
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. 7,979,277

Issued: July 12, 2011

Inventor: Guy Larri, et al.

Title: SPEECH RECOGNITION CIRCUIT AND METHOD

Inter Partes Review No. IPR2023-01193

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 7,979,277**

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 '277 PATENT	1
A. Patent Description	1
B. Summary of Unpatentability	2
C. Priority Date	4
D. Level of Skill of a POSITA.....	4
III. REQUIREMENTS FOR IPR UNDER 37 C.F.R. §42.104.....	4
A. Grounds for Standing (37 C.F.R. §42.104(a))	4
B. Identification of Challenge (37 C.F.R. §42.104(b)).....	4
C. Claim Construction (37 C.F.R. §42.104(b)(3)).....	5
1. Construction Pursuant to 35 U.S.C. §112, ¶ 6	6
IV. SHOWING OF ANALOGOUS PRIOR ART	7
A. Jiang.....	7
B. Baumgartner	8
C. Brown	9
D. Kazeroonian.....	9
E. Vensko	10
F. Smyth.....	10
V. GROUND 1: CLAIMS 1, 5, 7, 12, AND 14-16 ARE OBVIOUS OVER JIANG, BAUMGARTNER, AND BROWN	11
A. Claim 1	11
1. Claim 1(Pre).....	11

TABLE OF CONTENTS (Continued)

	Page
2. Claim 1(a)	12
3. Claim 1(b)	14
a) The '277 Patent's Discussion of "extracted" or "derived quantities"	14
b) <i>Jiang's</i> Teachings	15
4. Claim 1(c)	16
a) '277 Patent's Discussion of "distances"	17
b) <i>Jiang's</i> Teachings	19
(1) "a calculating circuit"	19
(2) "a plurality of predetermined acoustic states of an acoustic model"	20
(3) "a feature vector"	21
(4) "calculating distances ..."	22
c) <i>Baumgartner</i> Teaches a "calculating circuit"	27
d) Motivation to Combine <i>Jiang</i> with <i>Baumgartner</i>	29
5. Claim 1(d)	29
a) "a lexical tree"	29
b) "the lexical tree comprising a model of words"	30
c) "a search stage for using said calculated distances to identify words within a lexical tree"	31
6. Claim 1(e)	34
a) Overview of Mapping	35

TABLE OF CONTENTS (Continued)

	Page
b) <i>Baumgartner</i> Teaches an “audio front end,” a “calculating circuit,” and a “search stage”	37
c) <i>Baumgartner</i> Teaches an Audio Front End “implemented using a first processor”	39
d) <i>Baumgartner</i> Teaches a Calculating Circuit “implemented using a second processor”	40
e) <i>Baumgartner</i> Teaches a Search Stage “implemented using a first processor”	41
f) Motivation to Combine <i>Jiang</i> and <i>Baumgartner</i>	42
7. Claim 1(f)	45
a) <i>Brown’s</i> Teachings	45
b) Motivation to Combine <i>Brown</i> with <i>Jiang-</i> <i>Baumgartner</i>	46
B. Claim 5	48
C. Claim 7	49
D. Claim 12	49
E. Claim 14	50
1. Claim 14(Pre)	50
2. Claim 14(a)	50
3. Claim 14(b)	50
4. Claim 14(c)	50
5. Claim 14(d)	51
6. Claim 14(e)	51

TABLE OF CONTENTS
(Continued)

	Page
F. Claim 15	51
1. Claim 15(Pre).....	51
2. Claim 15(a)	51
3. Claim 15(b)	51
4. Claim 15(c)	51
5. Claim 15(d)	51
6. Claim 15(e)	51
G. Claim 16	52
1. Claim 16(Pre).....	52
2. Claim 16(a)	52
3. Claim 16(b)	53
4. Claim 16(c)	53
5. Claim 16(d)	53
6. Claim 16(e)	53
VI. GROUND 2: CLAIM 4 IS OBVIOUS OVER <i>JIANG</i> , <i>BAUMGARTNER, BROWN, AND KAZEROONIAN</i>	53
VII. GROUND 3: CLAIMS 9-10 ARE OBVIOUS OVER <i>JIANG</i> , <i>BAUMGARTNER, BROWN, AND VENSKO</i>	55
A. Claim 9	55
B. Claim 10	59
VIII. GROUND 4: CLAIMS 1, 5, 7, 12, AND 14-16 ARE OBVIOUS OVER <i>JIANG, BAUMGARTNER, BROWN, AND SMYTH</i>	59

TABLE OF CONTENTS (Continued)

	Page
A. Claim 1	60
1. Claim 1(Pre).....	60
2. Claim 1(a)	60
3. Claim 1(b)	60
4. Claim 1(c)	60
a) “a calculating circuit”	60
b) “a plurality of predetermined acoustic states”	61
c) “[plurality of predetermined acoustic states] of an acoustic model”	63
d) “for calculating distances ...”	66
e) Motivation to Modify <i>Jiang-Baumgartner-Brown</i> with Smyth.....	68
5. Claim 1(d)	70
6. Claim 1(e)	70
7. Claim 1(f).....	70
B. Claim 5	70
C. Claim 7	70
D. Claim 12	70
E. Claim 14	71
1. Claim 14(Pre).....	71
2. Claim 14(a)	71

TABLE OF CONTENTS
(Continued)

	Page
3. Claim 14(b)	71
4. Claim 14(c)	71
5. Claim 14(d)	71
6. Claim 14(e)	71
F. Claim 15	72
1. Claim 15(Pre)	72
2. Claim 15(a)	72
3. Claim 15(b)	72
4. Claim 15(c)	72
5. Claim 15(d)	72
6. Claim 15(e)	72
G. Claim 16	72
1. Claim 16(Pre)	72
2. Claim 16(a)	72
3. Claim 16(b)	72
4. Claim 16(c)	73
5. Claim 16(d)	73
6. Claim 16(e)	73
IX. GROUND 5: CLAIM 4 IS OBVIOUS OVER <i>JIANG</i> , <i>BAUMGARTNER</i> , <i>BROWN</i> , <i>SMYTH</i> , AND <i>KAZEROONIAN</i>	73

TABLE OF CONTENTS
(Continued)

	Page
X. GROUND 6: CLAIMS 9-10 ARE OBVIOUS OVER JIANG, BAUMGARTNER, BROWN, SMYTH AND VENSKO	73
A. Claim 9	73
B. Claim 10	73
XI. THE BOARD’S DISCRETION UNDER 35 U.S.C. §314(a) and §325(d).....	74
XII. CONCLUSION.....	75
XIII. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8(A)(1)	76
A. Real Party-In-Interest	76
B. Related Matters.....	76
C. Lead and Back-Up Counsel.....	76
D. Service Information.....	77
CLAIM LISTING APPENDIX	78

TABLE OF AUTHORITIES

	Page(s)
CASES	
<i>Apple Inc. v. Fintiv, Inc.</i> , IPR2020-00019 (PTAB, Mar. 20, 2021)	73
<i>General Plastic Industrial Co., Ltd. v. Canon Kabushiki Kaisha</i> , IPR2016-01357, Paper 19 (PTAB Sept. 6, 2017).....	73
<i>Phillips v. AWH Corp.</i> , 415 F.3d 1303 (Fed. Cir. 2005) (<i>en banc</i>)	5
STATUTES AND RULES	
35 U.S.C. §102(b)	<i>passim</i>
35 U.S.C. §314(A)	73
35 U.S.C. §315(c)	1, 4
35 U.S.C. §325(D)	73, 74
OTHER AUTHORITIES	
37 C.F.R. §42.100(b)	5
37 C.F.R. §42.104	4, 5
37 C.F.R. §42.122	4
83 Fed. Reg. 197 (Oct. 11, 2018).....	5
<i>Dictionary of Electrical and Computer Engineering</i>	27, 37
Manual of Patent Examining Procedure (“MPEP”) 2114	56, 58
<i>Microsoft Computer Dictionary</i>	27

EXHIBIT LIST (37 C.F.R. § 42.63(E))

Exhibit	Description
1001	U.S. Patent No. 7,979,277 (“the ’277 Patent”)
1002	Prosecution History for the ’277 Patent (“’277 File History”)
1003	Declaration of Christopher Schmandt (“Dec.”)
1004	U.S. Patent No. 6,374,219 to Jiang (“Jiang”)
1005	U.S. Patent No. 5,819,222 to Smyth et al. (“Smyth”)
1006	U.S. Patent No. 6,959,376 to Boike et al. (“Boike”)
1007	U.S. Patent Application Publication No. 2002/0049582 to Baumgartner et al. (“Baumgartner”)
1008	Harry Newton, <i>Newton’s Telecom Dictionary</i> , 19 th Edition CMP Books, March 2003 (“Newton”)
1009	Kevin L. Kloker, <i>The Motorola DSP56000 Digital Signal Processor</i> , IEEE Micro. Vol. 6, Issue 6, pp. 29-48, December 1986 (“Kloker”)
1010	Christopher Schmandt, <i>Voice Communication with Computers</i> , Van Nostrand Reinhold, 1994 (“Schmandt”)
1011	U.S. Patent No. 6,499,012 to Peters et al. (“Peters”)
1012	Ph.D. Thesis of Mosur Ravishankar (“Ravishankar”)
1013	Intentionally Left Blank
1014	<i>Zentian, Ltd. v Apple Inc.</i> , WDTX Waco (6:22-cv-122) Complaint “Complaint”
1015	Lawrence Rabiner, <i>Fundamentals of Speech Recognition</i> , Pearson Education, 1993 (“Rabiner”)
1016	Richard L. Klevans, <i>Voice Recognition</i> , Artech House Inc. 1997 (“Klevans”)

Petition for *Inter Partes* Review of
U.S. Patent No. 7,979,277

Exhibit	Description
1017	U.S. Patent No. 7,120,582 to Young et al. (“ <i>Young</i> ”)
1018	John Holmes, <i>Speech Synthesis and Recognition</i> , 2 nd Edition, Taylor & Francis, 2001 (“ <i>Holmes</i> ”)
1019	U.S. Patent No. 4,926,488 to Nadas et al. (“ <i>Nadas</i> ”)
1020	Frederick Jelinek, <i>Statistical Methods for Speech Recognition</i> , The MIT Press, 1997 (“ <i>Jelinek</i> ”)
1021	Shigeru Katagiri, <i>Handbook of Neural Networks for Speech Processing</i> , Artech House Inc. 2000, (“ <i>Katagiri</i> ”)
1022	Lawrence R. Rabiner, A Tutorial on Hidden Markov Models and <i>Selected Applications in Speech Recognition</i> , Proceedings of the IEEE, VOL. 77, NO. 2, pp. 257-286, February 1989 (“ <i>Rabiner89</i> ”)
1023	U.S. Patent No. 6,182,037 to Maes (“ <i>Maes</i> ”)
1024	U.S. Patent Application Publication No. 2003/0220796 to Aoyama et al. (“ <i>Aoyama</i> ”)
1025	U.S. Patent No. 5,895,447 to Ittycheriah et al. (“ <i>Ittycheriah '447</i> ”)
1026	U.S. Patent No. 6,092,045 to Stubbley et al. (“ <i>Stubbley</i> ”)
1027	U.S. Patent No. 6,151,574 to Lee et al. (“ <i>Lee</i> ”)
1028	U.S. Patent No. 6,580,814 to Ittycheriah et al. (“ <i>Ittycheriah</i> ”)
1029	U.S. Patent No. 6,452,348 to Toyoda (“ <i>Toyoda</i> ”)
1030	U.S. Patent No. 6,195,634 to Dudemaine et al. (“ <i>Dudemaine</i> ”)
1031	U.S. Patent No. 5,893,058 to Kosaka (“ <i>Kosaka</i> ”)
1032	U.S. Patent Application Publication No. 2001/0010039 to Yang. (“ <i>Yang</i> ”)
1033	U.S. Patent No. 5,754,878 to Asghar et al. (“ <i>Asghar</i> ”)
1034	U.S. Patent No. 6,041,400 to Ozcelik et al. (“ <i>Ozcelik</i> ”)

Petition for *Inter Partes* Review of
U.S. Patent No. 7,979,277

Exhibit	Description
1035	Declaration of June Ann Munford (“ <i>Munford</i> ”)
1036	U.S. Patent No. 5,699,456 to Brown et al. (“ <i>Brown</i> ”)
1037	International Publication No. WO 99/41684 to Kazeroonian et al. (“ <i>Kazeroonian</i> ”)
1038	U.S. Patent No. 4,567,606 to Vensko et al. (“ <i>Vensko</i> ”)
1039	U.S. Patent No. 6,182,203 to Simar, Jr. et al. (“ <i>Simar</i> ”)
1040	[RESERVED]
1041	[RESERVED]
1042	[RESERVED]
1043	[RESERVED]
1044	[RESERVED]
1045	[RESERVED]
1046	[RESERVED]
1047	U.S. Patent No. 6,879,954 to Nguyen et al. (“ <i>Nguyen</i> ”)
1048	[RESERVED]
1049	[RESERVED]
1050	Binu K. Mathew et al., A Gaussian Probability Accelerator for SPHINX 3, (“Mathew I”)
1051	Binu K. Mathew et al., A Low-Power Accelerator for the SPHINX 3 Speech Recognition System, (“Mathew II”)
1052	U.S. Patent No. 4,718,094 to Bahl et al. (“ <i>Bahl</i> ”)
1053	Hermann Ney et al. Dynamic Programming Search Strategies: From Digit Strings to Large Vocabulary Word Graphs C.-H. Lee et al. (eds.), Automatic Speech and Speaker Recognition (“Ney”)

Petition for *Inter Partes* Review of
U.S. Patent No. 7,979,277

Exhibit	Description
1054	U.S. Patent No. 7,058,678 to Luick et al. (“ <i>Luick</i> ”)
1055	Dictionary of Electrical and Computer Engineering, 6 th Edition, McGraw-Hill, 2003 (“Dictionary of Electrical and Computer Engineering”)
1056	Microsoft Computer Dictionary, 5th Edition. Microsoft Press, 2002 (“Microsoft Computer Dictionary”)
1057	Intentionally Left Blank
1058	Abdulmesih Aktas et al, <i>Multi-DSP and VQ-ASIC Based Acoustic Front-End for Real-Time Speech Processing Tasks</i> , First European Conference on Speech Communication and Technology EUROSPEECH ’89, Paris, France, September 1989 (<i>Atkas 1989</i>)
1059	Abdulmesih Aktas et al., <i>Speaker Independent Continuous Hmm-Based Recognition of Isolated Words on A Real-Time Multi-Dsp System</i> , Second European Conference on Speech Communication and Technology EUROSPEECH ’91, Genova, Italy, September 1991 (<i>Atkas 1991</i>)
1060	U.S. Patent No. 6,029,221 to Wu et al. (“ <i>Wu</i> ”)
1061	<i>Zentian, Ltd. v Apple Inc.</i> , WDTX Waco (6:22-cv-122) Scheduling Order “ <i>Scheduling Order</i> ”
1062	Declaration of Les Atlas, Ph.D.
1063	<i>Zentian Ltd. v. Amazon.com Services LLC, et al.</i> , No. 6:22-cv-123 (W.D. Tex.) Mar. 22, 2023 Amended Scheduling Order

This Petition raises grounds of unpatentability identical to those in already-instituted IPR2023-00034. It neither adds to nor alters any arguments the Board has already considered, will not expand the instituted grounds, and is accompanied by a timely motion for joinder under 35 U.S.C. § 315(c).

I. INTRODUCTION

Petitioners Amazon.com Services LLC and Amazon Web Services, Inc. (“Petitioners” or “Amazon”) request *Inter Partes* Review of Claims 1, 4-5, 7, 9-10, 12, and 14-16 (“Challenged Claims”) of USPN 7,979,277, assigned to Zentian Limited upon information and belief. ’277 *Patent* (Ex. 1001). The purportedly distinguishing feature of the Challenged Claims—allocating speech recognition steps to separate processors and pipelining data—was well-known before the ’277 Patent, and the Challenged Claims are obvious over the prior art. Accordingly, IPR of the Challenged Claims should be instituted.

II. SUMMARY OF THE ’277 PATENT

A. Patent Description

The ’277 Patent claims a speech recognition circuit utilizing a front end generating feature vectors from an audio signal, a calculating circuit calculating distances indicating similarity between feature vectors and predetermined acoustic states of an acoustic model, and a search stage identifying words within a lexical tree. ’277 *Patent* (Ex. 1001), 1:4-5, 3:51–4:18, 10:35-61, 11:65–12:10, Fig. 16. Data

is pipelined from the front end to the calculating circuit to the search stage. *Id.*, 4:8-18.

B. Summary of Unpatentability

Speech recognition commonly includes three general steps: (1) extracting feature vectors from the audio signal; (2) computing a likelihood, for each of the feature vectors, indicating the similarity of the extracted feature vector to modeled spoken speech; and (3) searching a prefix tree using the indicated most likely spoken speech to identify a word. *Dec.*, 97.¹ The '277 Patent's claims recite these three general steps, each of which was well-known in the art. *Id.* The purported novelty is pipelining data through these steps and performing likelihood computation on a processor separate from the processor performing both feature vector extraction and word identification. *Id.* The '277 Patent even admits it was known that likelihood (or probability or distance) determination is "computationally expensive." *Id.*; '277 *Patent*, 1:38-46.

¹ Submitted herewith is the Declaration of Les Atlas, Ph.D., who has reviewed, agrees with, and adopts the opinions in the Schmandt Declaration (IPR2023-00034, EX1003). EX1062, ¶¶ 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.

The below mapping relies on *Jiang* teaching the three general speech recognition steps. *Dec.*, 98. *Jiang* teaches feature vector extraction performed on a first processor and distance (or likelihood) calculation and word identification performed by a “tree search engine 74” via a dedicated processor. *Id.*; *Jiang*, 6:53-57 (disclosing feature extraction module 66 is a “conventional array processor”), 6:38-42 (disclosing tree search engine 74 may be a “dedicated speech recognition processor”). *Jiang* does not expressly teach pipelining data or that likelihood calculations are executed on a separate processor. *Dec.*, 98. However, *Jiang* expressly teaches its speech recognition system may comprise various computer configurations, including multiprocessor systems. *Id.*

In related art, *Brown* teaches pipelining data through the three well-known speech recognition steps, while *Baumgartner* teaches executing likelihood calculations on a separate processor from the processor executing both feature extraction and word identification. *Dec.*, 99-100.

Grounds 4-6 are provided additionally relying on *Smyth* in view of Patent Owner’s infringement contentions in the current litigation. *Smyth* also expressly teaches performing likelihood calculations on its own processor. *Dec.*, 101.

C. Priority Date

The '277 Patent claims priority to a GB application filed September 14, 2004. '277 Patent, (30). For purposes of this Petition only, Petitioners apply September 14, 2004, as the priority date for the Challenged Claims.

D. Level of Skill of a POSITA

A POSITA at the time of the '277 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. *Dec.*, 28-30. Additional education or experience might substitute for the above requirements. *Id.*

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 '277 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 IPR2023-00034, this petition is timely under 35 U.S.C. §315(c) and 37 C.F.R. §42.122.

B. Identification of Challenge (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). Based on the prior

art identified below, IPR of the Challenged Claims should be granted. 37 C.F.R. §42.104(b)(2).

Proposed Grounds of Unpatentability	
<u>Ground 1:</u>	Claims 1, 5, 7, 12, and 14-16 are obvious under §103(a) over <i>Jiang, Baumgartner, and Brown</i>
<u>Ground 2:</u>	Claim 4 is obvious under §103(a) over <i>Jiang, Baumgartner, Brown, and Kazeroonian</i>
<u>Ground 3:</u>	Claims 9-10 are obvious under §103(a) over <i>Jiang, Baumgartner, Brown, and Vensko</i>
<u>Ground 4:</u>	Claims 1, 5, 7, 12, and 14-16 are obvious under §103(a) over <i>Jiang, Baumgartner, Brown, and Smyth</i>
<u>Ground 5:</u>	Claim 4 is obvious under §103(a) over <i>Jiang, Baumgartner, Brown, Smyth, and Kazeroonian</i>
<u>Ground 6:</u>	Claims 9-10 are obvious under § 103(a) over <i>Jiang, Baumgartner, Brown, Smyth, and Vensko</i>

Sections V-X identify where each element of the Challenged Claims is found in the prior art. 37 C.F.R. §42.104(b)(4). The exhibit numbers of the supporting evidence relied upon to support the challenges are provided above and the relevance of the evidence to the challenges raised are provided in Sections V-X. 37 C.F.R. §42.104(b)(5). Exhibits 1001-1063 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*). With the exception discussed below, 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 '277 Patent are indefinite or additional terms need construction.

1. Construction Pursuant to 35 U.S.C. §112, ¶ 6

Claim 14 recites a limitation in means-plus-function format, which creates a rebuttable presumption applicant intended to invoke §112, ¶ 6. Corresponding structure for the means-plus-function limitation is identified below. Petitioners propose the claimed function recited in the limitation be given the plain and ordinary meaning the term would have had to a POSITA.

<i>Claim Term</i>	<i>Support</i>	<i>Structure and Function</i>
Claims 14: “calculating means for calculating a distance indicating the similarity between a feature vector and a predetermined acoustic state of an acoustic model”	’277 Patent, 37:66-67 (stating “calculation means may comprise the distance calculation engine”), 24:6-16 (referring to distance calculation engine as a coprocessor), Fig. 17 (illustrating distance calculation engine 204 as part of distance accelerator subsystem) <i>Dec.</i> , 24-27.	<u>Structure:</u> a processor <u>Function:</u> calculating a distance indicating the similarity between a feature vector and a predetermined acoustic state of an acoustic model

IV. SHOWING OF ANALOGOUS PRIOR ART

A. Jiang

Jiang was neither cited nor considered during prosecution of the ’277 Patent. *Jiang* issued April 16, 2002, qualifying as prior art at least under 35 U.S.C. §102(b). Because *Jiang*, like the ’277 Patent, discloses a “system for recognizing speech based on an input data stream indicative of the speech [which] provides possible words represented by the input data stream as a prefix tree[,]” *Jiang* is in the same field of endeavor as the ’277 Patent. *Compare, Jiang*, Abstract, with ’277 Patent, Abstract; *Dec.*, 102-103.

Because *Jiang*, like the ’277 Patent, recognizes the challenges with performing speech recognition and discloses the use of various hardware configurations for speech recognition, *Jiang* is reasonably pertinent to a problem

described by the '277 Patent. *Compare, Jiang*, 4:60–5:3, with '277 Patent, 1:4-14, 2:3-6; *Dec.*, 102-103.

Therefore, *Jiang* is analogous art.

B. Baumgartner

Baumgartner was neither cited nor considered during prosecution of the '277 Patent. *Baumgartner* published April 25, 2002, qualifying as prior art under at least 35 U.S.C. §102(b).

Like the '277 Patent, *Baumgartner* is related to accelerators for speech recognition systems. *Dec.*, 104; '277 Patent, Abstract, Fig. 21, 30:40-65; *Baumgartner*, Abstract, Fig. 19, [0002]. Accordingly, *Baumgartner* is in the same field of endeavor as the '277 Patent. *Dec.*, 104.

The '277 Patent purportedly addressed the problem of distributing the computational load in large vocabulary speech recognition systems. *Dec.*, 104; '277 Patent, 1:39-46, 2:3-6. *Baumgartner* also addressed the problem of distributing the computational load in large vocabulary speech recognition systems. *Dec.*, 104; *Baumgartner*, [0005]. Accordingly, *Baumgartner* is reasonably pertinent to a problem described by the '277 Patent. *Dec.*, 104.

Therefore, *Baumgartner* is analogous art.

C. Brown

Brown was neither cited nor considered during prosecution of the '277 Patent. *Brown* issued December 16, 1997, qualifying as prior art at least under 35 U.S.C. §102(b).

Like the '277 Patent, *Brown* is related to speech recognition methods using pipelining. *Brown*, Abstract, 1:9-11, 3:35-48; '277 Patent, Abstract, 4:8-15, 30:41-42, Fig. 21. Accordingly, *Brown* is in the same field of endeavor as the '277 Patent. *Dec.*, 105.

Like the '277 Patent, *Brown* addresses the problem of reducing computational requirements in speech recognition systems. Compare '277 Patent, 1:15-35, with *Brown*, 1:34-40. Accordingly, *Brown* is reasonably pertinent to a problem described by the '277 Patent. *Dec.*, 105.

Therefore, *Brown* is analogous art.

D. Kazeroonian

Kazeroonian was neither cited nor considered during prosecution of the '277 Patent. *Kazeroonian* published August 19, 1999, qualifying as prior art at least under 35 U.S.C. §102(b).

Like the '277 Patent, *Kazeroonian* is related to processing data in speech recognizers, including providing multi-threaded processing for efficient processing of multiple tasks. '277 Patent, Abstract, 1:4-7, 2:41-44, 29:29-41; *Kazeroonian*,

Abstract, 6:31-33, 15:33–16:3, 19:26–20:5, 21:9-11, Fig. 2. Accordingly, *Kazeroonian* is in the same field of endeavor as the '277 Patent and reasonably pertinent to a problem described by the '277 Patent. *Dec.*, 106.

Therefore, *Kazeroonian* is analogous art.

E. Vensko

Vensko was neither cited nor considered during prosecution of the '277 Patent. *Vensko* issued January 28, 1986, qualifying as prior art at least under 35 U.S.C. §102(b).

Vensko teaches a speech recognition system and is directed to issuing interrupts for efficient speech recognition processing. *Compare* '277 Patent, Abstract, 27:43-46, *with Vensko*, Abstract, 6:29-33, 10:10-19. Therefore, *Vensko* is in the same field of endeavor and reasonably pertinent to a problem described by the '277 Patent. *Dec.*, 107.

Therefore, *Vensko* is analogous art.

F. Smyth

Smyth was neither cited nor considered during prosecution of the '277 Patent. *Smyth* published October 6, 1998, qualifying as prior art under at least 35 U.S.C. §102(b).

Like the '277 Patent, *Smyth* “relates to speech recognition circuits and methods.” *Compare* '277 Patent, 1:4-5, 12:4-50, 23:61-67, *with Smyth*, 5:20-46,

5:58–6:4, 6:36-63. Therefore, *Smyth* is in the same field of endeavor as the '277 Patent. *Dec.*, 108-109.

Both *Smyth* and the '277 Patent recognize the challenges with performing speech recognition tasks using limited computing resources; therefore, *Smyth* is reasonably pertinent to a problem described by the '277 Patent. *Dec.*, 108-109.

Therefore, *Smyth* is analogous art.

V. GROUND 1: CLAIMS 1, 5, 7, 12, AND 14-16 ARE OBVIOUS OVER JIANG, BAUMGARTNER, AND BROWN

A. Claim 1

1. Claim 1(Pre)

To the extent the preamble is limiting, *Jiang* teaches or renders obvious a speech recognition circuit. *Jiang* teaches “a speech recognition system.” *Jiang*, 2:7-10; *see also* 1:15-19, 4:8-9, 4:30-32, 4:51–6:20, Abstract, Figs. 1-2. *Jiang* discloses a “suitable computing environment in which the invention may be implemented” (4:51-53), which includes a variety of “computer system configurations” (4:60-65), such as multiprocessor systems (4:62-63), functions implemented via computer-executable instructions (4:54-57), and various electronic components and computer hardware (5:4–6:17). *Jiang*, 4:51–5:3, 6:18-45.

To the extent Patent Owner argues *Jiang*’s electronic components and computer hardware must be expressly embodied as a circuit, it would have been obvious to a POSITA to implement *Jiang*’s speech recognition system using

“circuitry.” *Dec.*, 110-111. Per Mr. Schmandt, the ’277 Patent does not ascribe any particular structural features or requirements to “circuit.” *Dec.*, 111. A POSITA would have recognized *Jiang* teaches 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 a circuit or circuitry. *Dec.*, 111. To the extent any description refers to computerized methods stored as software, a POSITA would have found obvious physically implementing those computerized methods using electronic circuitry. *Dec.*, 111. Therefore, a POSITA would have found it obvious and been motivated to embody the various electronic components disclosed in the references as a “circuit” or “circuitry” without undue experimentation. *Dec.*, 111.

2. Claim 1(a)

Jiang teaches or renders obvious an “audio front end,” namely feature extraction module 66 calculating a feature vector from an audio signal. *Jiang*, 1:58-67, 6:46–7:10.

First, *Jiang* teaches an “audio signal.” *Id.* *Jiang* teaches “speech is input into [the speech recognition] system 60 in the form of an audible voice signal provided by the user to microphone 62.” *Jiang*, 6:46-48, 6:18-20 (referencing the speech recognition system 60). The microphone provides an analog signal to an A/D converter, which “samples” the signal at a rate of approximately 16kHz into

“a plurality of digital samples.” *Jiang*, 6:46–7:1. The stream (a “plurality”) of digital samples are provided to a feature extraction module 66. *Jiang*, 6:50-53. Because *Jiang* teaches a plurality of digital samples converted from speech input to the speech recognition system, *Jiang* teaches an “audio signal,” as claimed. *Dec.*, 112-114.

Second, *Jiang* teaches feature extraction module 66 “calculate[s] a feature vector from an audio signal[.]” In detail, *Jiang* states:

Feature extraction module 66 divides the digital signal received from A/D converter 64 into **frames** which include **a plurality of digital samples**. Each frame is approximately 10 milliseconds in duration. The frames are then preferably **encoded by feature extraction module 66 into a feature vector** reflecting the spectral characteristics for a plurality of frequency bands...feature extraction module 66 also preferably encodes the feature vectors into one or more **codewords** using vector quantization techniques and a codebook derived from training data. Thus, **feature extraction module 66** provides, at its output the **feature vectors (or codewords)** for each spoken utterance.

Jiang, 6:62–7:10. *Jiang* thus teaches the feature extraction module encodes the frames of digital audio samples into feature vectors, which are further encoded (using vector quantization techniques) into codewords. *Dec.*, 115-116. Per Mr. Schmandt, the encoding of the frames of digital audio samples into feature vectors teaches or renders obvious the claimed “for calculating a feature vector ...,” as *Jiang* expressly teaches a feature vector is encoded, i.e., calculated. *Dec.*, 116.

Because *Jiang's* feature extraction module 66 calculates (or generates or encodes) a feature vector from a frame of a digital audio sample, *Jiang* teaches or renders obvious an “audio front end” performing the claimed functionality. *Dec.*, 115-117.

3. Claim 1(b)

Jiang teaches or renders obvious Claim 1(b).

a) The '277 Patent's Discussion of “extracted” or “derived quantities”

The '277 Patent specification does not use the terms “extracted” or “derived quantities.” '277 *Patent*, 5:35-40, 5:58-63, 6:4-10, 6:17-23, 6:32-34. Instead, the '277 Patent describes “calculating” a feature vector “by splitting the audio signal into **spectral components**.... [T]he audio input signal is split into 12 frequency bands and one energy band, and a feature vector is constructed using these 13 components, plus their first and second derivatives.” '277 *Patent*, 12:5-10.

Per Mr. Schmandt, a POSITA would have understood Claim 1(b) to be referencing the '277 Patent's discussion of the **spectral components** of the feature vector and the component derivatives. *Dec.*, 118-119 (further discussing that cepstral coefficients are the spectral components of an audio signal, known to indicate the extracted and derived quantities of the spectral components). Mr. Schmandt explains that an audio signal includes spectral components or characteristics split among frequency bands. *Dec.*, 119.

b) *Jiang's* Teachings

Jiang teaches or renders obvious the calculated feature vector comprises extracted and/or derived quantities in the form of “spectral characteristics[.]” *Jiang* teaches: “[t]he frames are then preferably encoded by feature extraction module 66 into a **feature vector** reflecting **the spectral characteristics for a plurality of frequency bands.**” *Jiang*, 6:62–7:10, 6:53-57. Like ’277 *Patent*, 12:5-10, *Jiang* also discloses computing a magnitude value for “each **frequency band** of a frequency spectrum.” *Jiang*, 6:53-57. *Jiang's* feature extraction module 66 thus *extracts* quantities that are encoded into a feature vector and are indicative of the spectral characteristics of the audio signal, as expressly taught by *Jiang. Dec.*, 120-123 (opining speech captured in 10ms audio samples are Fast Fourier Transformed to extract/derive quantities from the spectral components). Thus, because *Jiang* teaches feature extraction module 66 encodes frames of the plurality of digital samples into a feature vector reflecting spectral characteristics, *Jiang* teaches or renders obvious the feature vector comprises a plurality of extracted and/or derived quantities from the audio signal, as claimed. *Dec.*, 122.

Jiang further teaches the feature vector comprises extracted/derived quantities of the audio signal “during a defined audio time frame[.]” Specifically, *Jiang* teaches that the digital signal is divided into frames of approximately 10ms in duration, which are then encoded by the feature extraction module 66 into a feature vector,

where the feature vector “reflect[s] the spectral characteristics[.]” *Jiang*, 6:62–7:1. Therefore, each frame represents 10ms of speech input that is encoded as a feature vector. *Dec.*, 123.

Because *Jiang* teaches a frame of audio input of 10ms in duration, i.e., a “defined audio time frame[.]” and further because *Jiang* teaches encoding, using the feature *extraction* module (which is mapped as the “audio front end”), a feature vector of this frame that includes *extracted* quantities from the audio signal, *Jiang* teaches or renders obvious Claim 1(b).

4. *Claim 1(c)*

Jiang, alone or in combination with *Baumgartner*, teaches or renders obvious Claim 1(c). *Jiang* teaches or renders obvious a “calculating circuit,” in addition to teaching the claimed “calculating distances[.]” To the extent Patent Owner argues *Jiang* does not teach a “calculating *circuit*,” *Baumgartner* teaches a speech label accelerator (SLA), where the “heart of the SLA is a **circuit**” performing probability (distance) calculations.

Petitioners note that Claim 1(e) further recites the calculating circuit is “implemented using a second processor[.]” *Baumgartner* is relied on for Claim 1(e) for teaching such, as *Baumgartner* expressly discloses a calculating circuit implemented on a processor distinct from a processor performing the feature vector extraction and word identification. *See* Section V.A.6.

To the extent Patent Owner argues *Jiang* does not teach or render obvious a calculating circuit performing the claimed functionality of calculating distances, Grounds 3-4 relying on *Smyth* are provided.

a) '277 Patent's Discussion of "distances"

The '277 Patent describes "distances" as "probabilities" or "likelihoods": **"a distance is an indication of the probability or likelihood** that a feature vector corresponds to a particular state. An important class of distance computation in speech recognition is the calculation of **output state probabilities** in recognisers using **Hidden Markov Models.**" '277 Patent, 4:13-17, 3:51-59, 12:38-49 ("The distance calculator stage of the recognition process computes a probability or likelihood that a feature vector corresponds to a particular state."). Figure 16 illustrates the distance calculation engine outputting "Distances or Probabilities or Likelihoods":

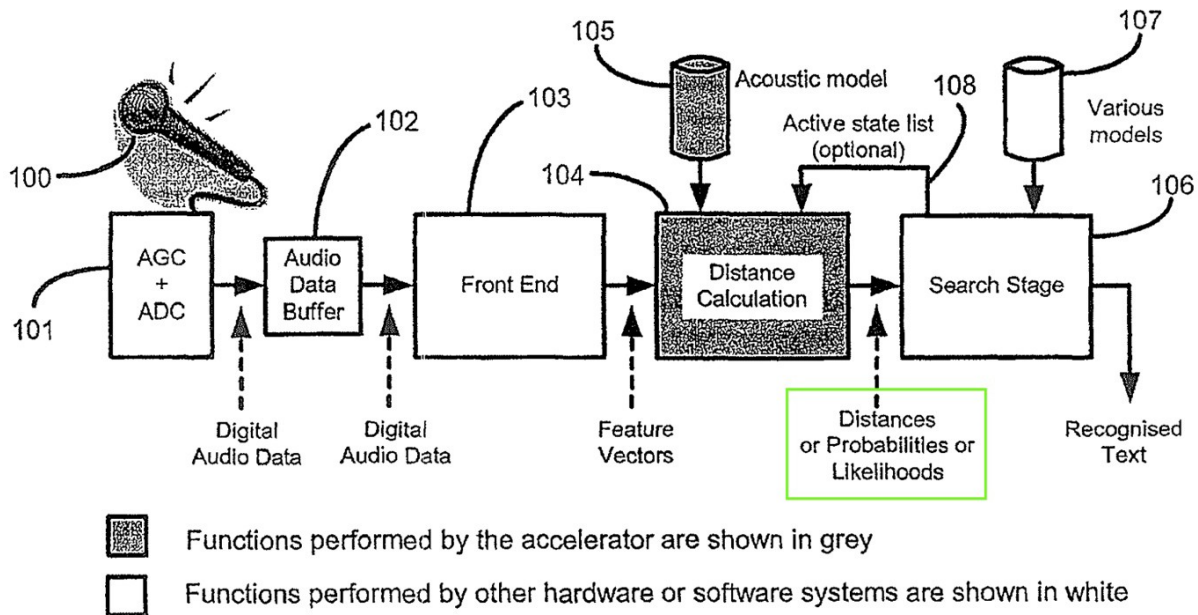


Figure 16

Id., Fig. 16 (annotated). The '277 Patent describes “a wide variety of distance metrics used” that are “widely documented in the literature.” *Id.*, 3:57-60.

Per Mr. Schmandt, the '277 Patent’s description of a “distance” as a probability or likelihood of a feature vector compared to a particular state of an HMM is consistent with the well-known use of HMMs in speech recognition. *Dec.*, 126-128. Specifically, for each state of HMM, a distance (i.e., a probability or likelihood) is computed, such that a distance indicates the similarity of the spoken speech to the state sequence or path. *Id.* (further opining that a POSITA would have understood that probability reflects a “distance” or similarity in that a high probability match indicates a small distance or difference between the expected

HMM output and the speech input feature vector, whereas a low probability reflects a greater distance).

b) *Jiang's* Teachings

(1) "a calculating circuit"

Jiang teaches or renders obvious a "calculating circuit[.]" *Jiang* teaches a tree search engine 74 for performing the functionality recited in Claim 1(c). *Jiang's* tree search engine 74 performs both the distance calculation of Claim 1(c) and the word identification of Claim 1(d). *Jiang*, 7:30-42, 8:16-42. *Jiang* discloses the tree search engine "is preferably implemented in CPU 21 (which may include one or more processors) or may be performed by a dedicated speech recognition processor employed by personal computer 20." *Jiang*, 6:39-42.

Mr. Schmandt opines *Jiang's* tree search engine 74, which may be a dedicated processor, teaches or renders obvious a circuit. *Dec.*, 129-131 (further opining a POSITA would have understood a "circuit" is met by a processor); Section V.A.1. Therefore, the tree search engine performing the distance calculation functionality recited in Claim 1(c) teaches or renders obvious a "calculating circuit[.]" *Id.*

As discussed for Claim 1(e), *Baumgartner* teaches the calculating circuit implemented as a second processor.

(2) **“a plurality of predetermined acoustic states of an acoustic model”**

Jiang teaches “a plurality of predetermined acoustic states of an acoustic model[.]”

Acoustic model: First, *Jiang* teaches an “acoustic model,” namely an HMM. *Jiang*, 7:30-47; *Dec.*, 132 (noting the ’277 Patent describes an HMM as an acoustic model). *Jiang* describes the HMMs as “phonetic speech unit models...which represent speech units to be detected by system 60.” *Jiang*, 7:33-36.

A POSITA would have understood *Jiang*’s HMMs, as phonetic speech unit models, are “acoustic models.” *Dec.*, 133. First, HMMs were well known to be acoustic models. *Dec.*, 133 (*citing Ravishankar*, 19-20, 27-28, 41). Second, *Jiang* expressly discloses the HMMs are **phonetic speech unit** models, and a POSITA would have understood **phonetic speech** is **acoustic**. *Jiang*, 2:53-56; *Dec.*, 133. Third, the ’277 Patent describes HMMs as acoustic models, and *Jiang* repeatedly teaches utilizing HMMs for speech recognition. ’277 *Patent*, 3:4-5, 10:56-67, 37:53-55; *Dec.*, 133. Therefore, *Jiang* teaches an “acoustic model.”

A Plurality of Predetermined Acoustic States [of an Acoustic Model]:

Jiang teaches each HMM includes a plurality of predetermined acoustic states: “A hidden Markov model includes a plurality of states” *Jiang*, 1:21-22, 1:43-48; *Dec.*, 134-135 (*citing Ravishankar*, 19-20, 27-28, 41). A POSITA would have understood the speech recognition system taught in *Jiang*, 7:30-47, which

employs HMMs, teaches the HMMs include a plurality of predetermined acoustic states. *Dec.*, 135. *Jiang* expressly states the discussion of speech recognition at 7:30–8:51 is “all done in a known manner[,]” including the known manner disclosed in *Jiang’s* Background that expressly teaches each HMM has a plurality of predetermined acoustic states. *Dec.*, 135 (*citing Jiang*, 1:15–2:15, 8:51).

The HMM states are *predetermined* because they are generated and stored prior to the claimed “distance” calculation. *Jiang*, 7:30-47; 11:25-47 (“One or more HMMs are created by training module 65 for each phoneme of each word in a desired vocabulary represented by the training data.”); *Dec.*, 136 (distinguishing the training phase from actual speech determination process, further explaining HMM states are predetermined because they exist in the system prior to the similarity indication, and *citing Ravishankar*, 19-20, 27-28).

(3) “a feature vector”

As discussed in Sections V.A.2-3, *Jiang* teaches encoding each frame of spoken speech into a feature vector. *Jiang*, 6:62–7:10 (“The frames are then preferably encoded by feature extraction module 66 into a **feature vector** reflecting the spectral characteristics for a plurality of frequency bands.”); *Dec.*, 137. *Jiang* also teaches “encod[ing] the feature vectors into one or more codewords using vector quantization techniques and a codebook derived from training data.” *Jiang*, 7:1-10. Per Mr. Schmandt, vector quantization was a well-known technique to reduce the

search space by clustering feature vectors into similar codewords, where each feature vector was vector quantized encoded into a codeword. *Dec.*, 138 (*citing Ravishankar*, 48-50 and *Brown*).

Jiang describes speech recognition as determining the phoneme(s) represented by the codeword(s) using HMM(s). *Jiang*, 7:38-43. A POSITA would have understood that because the codewords are vector quantized feature vectors, then *Jiang's* teaching of determining the phonemes of the codewords is likewise determining the phonemes of the feature vectors. *Dec.*, 139. Thus, *Jiang's* teaching of performing speech recognition of codewords also teaches performing speech recognition of feature vectors, because the codewords are encoded feature vectors. *Id.*

(4) “calculating distances ...”

Jiang teaches “calculating distances ...,” namely a *score* for each phoneme branch of the prefix tree, where the “**score** represents the **likelihood** that the particular **phoneme** encoded by the **codeword** corresponds to the phoneme for the branch under consideration.” *Jiang*, 8:29-33. *Jiang* describes how the score for each phoneme branch is determined:

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

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

Jiang, 7:30-47; *Dec.*, 140-142.

Jiang provides a discussion in its Background of using HMMs to calculate a distance indicating a similarity between a feature vector and a plurality of HMM states, where the distance in *Jiang* is referred to as a “score” or “probability score.” *Jiang*, 1:19–2:15; *Dec.*, 143-147 (explaining *Jiang*’s scores and how *Jiang*’s Background provides additional teachings informing a POSITA that the calculated “score” is a probability or likelihood representative of a similarity or distance measurement). For example, *Jiang* teaches utilizing HMMs to “model observed sequences of speech spectra, where specific spectra are **probabilistically associated with a state in an HMM.**” *Jiang*, 1:41-48. *Jiang* also teaches a “**probability score is computed for each state in the state sequences (i.e., the HMMs)** being considered.” *Jiang*, 2:5-7; *see also* Section V.A.4.a (describing the ’277 Patent “distance” as “an indication of the probability or likelihood that a feature vector corresponds to a particular state”). Thus, *Jiang*’s collective teachings indicate to a

POSITA that *Jiang*'s "score [] indicative of a likelihood that the particular phoneme being examined is the actual phoneme indicated by the codeword received from the feature extraction module" comprises a "distance" as recited in Claim 1(c). *Jiang*, 8:20-23; *Dec.*, 143-147. Therefore, *Jiang* teaches "calculating distances[,]" as claimed.

Jiang's teaching of determining the likelihood a *codeword* is a particular phoneme would also have been understood by a POSITA to teach calculating distances indicating the similarity between a *feature vector* and HMM states. *Dec.*, 148. As discussed above for Claim 1(a), *Jiang* teaches encoding the frames of spoken speech as feature vectors and, in some embodiments, encoding the feature vectors as codewords using vector quantization techniques. *Jiang*, 6:64–7:10, 1:58-67. The teachings at *Jiang*, 7:38-43 and 8:29-36 both refer to determining the likelihood of a phoneme represented by a *codeword*. As discussed for Claim 1(a), because the codeword is simply a vector quantized feature vector, a POSITA would have understood *Jiang*'s teachings of determining the likelihood the considered phoneme is indicated by the *codeword* to likewise be determining the likelihood the considered phoneme is indicated by one or more *feature vectors*. *Dec.*, 148-149. Thus, a score indicative of the likelihood that a codeword corresponds to the phoneme for the branch under consideration is a distance calculation (i.e., score

indicative of likelihood) between a feature vector (i.e., codeword) and a plurality of predetermined acoustic states of an acoustic model (i.e., states of HMMs).

Jiang's teachings at 7:30-45 and 8:16-51 would also have been understood to teach determining the similarity between the feature vector (or codeword) and “a plurality of predetermined acoustic states of an acoustic model[.]” As quoted above, *Jiang* teaches a “**probability score is computed for each state in the state sequences (i.e., the HMMs)** being considered.” *Jiang*, 2:5-7; *Dec.*, 149 (opining because *Jiang* teaches computing a probability score *for each state* of an HMM, *Jiang* teaches computing a probability score for a plurality of states of an acoustic model). HMMs are used to determine the most likely phoneme represented by the codeword and hence representative of the user's utterance. *Jiang*, 7:37-42. HMMs include a plurality of states, and HMMs are acoustic models. *See* Section V.A.4.b.2. A score is computed for each phoneme. These scores are then used to indicate the likelihood that the codeword corresponds to the branch under consideration in the lexical (“prefix”) tree. *Jiang*, 8:16-42. Because *Jiang* teaches HMMs are used to determine the most likely phoneme from a feature vector (*Jiang*, 7:30-45), *Jiang* teaches the calculated score indicates the similarity (referred to in *Jiang* as the probability or likelihood) between at least one feature vector (the at least one codeword) and the HMM states. *Dec.*, 150.

Jiang further teaches or renders obvious calculating distances (plural). As quoted above, *Jiang* teaches a “**probability score is computed for each state in the state sequences (i.e., the HMMs)** being considered.” *Jiang*, 2:5-7.

Further, *Jiang* teaches calculating a score for each respective node. *Dec.*, 151-154 (explaining that because *Jiang*, 8:28-32 teaches computing a score for **each** phoneme branch under consideration, then each phoneme branch is scored, resulting in calculating a plurality of distances, and discussing how *Jiang*’s threshold teachings at 11:1-15 further teach calculating more than one distance). Therefore, for each utterance (which is in turn encoded as a feature vector), a score is computed for each node indicating the likelihood the phoneme for the node is the phoneme encoded by the codeword. *Jiang*, 8:16-42; *Dec.*, 151-154. Because there are multiple nodes in the prefix tree (*see, e.g., Jiang*, Fig. 3), there are multiple scores, or distance calculations computed. *Dec.*, 151-154.

Finally, as mapped in Section V.A.4.b.1, *Jiang*’s tree search engine is a “calculating circuit[.]” *Jiang* further teaches the tree search engine is the component calculating distances because it is the component accessing the HMMs and determining the most likely phoneme. *Jiang*, 7:32-43, 8:28-32; *Dec.*, 152. As mapped for Claim 1(e), *Jiang* in combination with *Baumgartner* teaches the claimed architecture of the calculating circuit implemented using a second processor.

c) *Baumgartner* Teaches a “calculating circuit”

As noted above, *Baumgartner* is relied on for Claim 1(c) to the extent Patent Owner argues *Jiang* does not teach a “calculating circuit[.]” *Baumgartner* teaches a “calculating circuit,” namely a circuit implemented by Speech Label Accelerator (SLA) for determining the probability for each observation vector (i.e., feature vector). *Dec.*, 155-157 (opining *Baumgartner*’s “observation vectors” are “feature vectors” as known in the art). The SLA is mapped as the “second processor” recited in Claim 1(e).

Baumgartner teaches an “acoustic model” comprising a plurality of probability density functions (pdfs) identified by the function $P(\cdot | l)$, where each pdf is a leaf of a decision tree, and the acoustic model M comprises the pdfs. *Baumgartner*, [0040] (teaching an “acoustic model is a set of d-dimensional probability density functions or pdfs”), [0042] (teaching the P function is a leaf “because such functions are distributed among the leaves of a family of decision trees”). *Baumgartner* further teaches that given a single observation vector associated with a single frame of speech, the *likelihoods* of the observation vectors to the leaf are ordered by the probability function associated with each leaf in the acoustic model. *Baumgartner*, [0043]; *Dec.*, 158. Thus, *Baumgartner* teaches that each observation vector is associated with a probability or likelihood. *Dec.*, 155-158.

Baumgartner teaches:

The SLA operates on this sequence of observation vectors, one vector at a time. The SLA performs the work of labeling block 130. The labeling block 130 performs a computation that is discussed below and determines **probabilities** 135 for a number of phones. Phones are small, indivisible acoustic elements of speech.

Baumgartner, [0038]. Thus, *Baumgartner* teaches determining a probability for a particular phone, where the probability is determined by the SLA. *Dec.*, 159 (opining that probability indicates how good a match the observed data from the feature vector is to a particular HMM state, and hence, shows how closely the HMM state matches, i.e., the “distance”).

Baumgartner also teaches the SLA is or includes a calculating *circuit*, as claimed. *Baumgartner* teaches: “the heart of the SLA is a **circuit** that permits the fast approximate computation of a large collection of values $P=\{P(v|l)\}$ for a fixed observation vector v .” *Baumgartner*, [0043]; *Dec.*, 160. Per Mr. Schmandt, the SLA’s circuit performing the computation of the values P , i.e., the probabilities, for each observation vector (i.e., feature vector), is a “calculating *circuit*,” as claimed, because the SLA’s circuit is determining probabilities or likelihoods of observation vectors to leaves (or states) in an “acoustic model M .” *Dec.*, 159-160 (*citing Microsoft Computer Dictionary*, 5-8; *Dictionary of Electrical and Computer Engineering*, 6-7).

d) Motivation to Combine *Jiang* with *Baumgartner*

See Section V.A.6.f, which provides a motivation to modify *Jiang* to include *Baumgartner*'s hardware architecture and specifically a second processor (i.e., *Baumgartner*'s SLA including circuitry) for performing the distance or likelihood calculations, as taught for Claim 1(c) and mapped in the preceding section.

5. Claim 1(d)

Jiang teaches or renders obvious Claim 1(d).

a) “a lexical tree”

Jiang's prefix tree teaches a “lexical tree.” *Jiang* teaches speech recognition systems are trained, which create acoustic (or phonetic) HMM models, which are then organized into a tree:

One or more HMMs are created by training module 65 **for each phoneme** of each word in a desired vocabulary represented by the training data. **The HMMs thus represent model output distributions associated with the phonemes** in the vocabulary. **The prefix tree is then formed**, based upon the desired vocabulary, such that **phonemes are structured to provide a leaf** which is associated with each word in the desired vocabulary.

Jiang, 11:25-47, 7:30–8:51; *Dec.*, 162 (citing *Ravishankar*, 33, which teaches “organizing HMMs to be searched as a phonetic tree...is probably the most often cited improvement in search techniques in use currently”). Each phoneme is associated with a phoneme branch, and a collection of phonemes is structured to

provide a leaf, where each leaf is a word in the prefix tree. *Jiang*, 7:67–8:4, 9:37-42 (teaching each leaf on the tree represents a word), 11:40-43; *Dec.*, 163-164. Each branch is associated with one or more phonetic models, such as an HMM. *Jiang*, 2:53-67. Therefore, the prefix tree utilizes a plurality of HMMs, where each node and associated branch of the prefix tree is associated with one or more HMMs used to determine a phoneme. *Dec.*, 164.

A POSITA would have understood *Jiang*'s prefix tree is a “lexical tree,” as claimed. *Dec.*, 165. *Jiang* expressly teaches “[a] prefix tree 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.” *Jiang*, 2:47-52. Because the words of lexicon 70 are “presented to tree search engine 74 in the form of a prefix tree” comprising branches and nodes, as is conventional, *Jiang*'s prefix tree is a “lexical tree.” *Dec.*, 165 (*citing Ravishankar*, 33-34).

b) “the lexical tree comprising a model of words”

Jiang teaches the prefix tree comprises a model of words, namely all the words likely to be encountered by the speech recognition system:

A prefix tree 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.**

Jiang, 2:49-52. *Jiang* further teaches the lexicon provides the model of words:

[L]exicon 70 contains information which is representative of **all of the words in the vocabulary** of speech recognition system 60. The words are **preferably presented to tree search engine 74 in the form of a prefix tree** which can be traversed from a root to a leaf (or to an internal word node) to **arrive at the word most likely indicative of the utterance of the user.**

Jiang, 7:65–8:4; *see also* 7:48-54. Thus, *Jiang* expressly teaches the prefix tree is a model of the words in lexicon 70. *Dec.*, 166.

c) “a search stage for using said calculated distances to identify words within a lexical tree”

Jiang discloses tree search engine 74 includes a search stage of traversing prefix tree to identify words within the prefix tree. *Jiang*, 8:16-43. The search stage uses the calculated distances, i.e., the score indicative of the most likely phoneme for a particular node and associated phoneme branch, to identify “a word which most likely represents the codewords received by feature extraction module[.]” *Jiang*, 7:49-54.

Jiang, 8:5-43 teaches a prefix tree is traversed to identify the word, where the possible words are constrained by the lexicon 70, and the tree comprises a plurality of phoneme branches corresponding to the constrained set of words from the lexicon. *Jiang*, 8:5-15, Figs. 3-6. As the tree is traversed, a score is determined for each node and associated phoneme branch to determine the likelihood the phoneme being

examined is the phoneme indicated by the codeword. *Jiang*, 8:16-24, 8:29-33; *Dec.*, 167 (noting *Jiang*'s similarity to *Ravishankar*, 33).

Jiang teaches searching a lexicon 70 using the computed probability score of the most likely uttered phoneme to determine the word. *Jiang*, 7:48-54. The lexicon “contains information which is representative of all of the words in the vocabulary of speech recognition system 60.” *Jiang*, 7:65-67. Thus, the lexicon 70 contains the words searched by the tree search engine in the prefix tree. *Jiang*, 6:67–7:4. *Jiang* further teaches “[t]he **information** received by tree search engine 74 **based on its accessing of the phonetic speech unit models** in memory 72 **is used in searching lexicon 70 to determine a word which most likely represents the codewords received by feature extraction module 66” *Jiang*, 7:49-54.**

Per Mr. Schmandt, the “information...based on [] accessing of the phonetic speech unit models” includes, at the least, the likelihood score discussed at *Jiang*, 8:16-42 using HMMs discussed at *Jiang*, 7:30-47. *Dec.*, 168. That is, *Jiang* teaches using HMMs to calculate the score indicative of a likelihood the phoneme being examined in the prefix tree is the actual phoneme indicated by the codeword (i.e., the vector quantized feature vector). *Jiang*, 8:16-24. The determination of the phoneme indicated by the codeword utilizes the HMM states of *Jiang*, 7:34-43. Thus, the information received by tree search engine based on accessing the phonetic speech unit models (i.e., the “HMMs which represent phonemes”) (*Jiang*, 7:35-37)

is “used in searching lexicon 70 to determine a word” representative of the utterance.
Jiang, 7:51-52.

Jiang’s discussion of the prefix tree further supports that the score for the most likely phoneme is used in determining the word. *Jiang* teaches that feature extraction module 66 divides an uttered word into constituent phonemes, e.g., AO-R-IX-N-JH for the word ORANGE. *Jiang*, 8:25-28. *Jiang* teaches:

[T]ree search engine 74 computes a score for node 82 which indicates that the first codeword under consideration is **highly likely** to be represented by the **AO phoneme** corresponding to the branch under consideration. Tree search engine 74 also preferably computes a score for each of the other nodes 84 and 86 in tree 77, wherein the score is indicative of the likelihood that the codeword being analyzed is represented by the phonemes AE, and T. Under ideal circumstances, **the score assigned to nodes 84 and 86 is lower than the score assigned to node 82.**

Jiang, 8:32-43. Per Mr. Schmandt, in this example, the system is sequentially recognizing the phonemes AO-R-IX-N-JH to identify the word ORANGE. *Dec.*, 169. Each phoneme is identified using one or more distances. That is, the score for node 82 indicates the phoneme corresponding to the codeword is highly likely to be AO. *Dec.*, 169. Because the phoneme AO was scored as highly likely, then the prefix tree uses the AO-identified phoneme to perform word identification for the word ORANGE. *Dec.*, 169 (*citing Jiang*, 8:5-28).

Yet further, *Jiang* notes that because there are a finite number of phonemes, there is a relatively “small number of initial branches in the tree,” with the initial branches corresponding to the approximately 40-50 known phonemes. *Jiang*, 3:1-21. The score determined for the initial branch is, or is based on, the score used to determine the initial phoneme of the codeword (i.e., the feature vector). *Dec.*, 170 (citing *Jiang*, 8:5-43, discussing the initial nodes include nodes for the phonemes AO, AE, and T and explaining each node would have an assigned score). Thus, at the least, the initial branches of the prefix tree used for word identification use the calculated distances, i.e., one or more probability scores for the beginning one or more phonemes of the word, e.g., the AO, AE, and T shown in Fig. 3. *Id.*

Finally, *Jiang* teaches a “search stage,” as claimed, because the prefix tree is traversed (i.e., searched) to identify the spoken word amongst all the words in the lexicon 70. Indeed, the “search” is identified in the name of the component performing the search stage, i.e., the tree **search** engine is *searching* the tree to identify words within the tree. A POSITA would have understood the traversal of the tree by the tree *search* engine is a “search stage,” as claimed. *Dec.*, 170-171; *Jiang*, 7:67–8:4.

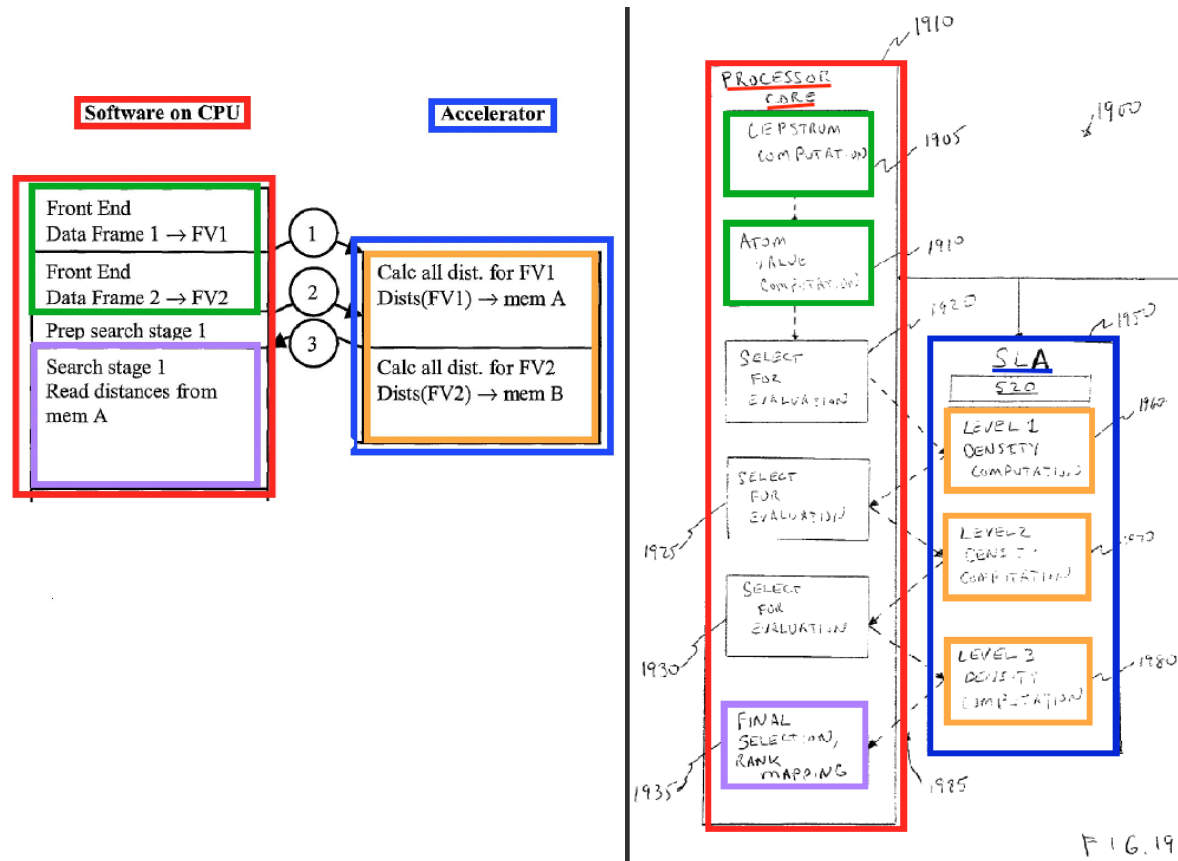
6. *Claim 1(e)*

As discussed above, *Jiang* alone or in combination with *Baumgartner* teaches or renders obvious an “audio front end,” “calculating circuit,” and “search stage.”

See Claims 1(a)-1(d). Although *Jiang* discloses various computer configurations, including multi-processor configurations (*Jiang*, 4:60-65), *Jiang* does not expressly teach the hardware architecture of Claim 1(e), i.e., the “audio front end” and “search stage” implemented using a “first processor” and a “calculating circuit” implemented using a “second processor[.]” In related art, *Baumgartner* teaches a similar front end and search stage as *Jiang*’s provided on a first processor and a similar calculating circuit as *Jiang*’s provided on a second processor. For the reasons explained below, a POSITA would have been motivated to modify *Jiang* to include the hardware architecture taught by *Baumgartner*. *Dec.*, 172.

a) Overview of Mapping

Comparing the ’277 *Patent*’s Fig. 21 with *Baumgartner*’s Fig. 19 illustrates how *Baumgartner* teaches the claimed hardware architecture of Claim 1(e). The ’277 *Patent*’s Figure 21 describes “a speech recognition circuit having a shared processor for front end and search stages[.]” ’277 *Patent*, 8:54-57. Figure 21 of the ’277 *Patent* and Figure 19 of *Baumgartner* are juxtaposed below, illustrating the “audio front end” and “search stage” implemented on a CPU (comprising a “first processor”) and a “calculating circuit” implemented on an accelerator (comprising a “second processor”).



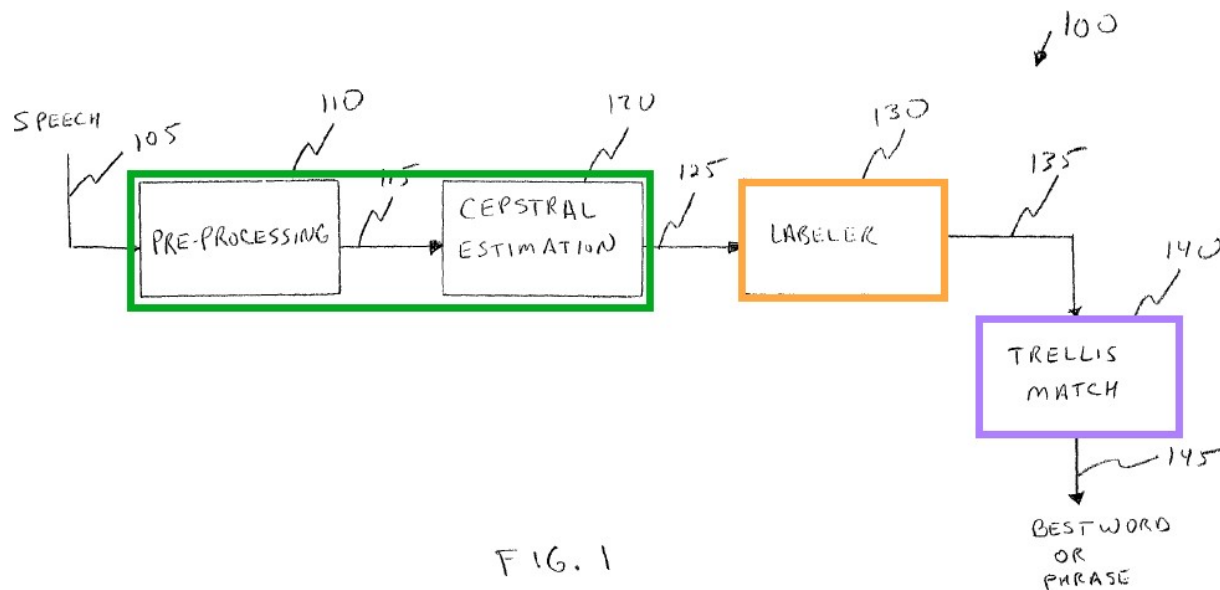
'277 Patent, Fig. 21 (annotated, excerpted, left); Baumgartner, Fig. 19 (annotated, excerpted, right).

Baumgartner's Fig. 19 teaches a speech recognition hardware architecture similar to the '277 Patent. Baumgartner, [0031]-[0033], [0156]-[0158], Fig. 19. Baumgartner's Fig. 19 illustrates a cepstrum computation 1905 and atom value computation 1910, serving as an audio front end, and at least step 1935 serving as a search stage. Each of the cepstrum computation 1905, atom value computation 1910 and final selection and rank mapping 1935 are implemented on processor core 1910 (a "first processor"). Baumgartner's Fig. 19 also discloses density computations

1960-1980, which are distance calculations, implemented on Speech Label Accelerator (SLA) 520 (a “second processor”). Dec., 173-176.

b) **Baumgartner Teaches an “audio front end,” a “calculating circuit,” and a “search stage”**

Baumgartner’s “invention relates to speech recognition and, more particularly, to Speech Label Accelerators (SLAs) and techniques for using such SLAs.” *Baumgartner*, [0002]. *Baumgartner* teaches well-known speech recognition functionality, including processing an acoustic signal to a digital signal (110), sampling the digital signal to extract feature (observation) vectors (120), calculating distances indicating the similarity of the feature vector to a stored acoustic model (130), and identifying the most likely utterance (e.g., phoneme or word) via a trellis search (140):



Baumgartner, Fig. 1 (annotated), [0033]-[0043] (with pinpoint, [0037]-[0038]);
Dec., 177-178.

Baumgartner’s “audio front end”: *Baumgartner*’s Fig. 1 teaches an “audio front end,” namely pre-processing 110 and cepstral estimation 120, that convert the speech signal into frames and apply signal-processing operations, such as cepstral estimation, to obtain an observation vector. *Baumgartner*, [0037]-[0038], Fig. 1; *Dec.*, 179 (citing ’277 Patent, 11:65–12:10; Jiang, 6:46–7:1; Ravishankar, 27-28, 40-41; *Dictionary of Electrical and Computer Engineering*, 5; and opining *Baumgartner*’s signal processing and cepstral encoding are similar to *Jiang*’s teachings mapped as the “audio front end”).

Baumgartner’s “calculating circuit”: *Baumgartner*’s Fig. 1 illustrates a labeler 130 performed by SLA 520, where the labeler 130 determines probabilities for phones. *Baumgartner*, [0038]; *Dec.*, 180. The SLA determines “likelihoods” of a “[g]iven a single observation vector, v, associated with a single frame of speech[.]” *Baumgartner*, [0043]. *Baumgartner* teaches “the heart of the SLA is a circuit that permits the fast approximate computation of a large collection of values [] for a fixed observation vector v.” *Baumgartner*, [0043]; *Dec.*, 180 (citing ’277 Patent, 3:51-56, Fig. 16, and opining *Baumgartner*’s labeling block 130 is similar to *Jiang*’s tree search engine 74 determining a probability score for each considered phoneme as compared to the acoustic states of an HMM and mapped as the “calculating circuit”).

Baumgartner’s “search stage”: *Baumgartner’s* Fig. 1 discloses **trellis match 140**, that “performs a search and selects a best word or phrase[.]” *Baumgartner*, [0038]; *Dec.*, 181 (citing ’277 Patent, 13:33–15:57; *Jiang*, 1:33-40, and opining *Baumgartner’s* trellis search is similar to *Jiang’s* tree search engine identifying a word and mapped as the “search stage”).

Baumgartner teaches the well-known speech recognition functionality described for Fig. 1 “may be designed as an apparatus” including a “wide variety of hardware designs that are possible for the same basic idea” and further describes “various extensions and refinements” of such hardware designs. *Baumgartner*, [0033]. Figure 19 represents one such design. *Dec.*, 182.

**c) *Baumgartner* Teaches an Audio Front End
“implemented using a first processor”**

Baumgartner’s **cepstrum computation 1905 and atom value computation 1910** on **processor core 1910** teach an audio front end “implemented using a first processor[.]” *Baumgartner*, [0037]-[0038], [0156]-[0157], Fig. 19; *Dec.*, 183-184 (explaining how *Baumgartner’s* “cepstrum” computation 1905 is akin to *Baumgartner’s* “cepstral” estimation 120, citing *Ravishankar*, 40-41, Fig. 3.1, and explaining how *Baumgartner’s* “atom value computation” is a front-end process). Per Mr. Schmandt, *Baumgartner’s* “**processor core 1910**” comprises a processor (as recited in Claim 1(e)) performing certain functionality related to signal processing and vector extraction, as recited in Claims 1(a)-1(b). *Dec.*, 185 (discussing the ’277

Patent, 30:40–33:67, Fig. 21 describes an exemplary CPU performing the audio front end functions and *Baumgartner*, [0156] describes the processor core as executing a sequence of steps).

d) ***Baumgartner* Teaches a Calculating Circuit
“implemented using a second processor”**

As discussed in Section V.A.4.c, *Baumgartner* teaches a calculating circuit, the mapping for which is incorporated by reference for Claim 1(e). Per *Baumgartner*, the “heart” of the SLA is a **circuit** for performing the probability determinations. *Baumgartner*, [0043]. *Baumgartner* teaches the SLA performs calculations later used in *Baumgartner*’s trellis match on *Baumgartner*’s processor core 1910 to “perform[] a search and select[] a best word or phrase 145.” *Baumgartner*, [0038].

Per Mr. Schmandt, a POSITA would have understood the SLA comprises a “processor.” *Dec.*, 186-188. The SLA comprises memory and adder circuitry and optionally an accumulator, mux, and a control unit. *Baumgartner*, [0008]. The SLA is described as allowing “fast and efficient **computation**” and that the heart of the SLA is a circuit, further indicating the SLA comprises a processor. *Baumgartner*, [0008], [0043]; *Dec.*, 186-188. The functionality performed by the SLA, including computationally heavy mathematical processing, would have been understood to be performed by a processor. *Dec.*, 186-188 (*citing Baumgartner*, [0003], [0005], [0007], [0033]). Finally, *Baumgartner* teaches speeding processing during the labeling stage of speech recognition and to free “one or more processors from

performing labeling.” *Baumgartner*, [0007]. Therefore, the SLA including the circuit for determining probabilities is or comprises a processor, as claimed. *Dec.*, 186-188.

Because *Baumgartner* teaches the SLA includes as its “heart” the circuit that permits computation of a large collection of values, P for a fixed observation vector v , and further because the SLA comprises a processor, *Baumgartner* teaches the calculating circuit is “implemented using a second processor,” as claimed. *Dec.*, 186-188.

e) *Baumgartner* Teaches a Search Stage “implemented using a first processor”

Baumgartner’s step 1935 performed by processor core 1910 teaches the search stage “implemented using a first processor[.]” *Baumgartner*, [0038], [0155]–[0157], Figs. 1, 19.

As explained above, *Baumgartner*’s processor core 1910 teaches the claimed “first processor,” and *Baumgartner* further teaches the functionality of an “audio front end” performed by the processor core. *Baumgartner*’s Fig. 19 also teaches processor core 1910 implements a search functionality similar to *Jiang*’s mapped “search stage.” Specifically, *Baumgartner*’s step 1935 takes the output computation from *Baumgartner*’s SLA to perform a search and select the best word or phrase. *Baumgartner*, [0038] (referring to a “trellis match”), [0156]–[0158], [0206] (“The final summation values written back to system memory by the SLA core are used by

software to determine the best match or to define the next set of indices to be applied to further refine the search for the best match.”), [0210], Figs. 1, 19; *Dec.*, 189-191.

Because *Baumgartner*’s processor core 1910 performs a search through a trellis (step 1935, [0038]), and because the processor core comprises the claimed “first processor,” *Baumgartner* teaches or renders obvious a search stage “implemented using a first processor,” as claimed. *Dec.*, 188-191.

f) Motivation to Combine *Jiang* and *Baumgartner*

A POSITA would have found it obvious and been motivated to modify *Jiang*, such that *Jiang*’s mapped audio front end and search stage were implemented using a first processor and *Jiang*’s mapped calculating circuit was implemented using a second processor, as taught by *Baumgartner*. *Dec.*, 193-203.

Jiang itself teaches its speech recognition system may be “practiced with other computer system configurations, including ... multi-processor systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, main-frame computers, and the like.” *Jiang*, 4:60-65, 6:32-36, 6:39-42 (further describing exemplary hardware architectures). Thus, a POSITA would have understood *Jiang* contemplated modifications to its hardware architecture, including using multiple processors. *Dec.*, 192-195 (*citing Jiang*, 10:65–11:25 and discussing *Jiang*’s desirable “increase in computational savings” afforded by threshold pruning).

A POSITA would have found it obvious and been motivated to modify *Jiang* such that distance calculation was performed on a processor distinct from the processor performing both feature extraction (already taught in *Jiang*, 6:53-57) and word identification (i.e., the processor for the tree search engine). *Dec.*, 192-195. *Jiang* refers to distinct processors performing distinct speech recognition functionality repeatedly, and the feature vector extraction and search stage are already indicated as being performed by dedicated processors, indicating to a POSITA that *Jiang* envisions dedicated processors for performing specific speech recognition functionality. *Dec.*, 192-195.

Given a POSITA would have understood *Jiang*'s disclosures contemplate various hardware architectures, a POSITA designing a speech recognition system at the time of the '277 Patent would have looked to *Baumgartner*'s explanations that "[s]peech recognition ... is currently a computationally intensive task[,] "it would be beneficial to reduce the load on the processors of the system when the system performs speech recognition[,] " and "a need exists for techniques to reduce the computational load of a processor during speech recognition." *Baumgartner*, [0003]-[0006]; *Dec.*, 196-201. A POSITA would have recognized *Baumgartner*'s SLAs "speeds processing during the labeling stage of a speech recognizer[,] and frees one or more processors from performing labeling." *Dec.*, 201 (*quoting Baumgartner*, [0007]). A POSITA would have further understood *Baumgartner*'s hardware

architecture—utilizing a separate processor for computationally-expensive distance calculations—provides an efficient solution to the computational load issues identified by the prior art. *Dec.*, 201; *Baumgartner*, [0200]–[0205].

It was well-known that calculating distances of feature vectors’ similarities to pre-determined acoustic states was a computationally-intensive task (as taught by *Baumgartner*). *Dec.*, 198-201 (*citing* numerous background references). Accordingly, a POSITA would have been motivated to utilize a separate processor, such as *Baumgartner’s* SLA, for the distance calculation phase of *Jiang’s* speech recognition system to increase overall processing speeds. *Dec.*, 197-201.

A POSITA would have had a reasonable expectation of success implementing *Jiang’s* distance calculation on a separate processor without undue experimentation at least because *Baumgartner* and the art cited by Mr. Schmandt expressly illustrated its feasibility. *Dec.*, 202. Additionally, using separate processors, such as *Baumgartner’s* SLA, or co-processors, was a well-known technique to offload computation tasks prior to the ’277 Patent. *Dec.*, 202. The determination of whether to use a separate processor for computationally-expensive tasks—as opposed to using a sole processor for the entire system—would have been a simple design choice within the skill of a POSITA. *Dec.*, 202. Given the ubiquity of separate processors at the time, a POSITA would have been able to implement a separate

processor into a speech recognition system without undue experimentation. *Dec.*, 202.

7. *Claim 1(f)*

The *Jiang-Baumgartner* system teaches, or at least renders obvious, transmitting data from a front end to a calculating circuit to a search stage. *See* Claims 1(a)-1(d); *Dec.*, 203. The *Jiang-Baumgartner* system does not expressly teach pipelining. However, as discussed in detail below, *Jiang-Baumgartner* in view of *Brown* teaches Claim 1(f). *Dec.*, 203-205 (discussing the '277 Patent's description of pipelining).

a) *Brown's Teachings*

Brown's speech recognition system 100 includes functional blocks performing feature extraction, phone probability processing, and word probability processing. *Brown*, 3:32-41. *Brown* teaches processing stages may be pipelined:

The functionality of each block of the recognition system may be performed by a respective different processor, or the functionality of several or all blocks may be performed by the same processor. ... The **stages are pipelined and their operations are synchronized with the rate at which frames of the incoming speech signal are generated.**

Brown, 3:41-48. A POSITA would have understood *Brown's* teachings of pipelining that “synchronize[s] with the rate at which frames of the incoming speech signal are generated” teaches or at least renders obvious the claimed data “is pipelined...” as

claimed. *Dec.*, 206-209. Mr. Schmandt opines that *Brown's* teachings inform a POSITA that as the frames of data are generated, the data is processed from one stage to another so that the operations are synchronized with the rate of frame generation. *Id.* Because *Jiang*, 6:62–7:1 teaches speech is input in 10ms frames and a feature vector for each 10ms frame is calculated (extracted), then further processing of each feature vector in a time-synchronized fashion (as expressly taught by *Brown*) is needed to process the speech efficiently. This processing method parallels the '277 Patent's description, where a time axis is divided into equal pipelined steps, such that the various processors performing each function are acting in parallel. '277 *Patent*, 26:60, 27:3-8; *Dec.*, 206-209.

b) Motivation to Combine *Brown* with *Jiang-Baumgartner*

A POSITA would have been motivated and found it obvious to modify the *Jiang-Baumgartner* system to pipeline data according to *Brown's* teachings. *Dec.*, 210-214. Pipelining data offered numerous well-known advantages including greater throughput of the processor, improved performance, decreased cost, and simplified programming. *Dec.*, 210. Utilizing the well-known pipelining technique would have improved the *Jiang-Baumgartner* system in the same way as *Brown's* speech recognition system. *Dec.*, 210.

A POSITA would have recognized that *Jiang* teaches the three stages of feature extraction, distance (probability) calculation, and word search, as explained

above. *Dec.*, 211-212. A POSITA would have also understood *Brown* teaches pipelining the speech recognition stages where the output from a prior stage is used as the input to the next stage. *Dec.*, 212. Because speech data is received in exemplary 10ms increments, a POSITA would have desired a steady data input and recognized a need to process data efficiently. *Id.* Per Mr. Schmandt, it would have been obvious to pipeline data among the three well-known speech recognition stages occurring on different processors (as in *Jiang-Baumgartner*) to keep processing efficient with respect to the exemplary 10ms time intervals. *Id.* As explained, *Jiang* already discusses the importance of time with respect to pruning. *Jiang*, 10:65–11:25. As also explained, *Jiang-Baumgartner* accounts for increased processing when utilizing pruning, thus achieving both higher accuracy and more efficient processing. A POSITA would have realized pipelining keeps all processors operating efficiently in a multi-processor system, such that wasteful idle-processing time is minimized. *Dec.*, 212-213. A POSITA would have recognized *Brown*'s pipelining was one such way to avoid wasted computational resources as *Brown* expressly teaches synchronization “with the rate at which frames of the incoming speech signal are generated.” *Id.* (quoting *Brown*, 3:44-48).

A POSITA would have also had a reasonable expectation of success in modifying the *Jiang-Baumgartner* system to implement *Brown*'s pipelining teachings. *Dec.*, 214. Pipelining was well-known in the art including in speech

recognition systems. *Dec.*, 214. A POSITA would have understood that *Jiang*, *Baumgartner*, and *Brown* teach speech recognition systems using similar hardware, such as various processors to perform similar speech recognition functions, such that modifying the *Jiang-Baumgartner* system to utilize pipelining would have been a trivial modification. *Id.*; *Jiang*, 4:60-65, 6:32-36, 6:39-42; *Baumgartner*, [0033], Fig. 19; *Brown*, 3:11-15. Additionally, a POSITA would have understood and found obvious that implementing pipelining would have “simplif[ied] programming,” as described in background art such as *Simar*, further providing a reasonable expectation of success when modifying the *Jiang-Baumgartner* system to utilize *Brown*’s pipelining. *Dec.*, 214; *Simar* (Ex. 1039), 20:22-23.

B. Claim 5

Jiang teaches “[a] probability score is computed for **each state** in the state sequences.” *Jiang*, 2:5-6, 1:21-25; *Dec.*, 215-217. As mapped for Claim 1(c), *Jiang* or alternatively *Baumgartner* teaches a calculating circuit calculating distances.

Additionally, *Jiang* teaches the calculating circuit is configured to *autonomously* calculate distances because tree search engine 74 (modified to have the search stage on a first processor and the distance calculation on a second processor, per Claim 1(e)) begins calculating the probability scores “[u]pon receiving the codewords from feature extraction module 66” and further by *accessing* the HMM and *determining* a most likely phoneme based on the probability

score for each state in the state sequence. *Dec.*, 215-217; *Jiang*, 7:30-42. Therefore, tree search engine *autonomously* performs the distance calculation for *every state* upon receiving the codewords. *Dec.*, 215-217.

Because *Jiang* teaches calculating the distance for **each state** of the acoustic model (HMM), and because the tree search engine calculates the distances (probabilities) upon receiving the codewords, the *Jiang-Baumgartner-Brown* system teaches, or at least renders obvious, Claim 5.

C. Claim 7

See Claim 1(b). *Jiang* teaches feature extraction module encodes feature vectors “reflecting the spectral characteristics[.]” *Jiang*, 6:62–7:1; *Dec.*, 218-220. As mapped for Claim 1(a), the feature vectors are encoded from the audio signal. The frame, and accordingly, the feature vector for the frame, is for a “predetermined time frame[.]” such as 10ms. *Jiang*, 6:64-65, 7:7-10; *Dec.*, 218-220.

D. Claim 12

Claim 1(a) maps *Jiang*’s feature extraction module 66 as the “audio front end.” *Jiang* further teaches the “speech is input into system 60” and converted to a digital signal, “which is provided to feature extraction module.” *Jiang*, 6:46-53; *Dec.*, 221-222.

E. Claim 14

1. Claim 14(Pre)

See Claim 1(Pre).

2. Claim 14(a)

See Claim 1(a).

3. Claim 14(b)

See Claim 1(b).

4. Claim 14(c)

See Claim 1(c), which maps the claimed function of calculating a distance. Claim 14(c) recites a limitation in means-plus-function format (*see* Section III.C.1). The construed “calculating means” includes at least a processor for performing the claimed function. *Jiang* discloses the distance calculation is performed by tree search engine, which may be performed by a “dedicated” processor. *Jiang*, 6:39-42, 8:16-24. *Jiang* also discloses tree search engine is implemented in CPU, and CPU is a processor. *Id.* Therefore, *Jiang*’s tree search engine being implemented by a dedicated processor and otherwise performing the claimed functionality (as mapped for Claim 1(c)) teaches or renders obvious Claim 14(c). *Dec.*, 226-227.

Additionally, *Baumgartner* teaches a processor for performing distance calculation, namely the SLA including a circuit, as also mapped for Claim 1(c) at Section V.A.4.c and Claim 1(e) at V.A.6.d (discussing SLA comprises a processor). *Dec.*, 228.

5. Claim 14(d)

See Claim 1(d).

6. Claim 14(e)

See Claim 1(f) regarding “pipelined data flow.” Additionally, *Baumgartner* teaches the components of Figure 19 are “all interconnected through bus 1948[,]” indicating the claimed components “are connected to each other,” as claimed. *Baumgartner*, [0156]; *Dec.*, 230-231.

F. Claim 15

1. Claim 15(Pre)

See Claim 1(Pre). *Jiang* discloses a *method* of recognizing speech. *Jiang*, Claim 1. *Dec.*, 232.

2. Claim 15(a)

See Claim 1(a); *Dec.*, 233.

3. Claim 15(b)

See Claim 1(b).

4. Claim 15(c)

See Claim 1(c).

5. Claim 15(d)

See Claim 1(d).

6. Claim 15(e)

See Claim 1(f).

G. Claim 16

1. Claim 16(Pre)

See Claims 1(Pre) (discussing speech recognition), 1(e) (discussing processors), and 15(Pre) (discussing speech recognition *method*).

The *Jiang* invention may be implemented “in the general context of computer-executable instructions,” and “[t]he drives and the associated computer-readable media provide nonvolatile storage of computer readable instructions[.]” *Jiang*, 4:54-56, 5:5-29; *Dec.*, 238-239 (explaining *Jiang*’s system memory 22, ROM 24, and RAM 25 represent examples of “non-transitory storage medium”). Further, a POSITA would have understood that speech recognition methods utilized storage medium and computer code to implement functions on processors, such as CPU and tree search engine serving as a dedicated processor. *Dec.*, 238-239 (*citing Ravishankar*, 5, 18-20, 49, 56, 79, Figs. 3.5, 4.6); *Jiang*, 6:39-42.

Accordingly, *Jiang*’s system memory 22 (or ROM 24 or RAM 25) teach the claimed “non-transitory storage medium[.]” and *Jiang*’s computer-executable instructions stored thereon teach Claim 16(Pre), to the extent such is limiting. *Dec.*, 238-239.

2. Claim 16(a)

See Claims 16(Pre), 1(a), 1(e). Because *Jiang* teaches computer-executable instructions for performing the disclosed functionality, a POSITA would have

understood the *Jiang* speech recognition system includes “code for” performing the claimed functionality of claims 16(a)-16(d). *Dec.*, 240-241.

3. Claim 16(b)

See Claims 16(Pre) (“code for”), 1(b), 1(e) (“processor”).

4. Claim 16(c)

See Claims 16(Pre), 1(c), 1(e).

5. Claim 16(d)

See Claims 16(Pre), 1(d)-1(e).

6. Claim 16(e)

See Claims 16(Pre), 1(e)-1(f). In the modified *Jiang*, the data is pipelined “pursuant to the code from the feature calculation,” i.e., pursuant to the calculated feature vector.

VI. GROUND 2: CLAIM 4 IS OBVIOUS OVER *JIANG*, *BAUMGARTNER*, *BROWN*, AND *KAZEROONIAN*

See Ground 1, Claim 1. Claim 1 does not recite “front ends” (plural) but only recites a “front end” (singular). For this Petition only, Petitioners map the “front ends” in Claim 4 as the claimed “front end” in Claim 1.

Kazeroonian teaches an automated real-time system for processing audio-video data including speech recognition functionality, including expressly teaching using multi-threaded operations for pipelined processing. *Kazeroonian* (Ex. 1037), Abstract, 15:33-16:1, 19:30–20:5, 21:9-11, Fig. 2.

Kazeroonian teaches an audio-video processing server (AVPS) 101 performing speech recognition. *Kazeroonian*, 15:33–16:1; *Dec.*, 246-249.

Kazeroonian teaches:

[AVPS] 101 can be a multi-processor allowing parallel processing of the modules and tasks shown in Figs. 2 and 3. The tasks and processing steps are divided between separate processing threads[.] ... AVPS 101 uses threads to achieve real-time video encoding throughput, high segmentation speed, and other characteristics that affect user interface responsiveness. **The pipelined architecture** supported by these multiple threads provides the end-users with access to the indexed videos only moments after the videos are received. [¶] Even on a **single processor system multi-threading can improve performance** by judicious assignment of various tasks (stages) to **separate threads**.

Kazeroonian, 19:26–20:5. Additionally, *Kazeroonian* teaches “different processor threads are assigned to different stages of processing.” *Kazeroonian*, 21:9-11.

Per Mr. Schmandt, *Kazeroonian* thus teaches a single processor performing multi-threaded operations with separate threads. *Dec.*, 246-249. *Kazeroonian*’s teachings related to AVPS 101 are also applicable to systems processing only audio data (as opposed to audio-video data), such as speech recognition. *Dec.*, 249.

A POSITA would have been motivated and found it obvious to modify the *Jiang-Baumgartner-Brown* system to utilize a multi-threaded operation for the claimed “first processor” to run the claimed “search stage” and “front end” as

separate threads to avoid data loss. *Dec.*, 250-251; *see* Claim 1. Per Mr. Schmandt, *Kazeroonian* expressly motivates the combination per its teachings that multiple threads achieve real-time throughput and high segmentation speed. *Dec.*, 250-251; *Kazeroonian*, 19:30-33. Yet further, such a modification would have advantageously “improve[d] performance by judicious assignment of various tasks (stages) to separate threads,” as expressly taught by *Kazeroonian*. *Kazeroonian*, 20:3-5; *Dec.*, 250-251. Indeed, *Kazeroonian* teaches “[u]sing threads is especially advantageous when a process consists of multiple tasks of varying urgency and processing needs.” *Kazeroonian*, 20:10-12. Such a use of a known multi-threading technique would have advantageously improved the modified *Jiang-Baumgartner-Brown* system in the same way as *Kazeroonian*’s system. *Dec.*, 251.

Furthermore, the modification would have been performed with a reasonable expectation of success. *Dec.*, 252. For example, multi-threading operations were well known in the art, within the skillset of a POSITA to implement, and were commonly used with pipelining operations, as taught by *Kazeroonian*. *Kazeroonian*, 19:33–20:2; *Dec.*, 252.

VII. GROUND 3: CLAIMS 9-10 ARE OBVIOUS OVER JIANG, BAUMGARTNER, BROWN, AND VENSKO

A. Claim 9

There is no antecedent basis for the claimed “speech accelerator.” For this Petition only, Petitioners maps “the speech accelerator” as the claimed “calculating

circuit” of Claim 1(c) implemented as a second processor, as recited in Claim 1(e). The ’277 Patent describes the “distance calculation engine may be included within an accelerator.” *’277 Patent*, 3:34-37.

Vensko teaches template processors that perform a dynamic programming algorithm (DPA) including performing a distance calculation between a character distance vector (i.e., feature vector) and an assigned template. *Vensko* (Ex. 1038), 6:22-29. A master processor causes character distance vectors to be transferred to each of the template processors for comparison. *Vensko*, 6:22-26. “The master processor via an interrupt commands the template processors to begin processing the character distance vectors on a frame basis.” *Vensko*, 6:26-29. Once the processing is complete, each template processor **“generate[s] an interrupt** to the master processor 160 informing it that the results are ready for transfer[,]” thus informing the master processor that each template processor is ready to receive and process the next character distance vector. *Vensko*, 6:29-33, 10:10-19; *Dec.*, 253-259. *Vensko* teaches after the interrupt signal—which informs the master processor results are ready—and a syntax analysis, “a new character distance vector is sent to the template processors and the process begins all over again until recognition is accomplished.” *Vensko*, 6:48-51. *Vensko* further teaches “[t]o continue, an interrupt of the master processor by a slave processor was described above. The interrupt is used to tell the master that the slave processor which is interrupting [has] finished a processing

task.” *Vensko*, 10:10-13. *Vensko*’s master processor then “decides on what data transfers are to take place” and sets up a data transfer based on the interrupt. *Vensko*, 10:13-19.

Per Mr. Schmandt, *Vensko* teaches or renders obvious utilizing interrupt signals amongst various processors of a speech recognition system. *Dec.*, 253-259. For example, *Vensko* teaches the master processor sending an interrupt to the template processor to command processing of the character distance vectors, and similarly, the template processor sends an interrupt to the master processor to indicate results are complete so that the master processor can transfer the results. *Dec.*, 253-259. Per Mr. Schmandt, *Vensko*’s teaching of interrupts to command speech recognition processing teaches or renders obvious Claim 9 because a POSITA would have understood *Vensko* suggests utilizing interrupt commands to both instruct processing and instruct transfer of data once processing is complete. *Dec.*, 253-259.

Petitioners additionally note Claim 9 is an apparatus claim (a circuit having an interrupt signal) performing a function (informing the front end ...). Per MPEP 2114, apparatus claims cover what a device is, not what a device does. MPEP 2114(II). Additionally, “[f]unctional claim language that is not limited to a specific structure covers all devices that are capable of performing the recited function.” MPEP 2114(IV). *Vensko* teaches a speech recognition system including a front end

for performing feature vector calculation and a calculating circuit (i.e., template processors) for calculating a distance. *Vensko* teaches processors sending and receiving interrupt signals to perform various functions. Therefore, *Vensko*'s disclosed system teaches the structure recited in Claim 9, including structure *capable of performing* the recited function. *Dec.*, 260. Thus, *Vensko* teaches or renders obvious Claim 9.

A POSITA would have been motivated and found it obvious to modify the *Jiang-Baumgartner-Brown* system to include *Vensko*'s interrupt functionality. *Dec.*, 261-262. A POSITA would have understood that pipelined systems process system stages concurrently—for example the claimed “audio front end,” “calculating circuit,” and “search stage.” *Dec.*, 261-262 (*citing Brown*, 3:32-48). Therefore, modifying the *Jiang-Baumgartner-Brown* system such that the calculating circuit implemented as a second processor performing distance calculation generates an interrupt to signal that the distance calculation processor is ready to receive a next feature vector for processing is a routine and obvious modification for obtaining efficient processing. *Dec.*, 261-262. A POSITA would have had a reasonable expectation of success without undue experimentation in making the modification at least because interrupt signals were well-known at the time and easy to implement. *Dec.*, 263.

B. Claim 10

See Claim 9. As mapped for Claim 1(f), data is pipelined from the calculating circuit (which computes distances) to the search stage, which uses the calculated distances. Because *Vensko*'s interrupt functionality on the template processor signals that results are ready for transfer in a pipelined system, in the modified *Jiang-Baumgartner-Brown-Vensko* system, it would have been obvious for the accelerator (i.e., the calculating circuit implemented as a second processor) to signal the search stage (i.e., the component that next uses the calculated distances) when distances for a new frame are available. *Dec.*, 264. Further, *Jiang* teaches tree search engine 74 accesses information stored in memory 72. *Jiang*, 7:31-34.

Petitioners further incorporate by reference its discussion of MPEP 2114 in the Claim 9 mapping. Because *Vensko* teaches a system for sending/receiving interrupt signals, *Vensko* teaches a system *capable of* performing the recited functionality of Claim 10.

**VIII. GROUND 4: CLAIMS 1, 5, 7, 12, AND 14-16 ARE OBVIOUS OVER
*JIANG, BAUMGARTNER, BROWN, AND SMYTH***

The *Jiang-Baumgartner-Brown* system teaches Claims 1, 5, 7, 12, and 14-16. Further, *Smyth* teaches a “calculating circuit for calculating distances indicating the similarity between a feature vector and a plurality of predetermined acoustic states of an acoustic model,” per Claim 1(c).

To the extent Patent Owner argues *Jiang*'s teachings explained above in Section V.A.4.b fail to teach the claimed “calculating circuit *for calculating distances ...*” functionality, it would have been obvious to a POSITA to configure the *Jiang-Baumgartner-Brown* system to perform such claimed functionality, as taught by *Smyth*.

A. Claim 1

1. Claim 1(Pre)

See Section V.A.1.

2. Claim 1(a)

See Section V.A.2.

3. Claim 1(b)

See Section V.A.3.

4. Claim 1(c)

Jiang, alone or in view of *Baumgartner*, teaches, or at least renders obvious, this limitation. *See* Section V.A.4. To the extent Patent Owner contends that neither *Jiang* nor *Baumgartner* teaches the claimed calculating circuit and recited functionality, *Smyth* teaches such.

a) “a calculating circuit”

Smyth teaches “a calculating circuit for calculating distances[,]” via classifier (or classification or classifying) processor 341, which calculates state probabilities

indicating the similarity that a feature vector corresponds to each acoustic state in state memory 342. *Smyth*, 5:55–6:4; *Dec.*, 268-279.

A POSITA would have understood *Smyth*'s classifier processor 341 teaches a “calculating circuit” or that such is obvious. *Dec.*, 270-273. The '277 Patent does not ascribe any particular structural features or requirements to the claimed “circuit.” *Id.* Additionally, a POSITA would have understood a “circuit” is met by a processor. *Id.*; *see also* Section V.A.1. Therefore, *Smyth*'s classifying processor performing the functionality recited in Claim 1(c) teaches or renders obvious the claimed “calculating circuit[.]”

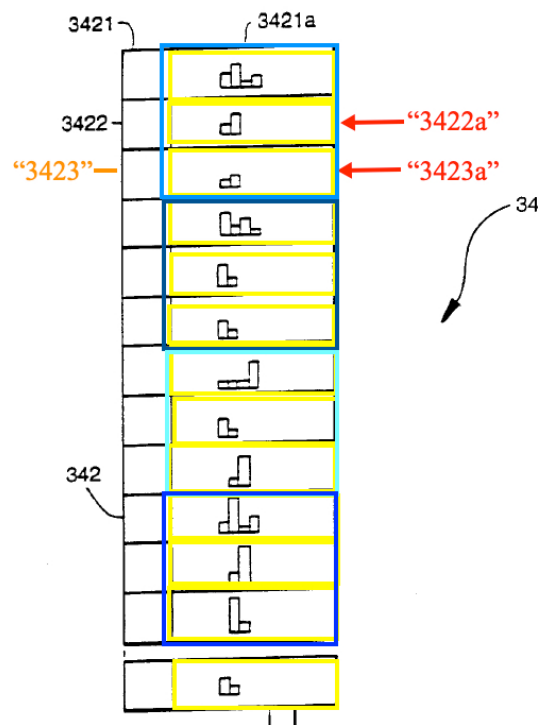
b) “a plurality of predetermined acoustic states”

Smyth's speech states stored in state fields of state memory teaches “a plurality of predetermined acoustic states[.]”.

Smyth teaches classifier 34 comprises classifier processor 341 and state memory 342. *Smyth*, 5:55-57, Fig. 3. “The state memory 342 comprises a state field 3421, 3422, ...” *Smyth*, 5:58-59. Each state field stores a speech state. *Smyth*, 5:58-59 (“a state field 3421, 3422 ..., **for each of the plurality of speech states.**”). *Smyth*'s original Fig. 3 identified by reference numeral only two fields (3421, 3422) and one speech state (3421a). However, a POSITA would have understood that the cells following 3421, 3422 are similarly state fields, and that each state field stores a speech state similar to speech state 3421a. *Dec.*, 274-276 (*citing Smyth*, 5:58-59

and noting *Smyth*'s use of ellipses that would also indicate similar state fields with respective speech states). Fig. 3 below depicts state memory 342, state fields 3421–3422, speech state 3421a, and groups of three speech states corresponding to an allophone (in shades of blue):

Fig. 3



Smyth, Fig. 3 (annotated). *Smyth* does not identify element 3421a in its written description (although 3421a is identified in Fig. 3). However, *Smyth* teaches a “plurality of speech states” with each state field having a speech state. *Smyth*, 5:58-62. Consequently, a POSITA would have understood that element 3421a was a “speech state” of state field 3421, and added elements 3422a, 3423a were

correspondingly “speech states” of respective state field 3422. *Dec.*, 274-277; *Smyth*, 5:58-62. Therefore, *Smyth* teaches a plurality of states.

Smyth expressly teaches the speech states are “acoustic” states: “each allophone to be recognised by the recognition processor comprises 3 states, and accordingly 3 state fields are provided in the state memory 342 for each allophone.” *Smyth*, 5:60-62. An allophone is well-known as an “acoustic” subword unit providing alternative pronunciations of phonemes. *Dec.*, 277-278.

The speech states are *predetermined* because the states exist prior to the distance calculations via training the speech recognition system. *Smyth* expressly teaches storing a template or model of a word or phrase containing a reference feature matrix. *Smyth*, 1:34-39; *Dec.*, 279 (*citing Ravishankar*, 19-20, 27-28, 32). *Smyth* also expressly teaches *storing* the speech states in state memory. *Smyth*, 5:58-67. Because *Smyth* teaches generating a model of spoken speech, the model contains speech states (*see* Section VIII.A.4.c), *Smyth* teaches storing speech states, and training a model with speech states was well-known, *Smyth* teaches the states are “predetermined[.]” *Dec.*, 279.

c) “[plurality of predetermined acoustic states] of an acoustic model”

Smyth teaches an “acoustic model” via its “template or model,” including a “plurality of model states,” used to determine a measure of similarity and recognition results for spoken speech. *Smyth*, 1:34-39, 5:7-9. *Smyth*’s discloses:

A **stored representation of the word or phrase**, known as a template or **model**, comprises a **reference feature matrix** of that word as previously derived from, in the case of speaker independent recognition, multiple speakers. The **input feature vector is matched with the model** and a **measure of similarity** between the two is produced.

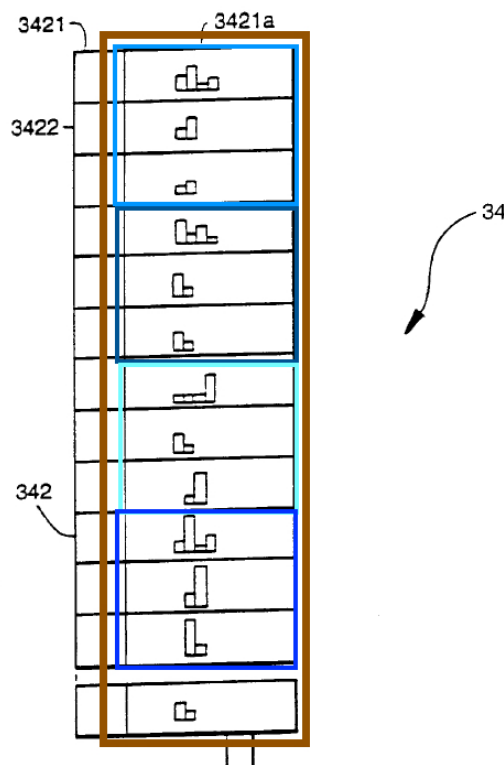
Smyth, 1:34-39. *Smyth* further teaches “classifier 34 receiv[es] the succession of feature vectors and operat[es] on each with a **plurality of model states**, to generate recognitions results[.]” *Smyth*, 5:6-9.

Speech recognition systems were well-known to generate models or templates of spoken input, as taught by *Smyth*, 1:34-39. *Dec.*, 280-281 (*citing* Ney, p. 4 (teaching estimating probabilities of acoustic vectors “during the training phase of the recognition system”), *Jiang*, 11:25-29, *Ravishankar*, 19-20, 27-28 (discussing training acoustic models in subword units, such as phonemes)). Per Mr. Schmandt, *Smyth*’s “plurality of model states” is an *acoustic model*, where reference to the states of the model (i.e., *Smyth*’s speech states, 5:58-62) are used to determine the likelihood that a feature vector corresponds to one of the speech states. *Dec.*, 281-282. For example, *Smyth* teaches calculating probabilities for each state in state memory 342, where each probability indicates this likelihood (further discussed below for Claim 1(c)). *Smyth*, 5:63–6:4. Additionally, the recognition system “operat[es] on each [feature vector] with a plurality of model states,” further

indicating the plurality of model states are an acoustic model used to determine probabilities. *Smyth*, 5:7-9. Thus, by referencing the states, (the “plurality of model states”) feature vector probabilities are calculated to identify allophones. *Id.*

Smyth further teaches words are determined by stringing together identified allophones. *Smyth*, 6:15-19; *Dec.*, 283-284. The stored representation of the word or phrase, i.e., the “template or model” per *Smyth*, 1:34-39, is the “string of allophones” that, when concatenated, form the word. *Smyth*, 6:15-19; *Dec.*, 283-284. Annotated Fig. 3 below depicts *Smyth*’s allophones (annotated shades of blue) as a collection of acoustic states (annotated brown).

Fig. 3



Smyth, Fig. 3 (excerpt, annotated). This collection of speech states (where three speech states are provided for each allophone) teaches or renders obvious an “acoustic model.”

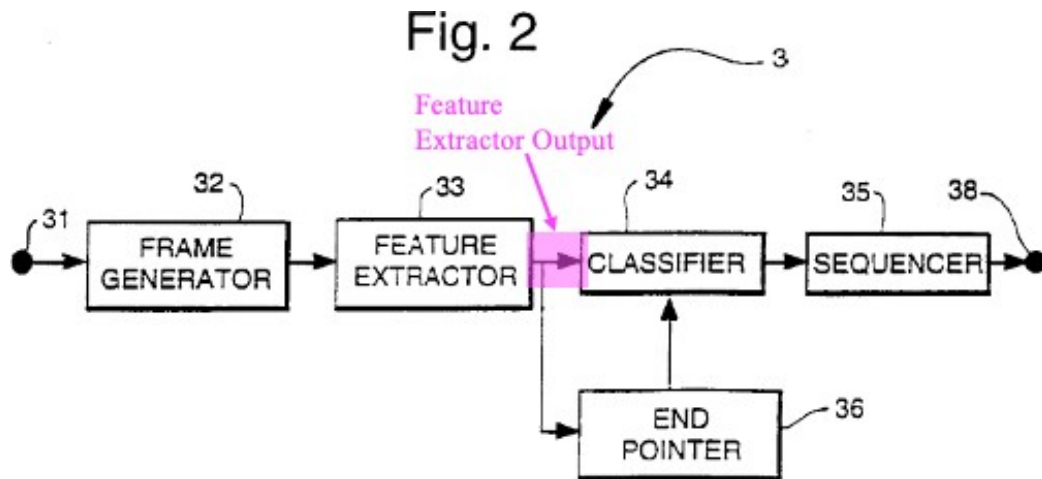
d) “for calculating distances ...”

Smyth teaches calculating distances for indicating the similarity, stating:

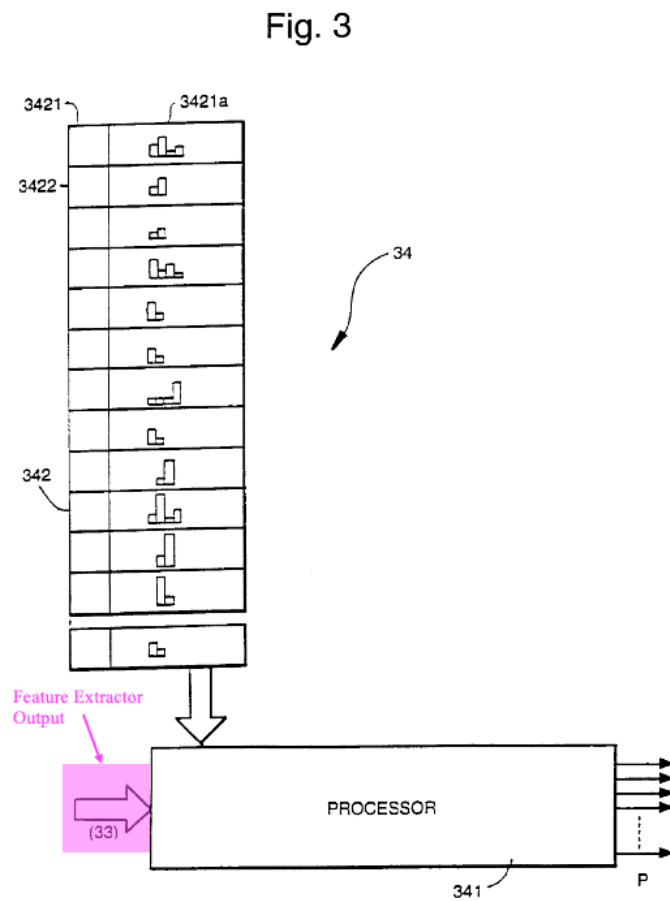
The **classification processor 34[1]** is arranged to **read each state field** within the memory 342 in turn, and ***calculate for each***, using the current input **feature coefficient set**, the ***probability*** that the **input *feature* set or *vector* corresponds to the corresponding state**. [¶]

Accordingly, the **output** of the classification processor is ***a plurality of state probabilities P, one for each state*** in the state memory 342, ***indicating the likelihood that the input feature vector corresponds to each state***.

Smyth, 5:63–6:4. *Smyth* further discloses “a classifier 34 **receiving the succession of feature vectors and operating on each with a plurality of model states**.” *Smyth*, 5:7-9, 5:55-57 (“the classifier 34 comprises a classifying processor 341 and a state memory 342”). Figs. 2-3 depict the output of feature extractor 33 as the input to classifier processor 341.



Smyth, Fig. 2 (annotated).



Smyth, Fig. 3 (annotated).

Smyth teaches or renders obvious the claimed “calculating distances[.]” *Dec.*, 285-288. *Smyth*’s calculated probabilities *P* is a “distance” because the probability indicates the likelihood, i.e., similarity, between the feature vector and a stored acoustic state. *Dec.*, 287; *Smyth*, 5:63–6:4. *Smyth* teaches calculating a plurality of probabilities (i.e., distances) for a “*plurality* of predetermined acoustic states” because a probability is calculated for “each state” of the acoustic model. *Dec.*, 287-288; *Smyth*, 6:1-4 (calculating a plurality of probabilities “**one for each state** in the state memory”). The classifier processor 341 (“calculating circuit”) is *for calculating distances* because processor 341 receives the feature vector from feature extractor 33 and calculates the probability indicating the likelihood the feature vector corresponds to the state. *Dec.*, 287-288; *Smyth*, 5:7-9. The probability (“distance”) is for a “plurality of predetermined *acoustic states* of an *acoustic model*” per Sections VIII.A.4.b-c, which discuss *Smyth*’s model is an acoustic model comprising “a plurality of model states,” namely “speech states.” *Smyth*, 1:34-37, 5:9, 5:59; *Dec.*, 288.

e) **Motivation to Modify *Jiang-Baumgartner-Brown* with *Smyth***

A POSITA would have found it obvious and been motivated to modify *Jiang-Baumgartner-Brown* to include *Smyth*’s “calculating circuit for calculating distances....” Although *Jiang* expressly teaches calculating a probability or likelihood that an observed feature vector matches an HMM state, *Smyth* provides

additional disclosures relating to how the distance calculation is performed, e.g., via probabilities “indicating the likelihood that the input feature vector corresponds to each state.” *Smyth*, 6:1-4. In the modified *Jiang* system, the speech recognition system includes *Baumgartner*’s SLA (the claimed “second processor”) implementing *Smyth*’s “calculating circuit for calculating distances...”

A POSITA would have been motivated to replace *Jiang*’s probability scores for each phoneme branch (where the “score represents the likelihood that the particular phoneme encoded by the codeword corresponds to the phoneme for the branch under consideration”), with *Smyth*’s probability calculation that an input feature vector corresponds to a given state. *Jiang*, 8:29-33; *Smyth*, 5:63-67; *Dec.*, 289-97. *Jiang* provides the background architecture for performing distance calculations, including utilizing HMMs to identify a “most likely phoneme represented by the codeword received from feature extraction module[.]” *Jiang*, 7:36-42. Performing distance calculations to determine the most likely phonemes would have entailed use of a known technique, i.e., calculating distances by comparing a feature vector to a state of an acoustic model, such as an HMM, to improve similar devices, i.e., *Jiang*’s speech recognition system. *Dec.*, 295. Instead of performing *Jiang*’s “most likely phoneme” based on stored HMMs (which *Jiang* further elaborates on, 1:19–2:24), it would have been obvious to substitute *Smyth*’s probability calculation for an input feature vector compared to a state, as such as

substitution of one known element for another would have obtained predictable results. *Dec.*, 296. Because distance calculation was well-known and commonly used (*Ravishankar*, 21, 28-29), the simple substitution would have been readily implemented by a POSITA. *Dec.*, 296.

A POSITA would have also had a reasonable expectation of success in implementing *Smyth*'s "calculating circuit" on *Baumgartner*'s SLA because *Baumgartner* teaches the "heart" of the SLA is a "circuit." *Dec.*, 297 (citing *Baumgartner*, [0043]).

5. Claim 1(d)

See Section V.A.5.

6. Claim 1(e)

See Section V.A.6.

7. Claim 1(f)

See Section V.A.7.

B. Claim 5

See Section V.B.

C. Claim 7

See Section V.C.

D. Claim 12

See Section V.D.

E. Claim 14

1. Claim 14(Pre)

See Section V.E.1.

2. Claim 14(a)

See Section V.E.2.

3. Claim 14(b)

See Section V.E.3.

4. Claim 14(c)

See Sections V.E.4, VIII.A.4.

Additionally, *Jiang*, *Baumgartner*, and *Smyth* each teaches a “calculating means” comprising a processor, as properly construed. *See* Section III.C.1. The incorporated sections map the construed structure (processor) and function. Specifically, *Jiang*, 6:39-42 teaches tree search engine may be a dedicated processor, *Baumgartner*, [0202] teaches processor may be equipped with the SLA, and *Smyth*, 6:5-7 teaches classifier processor 341 may be a processor such as a DSP.

5. Claim 14(d)

See Section V.E.5.

6. Claim 14(e)

See Section V.E.6.

F. Claim 15

1. Claim 15(Pre)

See Section V.F.1.

2. Claim 15(a)

See Section V.F.2.

3. Claim 15(b)

See Section V.F.3.

4. Claim 15(c)

See Sections V.F.4, VIII.A.4.

5. Claim 15(d)

See Section V.F.5.

6. Claim 15(e)

See Section V.F.6.

G. Claim 16

1. Claim 16(Pre)

See Section V.G.1.

2. Claim 16(a)

See Section V.G.2.

3. Claim 16(b)

See Section V.G.3.

4. Claim 16(c)

See Sections V.G.4, VIII.A.4.

5. Claim 16(d)

See Section V.G.5.

6. Claim 16(e)

See Section V.G.6.

IX. GROUND 5: CLAIM 4 IS OBVIOUS OVER *JIANG, BAUMGARTNER, BROWN, SMYTH, AND KAZEROONIAN*

The *Jiang-Baumgartner-Brown-Kazeroonian* system teaches or renders obvious Claim 4. See Section VI.

Further, *Smyth* teaches or renders obvious the “calculating circuit for calculating distances ...” limitation. See Section VIII.A.4.

X. GROUND 6: CLAIMS 9-10 ARE OBVIOUS OVER *JIANG, BAUMGARTNER, BROWN, SMYTH AND VENSKO*

A. Claim 9

The *Jiang-Baumgartner-Brown-Vensko* and *Jiang-Baumgartner-Brown-Smyth* circuits each teaches Claim 9. See Sections VII.A, VIII.A.4.

B. Claim 10

The *Jiang-Baumgartner-Brown-Vensko* and *Jiang-Baumgartner-Brown-Smyth* circuits each teaches Claim 10. See Sections VII.B, VIII.A.4.

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

§314(a). None of the factors articulated in *General Plastic Industrial Co., Ltd. v. Canon Kabushiki Kaisha*, IPR2016-01357, Paper 19 at 16 (PTAB Sept. 6, 2017) favor discretionary denial. Regarding Factors 1–5, Petitioners have filed no prior petition regarding the ’277 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 (EX1063), 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. (IPR2023-00034, Paper 10; Factor 6).

§325(d). Petitioners seek to join already instituted IPR2023-00034, based on the same grounds as applied to the same claims. Petitioners agree to assume an understudy role and will only actively participate in proceedings if Apple's participation terminates.

XII. CONCLUSION

For the forgoing reasons, Petitioners respectfully request *inter partes* review of the Challenged Claims of the '277 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

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

A. Real Party-In-Interest

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

B. Related Matters

The '277 Patent is currently the subject of *Apple Inc. v. Zentian Ltd.*, Inter Partes Review No. IPR2023-00034 (P.T.A.B.) (instituted Jun. 12, 2023). This Petition is substantively identical to the petition submitted in IPR2023-00034 and Petitioners have concurrently filed a motion to join that proceeding.

The '277 Patent is also asserted in the following pending patent infringement lawsuits filed by Zentian Ltd:

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

C. Lead and Back-Up Counsel

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

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

Petition for *Inter Partes* Review of
U.S. Patent No. 7,979,277

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

D. Service Information

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

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

CLAIM LISTING APPENDIX

U.S. Patent No. 7,979,277 for Claims 1, 4-5, 7, 9-10, 12, and 14-16

Claim Designation	Claim Language
Claim 1(Pre)	A speech recognition circuit, comprising:
Claim 1(a)	an audio front end for calculating a feature vector from an audio signal,
Claim 1(b)	wherein the feature vector comprises a plurality of extracted and/or derived quantities from said audio signal during a defined audio time frame;
Claim 1(c)	a calculating circuit for calculating distances indicating the similarity between a feature vector and a plurality of predetermined acoustic states of an acoustic model; and
Claim 1(d)	a search stage for using said calculated distances to identify words within a lexical tree, the lexical tree comprising a model of words;
Claim 1(e)	wherein said audio front end and said search stage are implemented using a first processor, and said calculating circuit is implemented using a second processor, and
Claim 1(f)	wherein data is pipelined from the front end to the calculating circuit to the search stage.
Claim 4	A speech recognition circuit as claimed in claim 1, wherein the first processor supports multi-threaded operation, and runs the search stage and front ends as separate threads.
Claim 5	A speech recognition circuit as claimed in claim 1, wherein the said calculating circuit is configured to autonomously calculate distances for every acoustic state defined by the acoustic model.
Claim 7	The speech recognition circuit of claim 1, wherein the feature vector comprises a plurality of spectral components of an audio signal for a predetermined time frame.

Claim Designation	Claim Language
Claim 9	The speech recognition circuit of claim 1, wherein the speech accelerator has an interrupt signal to inform the front end that the accelerator is ready to receive a next feature vector from the front end.
Claim 10	The speech recognition circuit of claim 1, wherein the accelerator signals to the search stage when the distances for a new frame are available in a result memory.
Claim 12	The speech recognition circuit of claim 1, wherein the audio front end is configured to input a digital audio signal.
Claim 14(Pre)	A speech recognition circuit, comprising:
Claim 14(a)	an audio front end for calculating a feature vector from an audio signal,
Claim 14(b)	wherein the feature vector comprises a plurality of extracted and/or derived quantities from said audio signal during a defined audio time frame;
Claim 14(c)	calculating means for calculating a distance indicating the similarity between a feature vector and a predetermined acoustic state of an acoustic model; and
Claim 14(d)	a search stage for using said calculated distances to identify words within a lexical tree, the lexical tree comprising a model of words;
Claim 14(e)	wherein said audio front end, said calculating means, and said search stage are connected to each other to enable pipelined data flow.
Claim 15(Pre)	A speech recognition method, comprising:
Claim 15(a)	calculating a feature vector from an audio signal using an audio front end,

Claim Designation	Claim Language
Claim 15(b)	wherein the feature vector comprises a plurality of extracted and/or derived quantities from said audio signal during a defined audio time frame;
Claim 15(c)	calculating a distance indicating the similarity between a feature vector and a predetermined acoustic state of an acoustic model using a calculating circuit; and
Claim 15(d)	using a search stage to identify words within a lexical tree using said calculated distances, the lexical tree comprising a model of words;
Claim 15(e)	wherein data is pipelined from the front end, to the calculating circuit, and to the search stage.
Claim 16(Pre)	A non-transitory storage medium storing processor implementable code for controlling at least one processor to implement a speech recognition method, the code comprising:
Claim 16(a)	code for controlling the processor to calculate a feature vector from an audio signal,
Claim 16(b)	wherein the feature vector comprises a plurality of extracted and/or derived quantities from said audio signal during a defined audio time frame;
Claim 16(c)	code for controlling the processor to calculate a distance indicating the similarity between a feature vector and a predetermined acoustic state of an acoustic model; and
Claim 16(d)	code for controlling the processor to identify words within a lexical tree using said calculated distances, the lexical tree comprising a model of words,
Claim 16(e)	wherein data is pipelined by the processor pursuant to the code from the feature calculation, to the distance calculation, and to the word identification.

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 13,984 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: July 12, 2023

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

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

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

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

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