognition circuit according to claim 58, wherein said processors prune the processing of the speech parameters in dependence upon the determined scores.

Claim Designation	Claim Language
Claim 64	A speech recognition circuit according to claim 46, wherein said input buffer is arranged to receive said speech parameters as feature vectors.
Claim 67	A speech recognition circuit according to claim 46, wherein said lexical memories store said lexical data as Hidden Markov Models, and each processor performs the Viterbi search algorithm using a respective part of said lexical data.

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 13,998 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 Petitioner
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:

Blank Rome LLP
1825 Eye Street, NW
Washington DC 20006-5403

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 Petitioner
Amazon Web Services, Inc. and
Amazon.com Services LLC