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

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

Petitioners

- vs. -

ZENTIAN LIMITED

Patent Owner

Patent No. 10,839,789

Issued: November 17, 2020

Inventor: Guy Larri, et al.

Title: SPEECH RECOGNITION CIRCUIT AND METHOD

Inter Partes Review No. IPR2023-01195

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 10,839,789**

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 '789 PATENT	1
A. Patent Description.....	1
B. Summary of Unpatentability	2
C. Priority Date.....	3
D. Level of Skill of a POSITA	3
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 Challenges (37 C.F.R. §42.104(b))	4
C. Claim Construction (37 C.F.R. §42.104(b)(3)).....	5
IV. SHOWING OF ANALOGOUS PRIOR ART	6
A. Smyth.....	6
B. Mozer.....	6
C. Nguyen.....	7
D. Houser.....	8
E. Vensko	8
F. Comeau	8
G. Boike.....	9
H. Whittaker	9
I. Ichikawa.....	9
J. Seshadri	10

TABLE OF CONTENTS
(Continued)

	Page
K. Steinbiss.....	10
V. GROUND 1: CLAIMS 1, 10-11, 14, AND 17 ARE OBVIOUS OVER <i>SMYTH</i> AND <i>MOZER</i>	10
A. Claim 1(Pre).....	10
1. “audio signal”	11
2. “data associated with an audio signal”	11
3. “acoustic coprocessor”	11
B. Claim 1(a)	18
1. Smyth’s Teachings	18
a) “wherein the feature vector is determined from the audio signal”	18
b) “a first interface for receiving at least one feature vector”	20
2. Mozer’s Teachings	23
C. Claim 1(b)	25
1. “a plurality of acoustic states”	25
2. “an acoustic model memory for storing an acoustic model defining a plurality of acoustic states”	27
D. Claim 1(c)	30
1. “a calculating apparatus”	30
2. “[a calculating apparatus] for calculating distances indicating a similarity ...”	31
E. Claim 1(d)	34

TABLE OF CONTENTS
(Continued)

	Page
1. Smyth’s Teachings	34
2. Mozer’s Teachings	38
F. Claim 1(e)	39
1. Smyth’s Teachings	39
2. Mozer’s Teachings	41
3. Motivation to Combine Smyth in View of Mozer.....	42
G. Claim 10(Pre).....	45
H. Claim 10(a)	45
I. Claim 10(b)	45
J. Claim 10(c)	45
K. Claim 11	45
L. Claim 14.....	45
M. Claim 17.....	47
VI. GROUND 2: CLAIMS 2, 4, 6-9, 12, 16, 18, AND 20-21 ARE OBVIOUS OVER <i>SMYTH</i> , <i>MOZER</i> , AND <i>NGUYEN</i>	48
A. Claim 2.....	48
1. Nguyen’s Teachings	48
2. Motivation to Combine Smyth and Nguyen.....	50
B. Claim 4.....	51
C. Claim 6.....	51
D. Claim 7.....	51

TABLE OF CONTENTS
(Continued)

	Page
E. Claim 8.....	52
F. Claim 9(Pre).....	54
G. Claim 9(a)	54
H. Claim 9(b)	56
1. Nguyen’s Teachings	57
2. Motivation to Combine Smyth and Nguyen.....	58
I. Claim 12.....	59
J. Claim 16.....	59
1. Claim 16(Pre)	59
2. Claim 16(a).....	59
3. Claim 16(b).....	59
K. Claim 18.....	60
L. Claim 20.....	60
M. Claim 21	60
VII. GROUND 3: CLAIM 13 IS OBVIOUS OVER <i>SMYTH</i> , <i>MOZER</i> , AND <i>HOUSER</i>	61
A. Claim 13.....	61
VIII. GROUND 4: CLAIMS 22-24 ARE OBVIOUS OVER <i>SMYTH</i> , <i>MOZER</i> , <i>NGUYEN</i> , AND <i>VENSKO</i>	61
A. Claim 22.....	61
B. Claim 23.....	62

TABLE OF CONTENTS **(Continued)**

	Page
C. Claim 24.....	63
IX. GROUND 5: CLAIMS 25-28 AND 37 ARE OBVIOUS OVER <i>SMYTH, MOZER, NGUYEN, VENSKO, AND COMEAU</i>	64
A. Claim 25.....	64
B. Claim 26.....	66
C. Claim 27.....	67
D. Claim 28(Pre).....	68
E. Claim 28(a)	69
F. Claim 28(b)	69
G. Claim 37.....	69
X. GROUND 6: CLAIM 29 IS OBVIOUS OVER <i>SMYTH,</i> <i>MOZER, NGUYEN, VENSKO, COMEAU, AND BOIKE</i>	69
A. Claim 29.....	69
XI. GROUND 7: CLAIM 35 IS OBVIOUS OVER <i>SMYTH,</i> <i>MOZER, NGUYEN, VENSKO, COMEAU, AND WHITTAKER</i>	72
A. Claim 35.....	72
XII. GROUND 8: CLAIMS 38-39 ARE OBVIOUS OVER <i>SMYTH,</i> <i>MOZER, NGUYEN, VENSKO, COMEAU, AND ICHIKAWA</i>	73
A. Claim 38.....	73
B. Claim 39.....	74
XIII. GROUND 9: CLAIMS 42-44 ARE OBVIOUS OVER <i>SMYTH,</i> <i>MOZER, NGUYEN, VENSKO, COMEAU, ICHIKAWA, AND</i> <i>SESHADRI</i>	74

TABLE OF CONTENTS
(Continued)

	Page
A. Claim 42.....	74
B. Claim 43.....	75
C. Claim 44.....	76
XIV. GROUND 10: CLAIM 45 IS OBVIOUS OVER SMYTH, MOZER, NGUYEN, VENSKO, COMEAU, ICHIKAWA, AND STEINBISS.....	77
A. Claim 45.....	77
XV. THE BOARD’S DISCRETION UNDER 35 U.S.C. §314(a) AND §325(d).....	78
XVI. CONCLUSION.....	79
XVII. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8(A)(1)	80
A. Real Party-In-Interest.....	80
B. Related Matters	80
C. Lead and Back-Up Counsel	80
D. Service Information	81
CLAIM LISTING APPENDIX	82

TABLE OF AUTHORITIES

	Page(s)
CASES	
<i>Apple Inc. v. Fintiv, Inc.</i> , IPR2020-00019 (PTAB, Mar. 20, 2021)	78
<i>General Plastic Industrial Co., Ltd. v. Canon Kabushiki Kaisha</i> , IPR2016-01357, Paper 19 (PTAB Sept. 6, 2017).....	78
<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	<i>passim</i>
35 U.S.C. § 315(c)	1, 4
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

EXHIBIT LIST (37 C.F.R. § 42.63(E))

Exhibit	Description
1001	U.S. Patent No. 10,839,789 (“’789 Patent”)
1002	Prosecution History for the ’789 Patent (“’789 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	[RESERVED]
1008	Harry Newton, <i>Newton’s Telecom Dictionary</i> , 19 th Edition, CMP Books, March 2003 (“Newton”)
1009	Kevin L. Kloker, “The Motorola DSP56000 Digital Signal Processor,” 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	J. Mariani, <i>Recent Advances in Speech Processing</i> , IEEE. International Conference on Acoustics, Speech, and Signal Processing, pp. 429-440. May 1989 (“Mariani”)
1014	Zentian, Ltd. v Apple Inc., 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”)

TABLE OF CONTENTS
(Continued)

Page

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, <i>A Tutorial on Hidden Markov Models and 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</i> ‘447’”)	
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> ”)	

TABLE OF CONTENTS
(Continued)

Page

Exhibit	Description
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> ”)
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	[RESERVED]
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	U.S. Patent No. 5,774,859 to Houser et al. (“ <i>Houser</i> ”)
1041	U.S. Patent No. 7,171,358 to Whittaker et al. (“ <i>Whittaker</i> ”)
1042	U.S. Patent No. 7,478,041 to Ichikawa et al. (“ <i>Ichikawa</i> ”)
1043	U.S. Patent No. 7,587,318 to Seshadri (“ <i>Seshadri</i> ”)
1044	U.S. Patent Application Publication No. 2002/0091826 to Comeau et al. (“ <i>Comeau</i> ”)
1045	U.S. Patent Application Publication No. 2003/0012347 to Steinbiss et al. (“ <i>Steinbiss</i> ”)
1046	U.S. Patent No. 6,832,194 to Mozer et al. (“ <i>Mozer</i> ”)
1047	U.S. Patent No. 6,879,954 to Nguyen et al. (“ <i>Nguyen</i> ”)
1048	U.S. Patent No. 5,150,330 to Hag (“ <i>Hag</i> ”)
1049	U.S. Patent No. 5,724,587 to Carmon et al. (“ <i>Carmon</i> ”)
1050	[RESERVED]
1051	[RESERVED]

TABLE OF CONTENTS
(Continued)

Page

Exhibit	Description	
1052	<i>[RESERVED]</i>	
1053	Hermann Ney et al., “ <i>Dynamic Programming Search Strategies: From Digit Strings To Large Vocabulary Word Graphs</i> ” C.-H. Lee et al. (eds.), Automatic Speech and Speaker Recognition (“ <i>Ney</i> ”)	
1054	U.S. Patent No. 7,058,678 to Luick et al. (“ <i>Luick</i> ”)	
1055	<i>[RESERVED]</i>	
1056	<i>[RESERVED]</i>	
1057	<i>[RESERVED]</i>	
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	Zentian, Ltd. v Apple Inc., WDTX Waco (6:22-cv-122) Scheduling Order (“ <i>Scheduling Order</i> ”)	
1062	U.S. Patent No. 5,734,604 to Akamatsu et al. (“ <i>Akamatsu</i> ”)	
1063	U.S. Patent No. 6,130,462 to Yang et al. (“ <i>Yang 462</i> ”)	
1064	U.S. Patent No. 6,128,311 to Poulis et al. (“ <i>Poulis</i> ”)	
1065	U.S. Patent No. 6,185,634 to Wilcox (“ <i>Wilcox</i> ”)	

TABLE OF CONTENTS
(Continued)

Page

Exhibit	Description
1066	U.S. Patent No. 6,567,953 to Pomerantz (“ <i>Pomerantz</i> ”)
1067	Declaration of Les Atlas, Ph.D.
1068	<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 IPR2023-00036. 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”) request *Inter Partes* Review of Claims 1-2, 4, 6-14, 16-18, 20-29, 35, 37-39, and 42-45 (“Challenged Claims”) of USPN 10,839,789, assigned to Zentian Limited upon information and belief. ’789 Patent (EX1001). The purportedly distinguishing feature of the Challenged Claims—a calculating apparatus and memory on a single integrated chip—was well-known before the priority date of the ’789 Patent, and the Challenged Claims are obvious over the prior art as detailed herein. Accordingly, IPR of the Challenged Claims should be instituted.

II. SUMMARY OF THE ’789 PATENT

A. Patent Description

The ’789 Patent is directed to an acoustic coprocessor for processing audio data, including calculating distances indicating the similarity of a feature vector to an acoustic state of an acoustic model. ’789 Patent, 4:12-17, 24:21-31. Figure 16 below depicts a speech recognition system comprising a distance calculation engine 104, feature vectors input to the distance calculation engine, an acoustic model

memory 105, and “Distances or Probabilities or Likelihoods” output from the distance calculation engine:

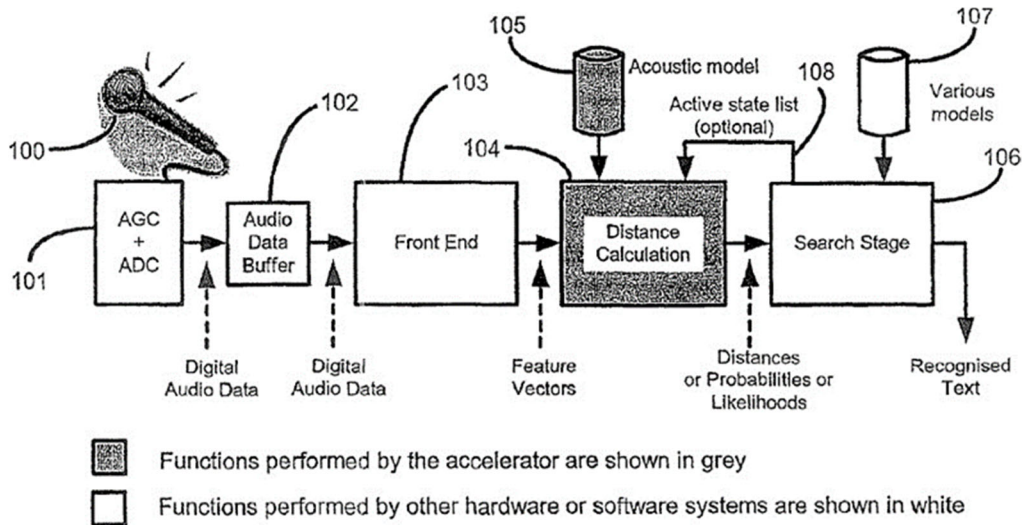


Figure 16

Id., Fig. 16.

B. Summary of Unpatentability

Claim 1 of the '789 Patent recites a well-known acoustic coprocessor for calculating distances (i.e., probabilities or likelihoods, per the '789 Patent, Fig. 16) indicating the similarity of a feature vector associated with an audio signal compared to a stored acoustic state of an acoustic model. '789 Patent, Claim 1. The '789 Patent even admits that coprocessors for calculating distances (i.e., probabilities) were known. *Id.* at 1:58-65 (discussing an article that uses a “special purpose co-processor for up-front calculation of the computationally expensive Gaussian output probabilities of audio frames corresponding to particular states in the acoustic

model”). The purported novelty is that a calculating apparatus and memory of the coprocessor is fabricated on a single IC. However, numerous prior art references, including *Mozer* cited in the Grounds, teach mounting a processor and memory associated with speech recognition on a single IC.

C. Priority Date

The ’789 Patent claims its earliest priority to a GB application filed September 14, 2004. ’789 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 ’789 Patent would have had a master’s degree in computer engineering, computer science, electrical engineering, or a related field, with at least two years of experience in the field of speech recognition, or a bachelor’s degree in the same fields with at least four years of experience in the field of speech recognition. Additional education or experience might substitute for the above requirements. *Dec.*, 24-25.¹

¹ Submitted herewith is the Declaration of Les Atlas, Ph.D., who has reviewed, agrees with, and adopts the opinions in the Schmandt Declaration (IPR2023-00036, EX1003). EX1067, ¶¶ 25-26. For consistency with Apple’s Petition, citations to Dr.

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 '789 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 IPR2023-00036, this petition is timely under 35 U.S.C. §315(c) and 37 C.F.R. §42.122.

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

In view of the prior art and evidence presented, the Challenged Claims are unpatentable and should be cancelled. 37 C.F.R. §42.104(b)(1). 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, 10-11, 14, and 17 are obvious under §103(a) over <i>Smyth</i> and <i>Mozer</i>
<u>Ground 2</u>	Claims 2, 4, 6-9, 12, 16, 18, and 20-21 are obvious under §103(a) over <i>Smyth</i> , <i>Mozer</i> , and <i>Nguyen</i>
<u>Ground 3</u>	Claim 13 is obvious under §103(a) over <i>Smyth</i> , <i>Mozer</i> , and <i>Houser</i>

Atlas's opinion are to paragraph numbers of EX1003 containing those opinions as expressed by Mr. Schmandt.

<u>Ground 4</u>	Claims 22-24 are obvious under §103(a) over <i>Smyth, Mozer, Nguyen, and Vensko</i>
<u>Ground 5</u>	Claims 25-28 and 37 are obvious under §103(a) over <i>Smyth, Mozer, Nguyen, Vensko, and Comeau</i>
<u>Ground 6</u>	Claim 29 is obvious under §103(a) over <i>Smyth, Mozer, Nguyen, Vensko, Comeau, and Boike</i>
<u>Ground 7</u>	Claim 35 is obvious under §103(a) over <i>Smyth, Mozer, Nguyen, Vensko, Comeau, and Whittaker</i>
<u>Ground 8</u>	Claims 38-39 are obvious under §103(a) over <i>Smyth, Mozer, Nguyen, Vensko, Comeau, and Ichikawa</i>
<u>Ground 9</u>	Claims 42-44 are obvious under §103(a) over <i>Smyth, Mozer, Nguyen, Vensko, Comeau, Ichikawa, and Seshadri</i>
<u>Ground 10</u>	Claim 45 is obvious under §103(a) over <i>Smyth, Mozer, Nguyen, Vensko, Comeau, Ichikawa, and Steinbiss</i>

Sections V-XIV 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-XIV. 37 C.F.R. §42.104(b)(5). Exhibits 1001–1068 are also attached.

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

In this proceeding, claims are interpreted under the same standard applied by Article III courts (i.e., the *Phillips* standard). 37 C.F.R. §42.100(b); 83 Fed. Reg. 197 (Oct. 11, 2018); *Phillips v. AWH Corp.*, 415 F.3d 1303, 1312 (Fed. Cir. 2005) (*en banc*). Petitioners apply the plain and ordinary meaning of all claim terms as understood by a POSITA for all terms not expressly construed below. Petitioners do

not waive any argument in any litigation that claim terms in the '789 Patent are indefinite or additional terms need construction.

IV. SHOWING OF ANALOGOUS PRIOR ART

A. Smyth

Smyth was neither cited nor considered during prosecution of the '789 Patent. *Smyth* issued October 6, 1998, qualifying as prior art under at least 35 U.S.C. §102(b). *Smyth* (EX1005).

Because *Smyth*, like the '789 Patent, discloses an acoustic coprocessor for processing data associated with an audio signal, *Smyth* “relates to speech recognition circuits and methods” and is in the same field of endeavor as the '789 Patent. '789 Patent, 1:20-21; Dec., 92-93 (*comparing Smyth*, 5:20-46, 5:55–6:4, 6:36-63, with '789 Patent, 12:58-65, 13:3-38, 24:14-20).

Both *Smyth* and the '789 Patent recognize the challenges with performing speech recognition tasks using limited computing resources; therefore, *Smyth* is reasonably pertinent to a problem described by the '789 Patent. Dec., 92-93.

Therefore, *Smyth* is analogous art.

B. Mozer

Mozer was neither cited nor considered during prosecution of the '789 Patent. *Mozer* was filed October 26, 2000, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(e). *Mozer* (EX1046).

Because *Mozer*, like the '789 Patent, discloses a speech recognition system using feature vectors and similarity calculation, and “relates to speech recognition circuits and methods,” *Mozer* is in the same field of endeavor as the '789 Patent. '789 Patent, 1:20-21; Dec., 94 (comparing *Mozer*, 7:65-8:9, 8:58-66, (describing feature vector extraction), 9:22-26, 9:38-42, 10:17-20 (describing similarity calculation), with '789 Patent, 12:58-65 (describing calculation of feature vectors), 13:3-38 (describing distance calculation)).

Both *Mozer* and the '789 Patent recognize the challenges with performing speech recognition tasks using limited computing resources, and therefore, *Mozer* is reasonably pertinent to a problem described by the '789 Patent. Dec., 94.

Therefore, *Mozer* is analogous art.

C. Nguyen

Nguyen was neither cited nor considered during prosecution of the '789 Patent. *Nguyen* was filed April 22, 2002, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(e). *Nguyen* (EX1047).

Because *Nguyen* discloses a method for improving a speech recognition system, *Nguyen* is reasonably pertinent to a problem described by the '789 Patent. *Nguyen*, Fig. 1 (excerpt), 2:21-29, 2:31-38, 3:11-23; Dec., 95-96.

Therefore, *Nguyen* is analogous art.

D. Houser

Houser was neither cited nor considered during prosecution of the '789 Patent. *Houser* issued June 30, 1998, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(b). *Houser* (EX1040).

Houser teaches a speech recognition system and is reasonably pertinent to a problem described by the '789 Patent. *Dec.*, 97. Therefore, *Houser* is analogous art.

E. Vensko

Vensko was neither cited nor considered during prosecution of the '789 Patent. *Vensko* issued January 28, 1986, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(b). *Vensko* (EX1038).

Vensko teaches a speech recognition system and is reasonably pertinent to a problem described by the '789 Patent. *Dec.*, 98. Therefore, *Vensko* is analogous art.

F. Comeau

Comeau was neither cited nor considered during prosecution of the '789 Patent. *Comeau* published July 11, 2002, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(b). *Comeau* (EX1044).

Comeau teaches a speech recognition system and is reasonably pertinent to a problem described by the '789 Patent. *Dec.*, 99. Therefore, *Comeau* is analogous art.

G. Boike

Boike was neither cited nor considered during prosecution of the '789 Patent. *Boike* was filed October 11, 2001, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(e). *Boike* (EX1006).

Boike teaches a speech recognition system and is reasonably pertinent to a problem described by the '789 Patent. *Dec.*, 100. Therefore, *Boike* is analogous art.

H. Whittaker

Whittaker was neither cited nor considered during prosecution of the '789 Patent. *Whittaker* was filed January 13, 2003, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(e). *Whittaker* (EX1041).

Whittaker teaches a speech recognition system and is reasonably pertinent to a problem described by the '789 Patent. *Dec.*, 101. Therefore, *Whittaker* is analogous art.

I. Ichikawa

Ichikawa was neither cited nor considered during prosecution of the '789 Patent. *Ichikawa* was filed March 12, 2003, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(e). *Ichikawa* (EX1042).

Ichikawa teaches a speech recognition system and is reasonably pertinent to a problem described by the '789 Patent. *Dec.*, 102. Therefore, *Ichikawa* is analogous art.

J. Seshadri

Seshadri was neither cited nor considered during prosecution of the '789 Patent. *Seshadri* was filed September 12, 2003, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(e). *Seshadri* (EX1043).

Seshadri teaches a speech recognition system and is reasonably pertinent to a problem described by the '789 Patent. *Dec.*, 103. Therefore, *Seshadri* is analogous art.

K. Steinbiss

Steinbiss was neither cited nor considered during prosecution of the '789 Patent. *Steinbiss* published January 16, 2003, qualifying as prior art to the '789 Patent under at least 35 U.S.C. §102(b). *Steinbiss* (EX1045).

Steinbiss teaches a speech recognition system and is reasonably pertinent to a problem described by the '789 Patent. *Dec.*, 104. Therefore, *Steinbiss* is analogous art.

**V. GROUND 1: CLAIMS 1, 10-11, 14, AND 17 ARE OBVIOUS OVER
SMYTH AND MOZER**

A. Claim 1(Pre)

To the extent the preamble is limiting, *Smyth* teaches Claim 1(Pre). *Smyth* discloses a “speech recognition system [that] recognizes connected speech using a plurality of vocabulary nodes[.]” *Smyth*, Abstract.

1. “audio signal”

Smyth teaches an “audio signal,” namely digitized, sampled input speech. *Smyth*’s recognition processor 3 “comprises an input 31 for receiving **speech in digital form**[.]”² and “a frame processor 32 for partitioning the **succession of digital samples** into a succession of frames of contiguous samples[.]” *Smyth*, 5:1-7. The speech in digital form and partitioned into digital samples is an “audio signal,” as claimed. *Dec.*, 106.

2. “data associated with an audio signal”

Smyth teaches data associated with an audio signal, such as feature vectors. *Smyth* teaches a “feature extractor 33 for generating from a frame of samples a corresponding **feature vector**[.]” *Smyth*, 5:6-7. Feature vectors are well-known to comprise cepstral data that is representative of the spectral information of an audio signal. *Smyth*, 1:24-28, 5:31-46; *Dec.*, 108-109. Therefore, the extracted feature vectors are, at the least, data associated with an audio signal. *Dec.*, 107-109.

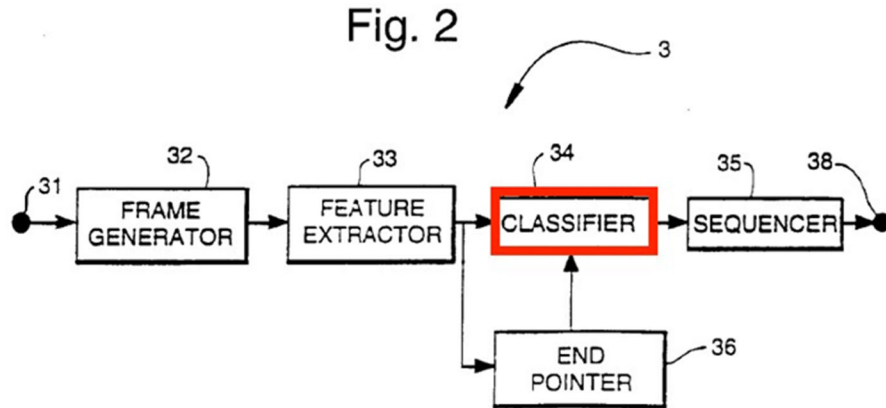
3. “acoustic coprocessor”

Claim 1 is directed to an acoustic coprocessor that includes first and second interfaces, a calculating apparatus, and an acoustic model memory. Claim 1’s recited hardware is generally mapped as follows:

² All emphases added, unless otherwise noted.

- *Smyth*'s classifier 34 and at least one of signal lines, or input/output for receiving/transmitting data, are collectively mapped as the "acoustic coprocessor";
- *Smyth*'s classifier and at least one of the signal line between the feature extractor and the classifier, or the input at the classifier, are collectively mapped as the first interface;
- *Smyth*'s classifier and at least one of the signal line between the classifier and the sequencer, or the output at the classifier, are collectively mapped as the second interface;
- *Smyth*'s classifier processor 341 is mapped as the "calculating apparatus"; and
- *Smyth*'s state memory 342 is mapped as the "acoustic model memory."

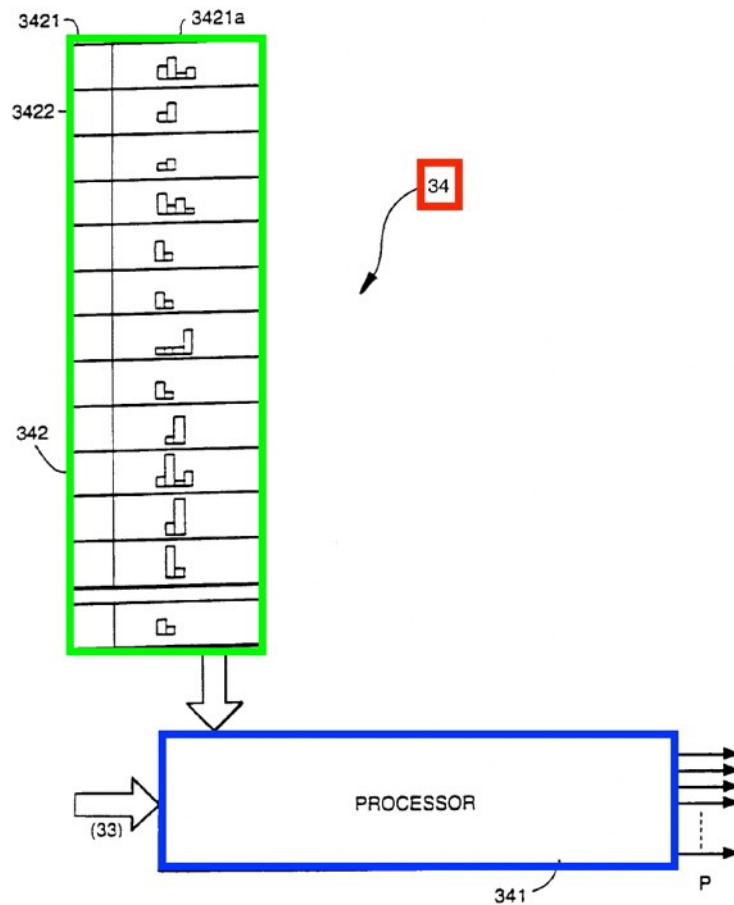
Smyth teaches a classifier 34 that receives feature vectors from feature extractor 33 and outputs calculated probabilities to sequencer 35. *Smyth*, 5:6-9, 5:12-15, 6:1-4, Figs. 2-3; *Dec.*, 116.



Smyth, Fig. 2 (annotated). Per Mr. Schmandt, classifier 34 is communicatively coupled to feature extractor 33 and sequencer 35 because data is transmitted from extractor 33 to classifier, and from classifier to sequencer. *Smyth*, 5:6-18, 5:63-67, 6:1-4, Fig. 2; *Dec.*, 117.

Classifier 34 comprises classifier processor 341 and state memory 342. *Smyth*, 5:55-57, 4:7-8.

Fig. 3



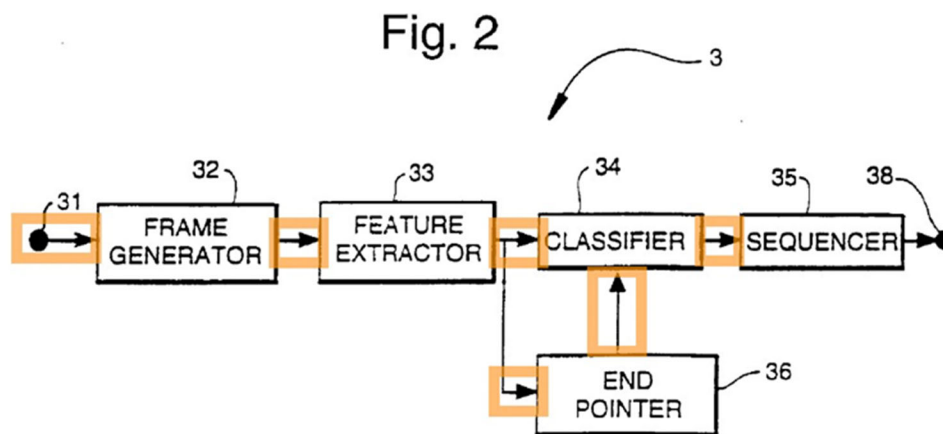
Smyth, Fig. 3 (annotated).

Smyth further teaches a frame generator 32 and the feature extractor 33 may be provided by a single DSP. *Smyth*, 5:49-53. The classifier processor 341 of classifier may also be a DSP, where the DSP may be the same type or model of DSP (e.g., Motorola DSP56000) as the DSP for frame generator and feature extractor. *Smyth*, 6:5-8; *Dec.*, 119-125.

Smyth thus teaches that classifier receives data (feature vectors) from feature extractor, and that sequencer receives data (probabilities) from classifier. *Dec.*, 126.

The transmission of data would have been understood by a POSITA to include transmitting along one or more signal lines or other communication paths and received at the classifier at an input. *Id.*

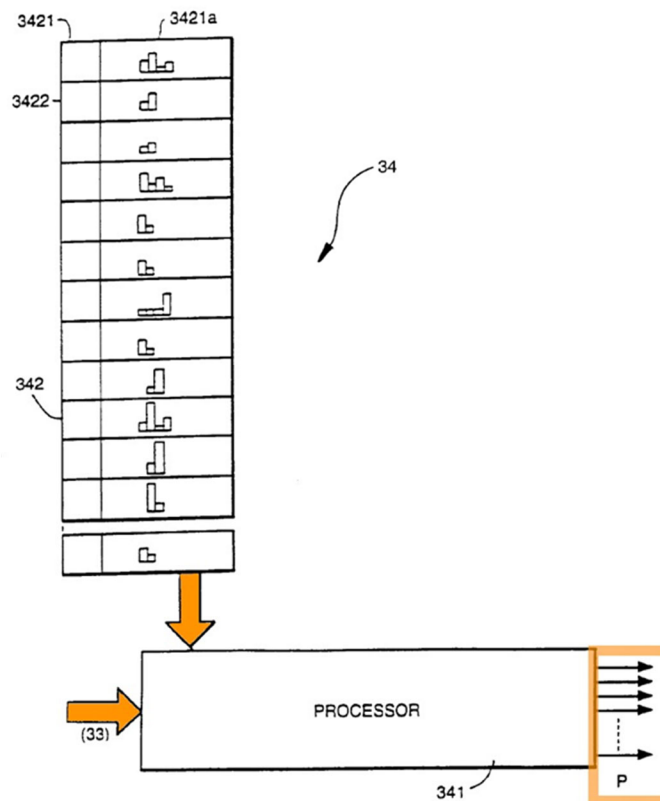
First, *Smyth* expressly states data is received by classifier (and likewise, sequencer). *Smyth*, 5:7-8 (“classifier 34 **receiving** the succession of feature vectors”), 5:12-14 (“sequencer 35 which is arranged to **receive** the classification results form [*sic*] the classifier”), 6:1-4 (referring to the **output** of the classification processor); *Dec.*, 126 (*citing Kloker*, pp. 3-4, 14-15 and noting the exemplary DSP in *Smyth* (Motorola DSP56000) includes I/O ports for transmitting and receiving data). *Smyth* also expressly illustrates the transmission of data between various components of the recognition processor:



Smyth, Fig. 2 (annotated).

Second, *Smyth* discloses that data-carrying signals are transmitted/received among the components of the recognition system. *Smyth*, 3:58-61 (discussing “signal processing” of path signatures), 4:38-45 (discussing transmitting both a “speech signal” and “a signal indicating identification or otherwise of a particular word or phrase”). For example, *Smyth* illustrates data being sent from memory 342 to classifier processor 341, and similarly from extractor 33 to processor 341 and from processor 341 out. *Dec.*, 128.

Fig. 3



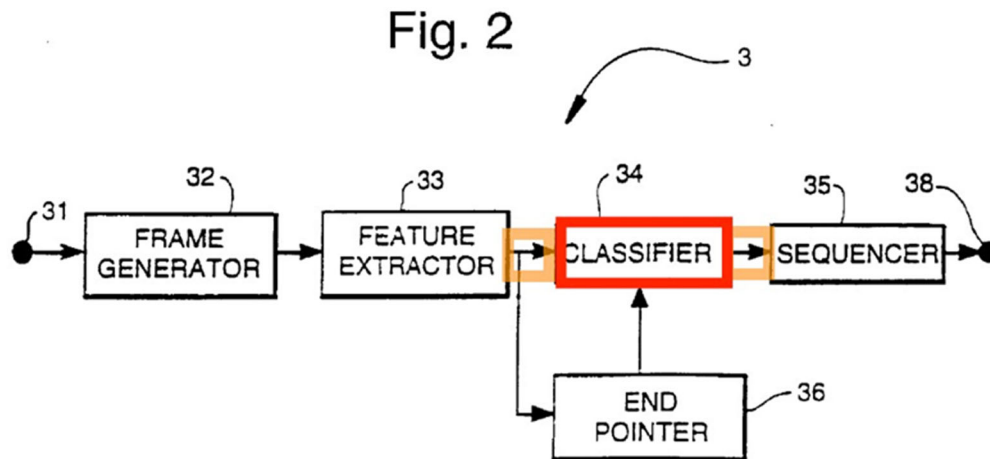
Smyth, Fig. 3 (annotated).

Third, *Smyth* teaches the classification processor 341 may be a “digital signal processing (DSP) device” and that it communicates with a DSP associated with the feature extractor and a microprocessor or another DSP associated with the sequencer. *Smyth*, 6:5-8, 5:49-53, 6:47-51; *Dec.*, 126-131. Per Mr. Schmandt, it is well-known that the transmission/receipt of data between electronic components, such as DSPs and microprocessors, entails communication along signal lines and inputs and outputs, such as ports. *Dec.*, 126.

Therefore, a POSITA would have understood classifier 34 is associated with (i) at least a first signal line or communication path between feature extractor 33 and classifier; and (ii) an input at classifier. *Dec.*, 126, 129. Additionally, classifier is associated with (i) at least a second signal line or communication path between classifier and sequencer; and (ii) an output at classifier. *Id.* Thus, the claimed “acoustic coprocessor” includes the classifier (which further includes classifier processor 341 and memory 342) and either or both of the associated signal lines or the input/output. *Dec.*, 126-131.

Finally, *Smyth*’s classifier and at least one of the associated signal lines or input/output are collectively a *coprocessor* because the classifier is designed to operate with the feature extractor and the sequencer (and other components of speech recogniser 3) to calculate distances and performed other claimed functionality. *Dec.*, 129. The classifier and associated signal lines are collectively an *acoustic*

coprocessor because the classifier is designed to operate on acoustic data, namely the input speech signal. *Smyth*, 5:1-18.



Smyth, Fig. 2 (annotated); *Dec.*, 129.

Thus, a POSITA would have understood that *Smyth*'s classifier 34 is an acoustic coprocessor for processing feature vectors (i.e., data associated with an audio signal, as claimed). *Dec.*, 126-131.

B. Claim 1(a)

Smyth alone or in combination with *Mozer* teaches or renders obvious Claim 1(a).

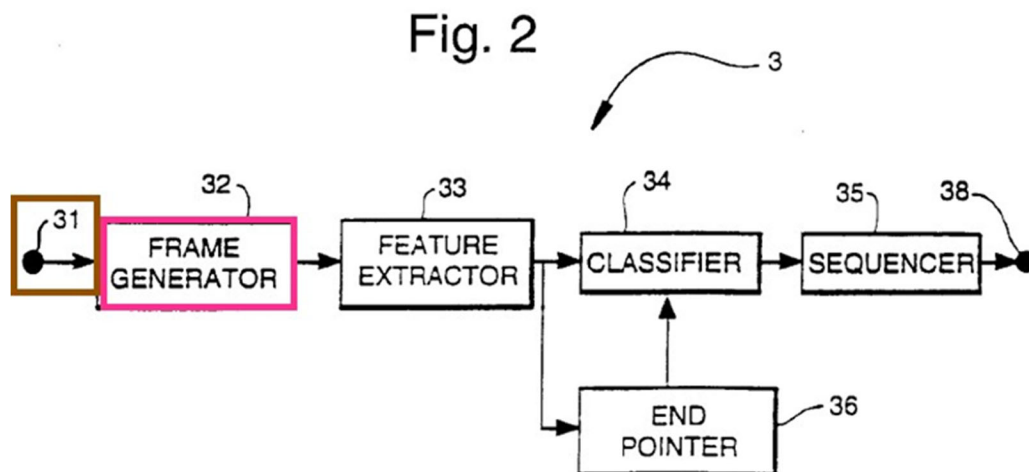
1. *Smyth*'s Teachings

a) “wherein the feature vector is determined from the audio signal”

Smyth teaches it is “common in speech recognition processing to input speech data, typically in digital form, to a so-called front-end processor, which derives from

the stream of input speech data a more compact, perceptually significant set of data referred to as a front-end **feature set or vector.**” *Smyth*, 1:24-28. *Smyth* teaches “speech is typically input via a microphone, sampled, digitised, segmented into frames of length 10-20 ms (e.g. sampled at 8 kHz) and, **for each frame, a set of coefficients is calculated.**” *Smyth*, 1:29-32; *Dec.*, 132.

Before determining feature vectors for input speech, *Smyth* teaches the input and digitized speech (i.e., the claimed “audio signal” per Section 1(Pre)) is partitioned into frames. Specifically, *Smyth* teaches “recognition processor 3 comprises an input 31 for receiving speech in digital form (either form [*sic*] a digital network or from an analog to digital converter), a frame processor 32 for partitioning the succession of digital samples into a succession of frames of contiguous samples[.]” *Smyth*, 5:1-5. *Smyth*’s Fig. 2, annotated below, illustrates input 31 for receiving speech data (annotated brown) and frame processor 32 (annotated pink) that receives speech samples and generates frames:



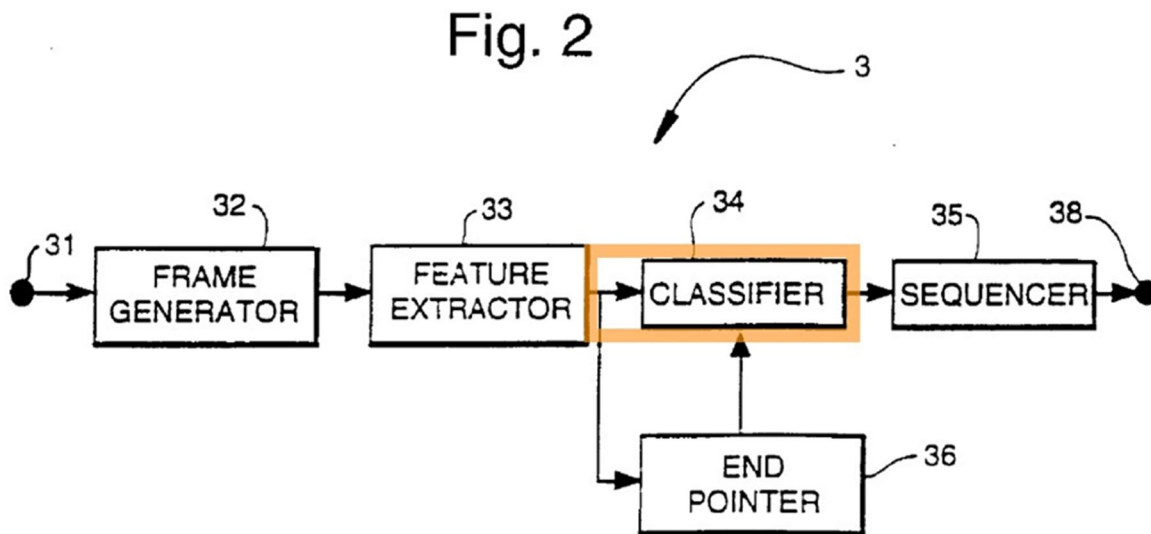
Smyth, Fig. 2 (annotated). After frame generator 32 generates the frames, feature extractor 33 receives frames of digitized speech and “***generates, in each case, a set or vector of features.***” *Smyth*, 5:31-33, 5:6-7 (“a feature extractor 33 for generating from a frame of samples a corresponding **feature vector**”). *Dec.*, 132.

Because feature extractor 33 generates feature vectors from the frames of digitized speech (*Smyth*, 5:6-7), a POSITA would have understood *Smyth* teaches the feature vector is determined from the audio signal, i.e., the “succession of digital samples.” *Smyth*, 5:3-5; *Dec.*, 132.

b) “a first interface for receiving at least one feature vector”

Petitioners incorporate by reference the Claim 1(Pre) mapping for the “acoustic coprocessor.”

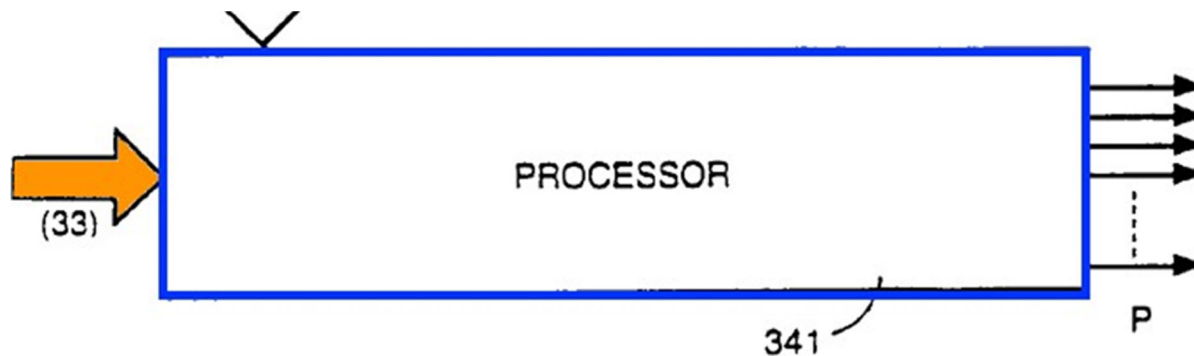
Smyth teaches “a first interface for receiving at least one feature vector,” namely the classifier 34 and at least one of the associated signal line or communication path between feature extractor and classifier transmitting the feature vectors to classifier 34 or the input at the classifier for receiving the feature vectors. *Smyth* teaches “classifier 34 **receiv[es] the succession of feature vectors**” *Smyth*, 5:6-9. Figure 2 shows the recognition processor’s elements (*Smyth*, 4:4-6) and depicts classifier 34 “receiving the succession of feature vectors” from feature extractor 33 via the signal line (annotated orange).



Smyth, Fig. 2 (annotated), 5:7-9. Therefore, *Smyth* teaches the collective classifier and signal line or communication path between feature extractor and classifier “receiv[es] at least one feature vector,” as claimed, because the signal line or communication path transmits feature vectors from extractor to classifier. *Dec.*, 134.

Smyth also teaches or renders obvious that the classifier and at least one of the associated signal line or communication path between extractor and classifier 34 or the input at classifier is a “first interface,” as claimed. *Dec.*, 134-136.

First, *Smyth* teaches classifier 34 includes classifier processor 341, which is mapped for Claim 1(c) as the claimed “calculating apparatus.” Figure 3, below, depicts classifier processor 341 (annotated blue) receiving feature vectors, represented by an arrow (annotated orange), from feature extractor 33.



Smyth, Fig. 3 (excerpt, annotated); *Dec.*, 134.

Per Mr. Schmandt, classifier 34 and the signal line or communication path from feature extractor 33 is an “interface,” as claimed, because *Smyth* discloses classifier 34 includes components, such as classifier processor 341 and state memory 342, for receiving the feature vectors and for performing processing thereon, and for further outputting data, such as calculated probabilities. *Dec.*, 135 (discussing the ’789 Patent’s usage of an “interface”), 134 (opining that classifier 34 including processor 341, which may be a DSP, and the communication path from feature extractor renders obvious an interface); *Smyth*, 5:12-15, 5:63-67 (discussing calculating probability based on input feature vector), 6:13-15 (discussing classifier processor 341 outputs probabilities), Fig. 2 (showing output from classifier 34 to sequencer 35), Fig. 3 (showing output of probabilities from classifier processor 341 to sequencer 35).

Second, the classifier and its input teaches or renders obvious a “first interface.” *Smyth* teaches a DSP for the classifier (the Motorola DSP56000) that

includes I/O ports for transmitting and receiving data, further indicating to a POSITA that the classifier with the *input* is a “first interface.” *Dec.*, 126, 136; *Smyth*, 6:5-8, 5:49-53; *Kloker*, pp. 3-4, 14-15.

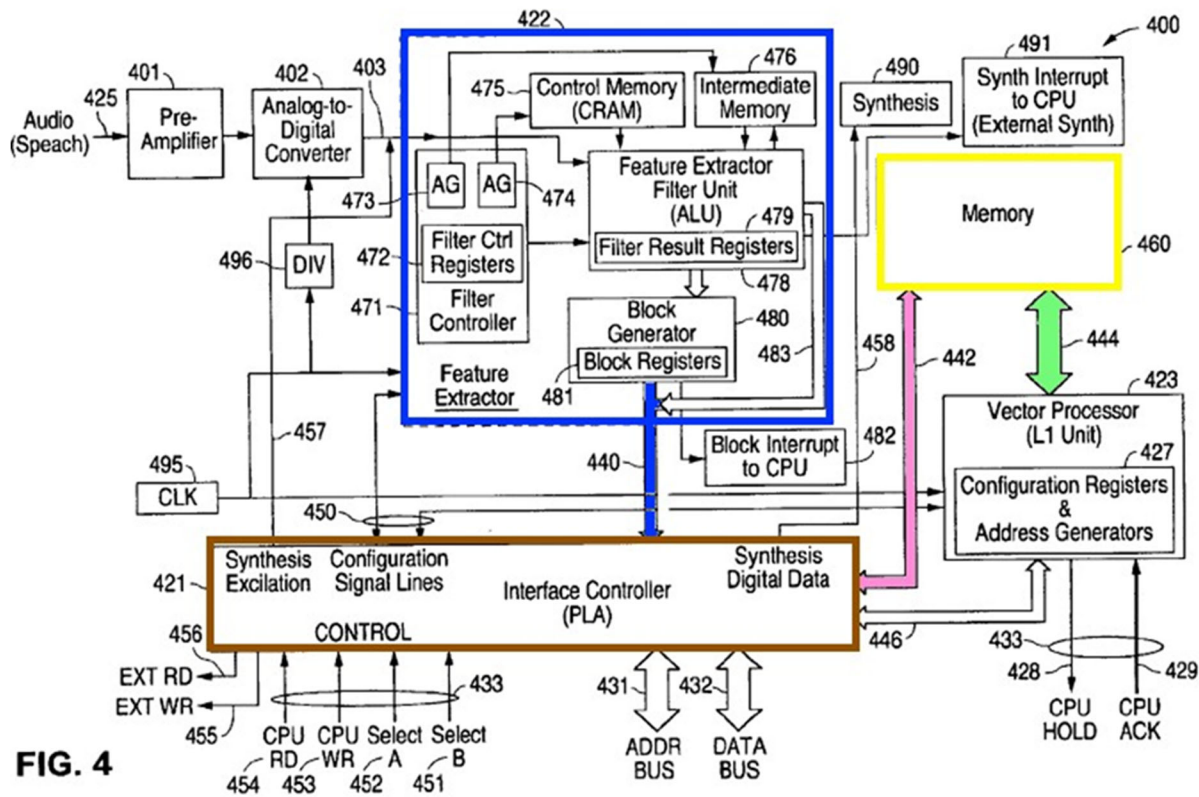
Because classifier both receives and outputs data from other electronic components (feature extractor 33 and sequencer 35), a POSITA would have understood the collective classifier and at least one of a signal line or communication path between extractor and classifier or the input at classifier is or includes an *interface* or that such is obvious. *Dec.*, 136.

2. *Mozer’s Teachings*

Mozer teaches a “first interface” for receiving at least one feature vector, namely signal lines such as “feature extractor output lines 440,” “memory input lines 442,” and memory interface line 444. To the extent Patent Owner contends *Smyth* does not teach or render obvious a “first interface” as claimed, it would have been obvious to modify *Smyth*’s system to include *Mozer*’s signal lines. *Dec.*, 137.

Mozer discloses a feature extractor 422 (annotated blue) sends feature vectors via feature extractor output lines 440 (annotated blue) through interface controller 421 (annotated brown) and **memory input lines 442** (annotated pink) to memory 460 (annotated yellow). *Dec.*, 137. A **memory interface line 444** (annotated green) is shown interfacing between memory and vector processor. *Id.* (opening line 442 is not discussed in *Mozer*’s written description (only Fig. 4) but that the similar

embodiment of Fig. 3, 6:55–7:27 discusses line 342 between memory and vector processor, where feature vectors are received along line 342). A POSITA would have understood either or both of lines 442 and 444 were the claimed “first interface for receiving at least one feature vector,” because these signal lines receive feature vectors. *Mozer*, 9:54-56 (“vector processor accesses the vector stored in memory 460 over memory input lines 444”), *Dec.*, 138.



Mozer, Fig. 4 (annotated).

A POSITA would have found it obvious and been motivated to modify *Smyth*’s classifier 34 to include or otherwise be associated with *Mozer*’s signal lines

between electronic components, such as *Smyth*'s feature extractor 33 and sequence extractor 35. *Smyth* already teaches classifier includes a processor 341 that *receives feature vectors* from feature extractor 33, thus indicating to a POSITA that feature vector information is received by classifier 34. *See* Sections V.A, V.B.1.b. Signal lines between electronic components for transmitting data therebetween is a routine design element of computer systems, per Mr. Schmandt and as evidenced by *Mozzer* (which simply provides additional detail on the signaling lines). *Dec.*, 139. To the extent *Smyth* does not teach or render obvious signal lines between components, the proposed modification would have entailed use of a known technique (i.e., adding such signal lines to transmit data) to improve similar devices (i.e., *Smyth*'s feature extractor 33) in the same way (provide for transmission of data). *Dec.*, 139.

C. Claim 1(b)

Smyth teaches or renders obvious Claim 1(b).

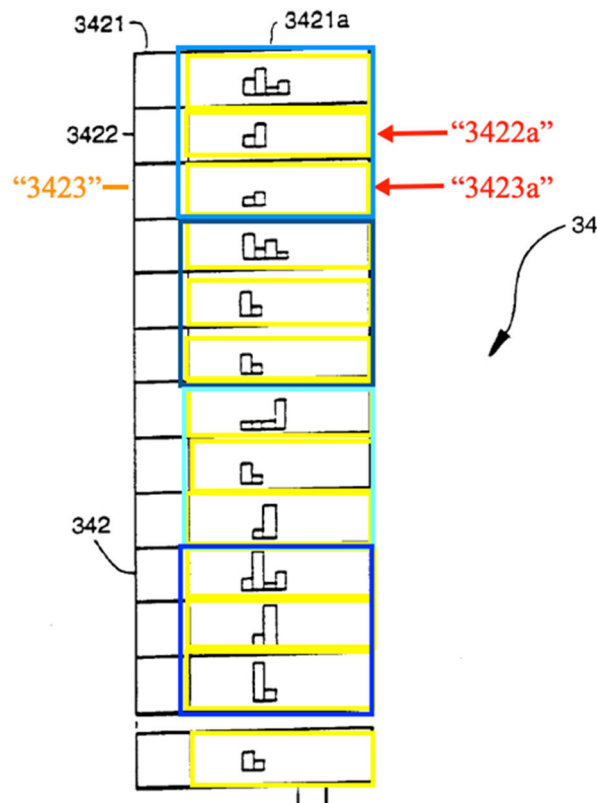
1. “a plurality of acoustic states”

Smyth teaches the claimed “plurality of acoustic states,” namely speech states stored in state fields of state memory.

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, per Mr. Schmandt, 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.*, 141-142 (*citing Smyth*, 5:58-61). Annotated Fig. 3 below depicts state memory 342, state fields 3421–3422 (with state field 3423 added and annotated orange), speech state 3421a (with speech states 3422a-3423a added and annotated red), and groups of three speech states corresponding to an allophone (annotations added 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 fields 3422. *Dec.*, 141-142; *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, namely alternative pronunciations of phonemes. *Dec.*, 143-146.

2. ***“an acoustic model memory for storing an acoustic model defining a plurality of acoustic states”***

Regarding the claimed “acoustic model,” *Smyth* teaches or renders such obvious. *Smyth* teaches a “template or model,” including a “plurality of model states,” that are used to determine a measure of similarity and recognition results for spoken speech. *Smyth*, 1:34-39, 5:7-9. *Smyth*’s Background of the Invention 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.

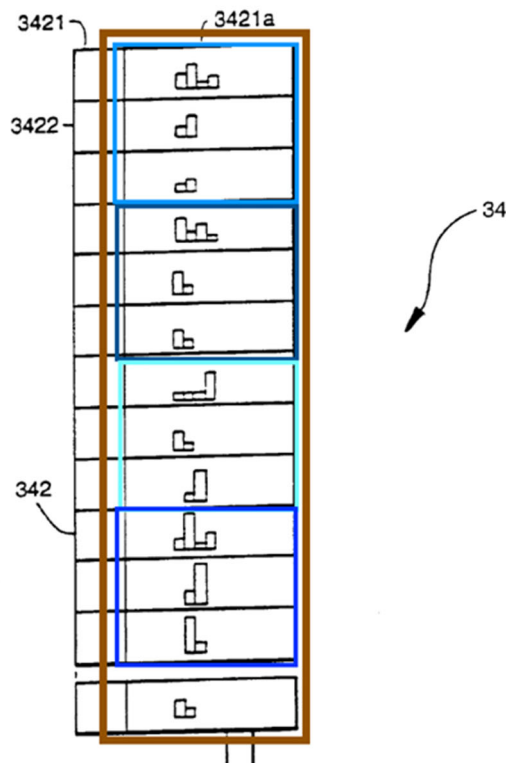
Smyth, 1:34-39. *Smyth* further teaches “classifier 34 receiv[es] the succession of feature vectors and operat[es] on each with **a plurality of model states**, to generate recognitions results[.]” *Smyth*, 5:6-9.

Mr. Schmandt explains that it is well-known to train the speech recognition system to generate models or templates of spoken input, as otherwise taught by *Smyth*, 1:34-39. *Dec.*, 148 (citing Ney, p. 4 (teaching estimating probabilities of acoustic vectors “during the training phase of the recognition system”), *Jiang*, 11:25-28, *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) is used to determine the likelihood that a feature vector corresponds to one of the speech states. *Dec.*, 148-149. 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, i.e., the disclosed “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.*, 152. 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.*, 152. 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 the claimed “acoustic model.”

Regarding the claimed “acoustic model *memory*,” *Smyth* teaches a plurality of allophones stored in state memory 342: “each allophone to be recognized 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. As discussed above, this collection of a plurality of allophones stored as states (where each allophone comprises three speech states stored in three state fields) is an “acoustic model.” Because state memory 342 stores model speech states comprising allophones representative of acoustic speech, *Smyth* teaches or renders obvious an “acoustic model memory,” as claimed, where the acoustic model comprises acoustic states, namely speech states 3241a, etc. *Dec.*, 153.

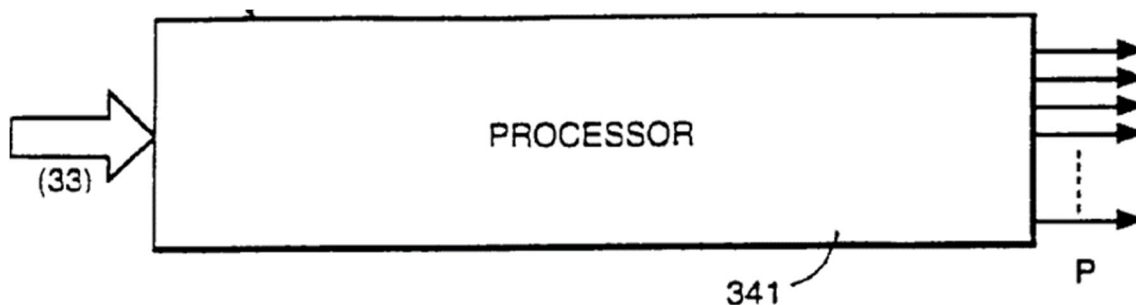
D. Claim 1(c)

As discussed above, *Smyth* teaches or renders obvious acoustic states of an acoustic model. *See* Claim 1(b).

1. “a calculating apparatus”

Smyth teaches a “calculating apparatus for calculating distances,” as claimed, namely classifier processor 341, which calculates “state probabilities P, one for each state in the state memory 342, indicating the likelihood that the input feature vector

corresponds to each [acoustic] state” of the acoustic model read from state memory 342 (i.e., acoustic model memory). *Smyth*, 5:55–6:4; *Dec.*, 155. *Smyth*’s Figure 3 identifies classifier processor 341 as a *processor*.



Smyth, Fig. 3 (excerpt), 5:55-56, 6:1, 6:5. Because the classifier processor is a *processor* that calculates state probabilities (discussed further in the following section), a POSITA would have understood that classifier processor 341 was a “calculating apparatus,” as claimed. *Dec.*, 155.

Additionally, or alternatively, *Smyth* teaches “classifier processor 341 may be a suitably programmed digital signal processing (DSP) device[.]” *Smyth*, 6:5-6. The classifier processor implemented as a DSP and calculating the probabilities would also have been understood to be a “calculating apparatus,” as claimed. *Dec.*, 156.

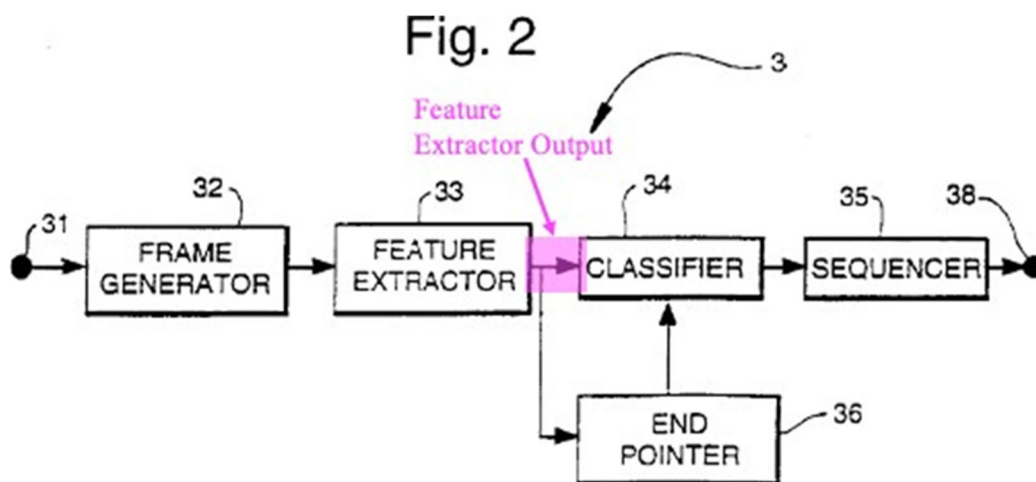
2. “[a calculating apparatus] for calculating distances indicating a similarity ...”

Smyth teaches classifier processor 341 (i.e., “calculating apparatus”) calculates distances indicating the similarity between a feature vector (i.e., feature vector calculated in feature extractor 33) and respective acoustic states (i.e., the

speech states in state fields 3421, 3422, etc.) of an acoustic model (i.e., the collection of speech states of the allophones in state memory 342). *Smyth* teaches:

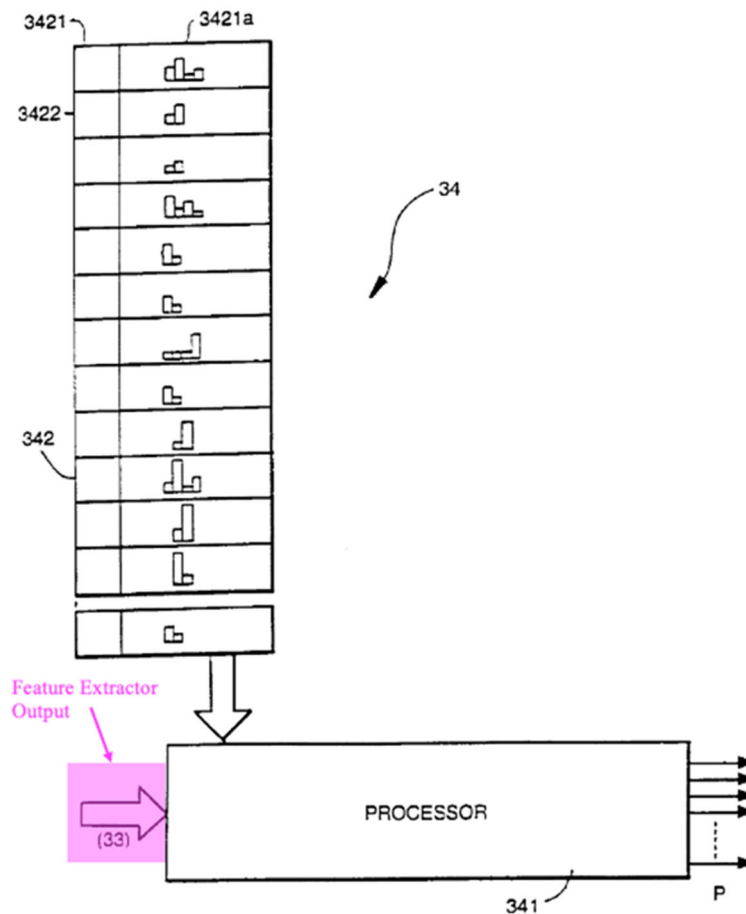
The **classification processor 34[1]** is arranged to **read each state field** within the memory 342 in turn, and **calculate for each**, using the current input **feature coefficient set**, the **probability** that the **input feature** set or **vector corresponds to the corresponding state**. [¶] 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 (disclosing “the classifier 34 comprises a classifier processor 341 and a state memory 342”). *Smyth*’s Figs. 2-3 depict the output of feature extractor 33 as the input to classifier processor 341.



Smyth, Fig. 2 (annotated).

Fig. 3



Smyth, Fig. 3 (annotated).

Per Mr. Schmandt, *Smyth* teaches or renders obvious the claimed “calculating distances.” *Dec.*, 158. Each of *Smyth*’s calculated probabilities P is a “distance,” as claimed, because the probability indicates the likelihood, i.e., similarity, between the feature vector and the stored acoustic state. *Dec.*, 158-159; *Smyth*, 5:63–6:4. *Smyth* teaches calculating a *plurality* of probabilities, i.e., distances (plural), for “respective acoustic states” because a probability is calculated for “each state” of the acoustic model. *Smyth*, 6:1-4 (teaching calculating a plurality of probabilities “**one for each**

state in the state memory”). The classification processor 341 (i.e., the claimed “calculating apparatus”) is *for calculating distances* because the processor 341 receives the feature vector from feature extractor 33 and calculates the probability indicating the likelihood the feature vector corresponds to the state. *Smyth*, 5:7-9 (discussing classifier receiving feature vectors and generating recognitions results by operating on a plurality of model states). The probability (i.e., distance) is for respective acoustic states “of the acoustic model read from the acoustic model memory,” as claimed, because *Smyth* expressly teaches the processor 341 “is arranged to **read each state field within the memory**” to calculate the probability, and each state field includes a speech state. *Smyth*, 5:58–6:4, 1:34-37, 5:9, 5:59, Fig. 2; *Dec.*, 159; Section V.A.3 (mapping *Smyth*’s acoustic states of an acoustic model).

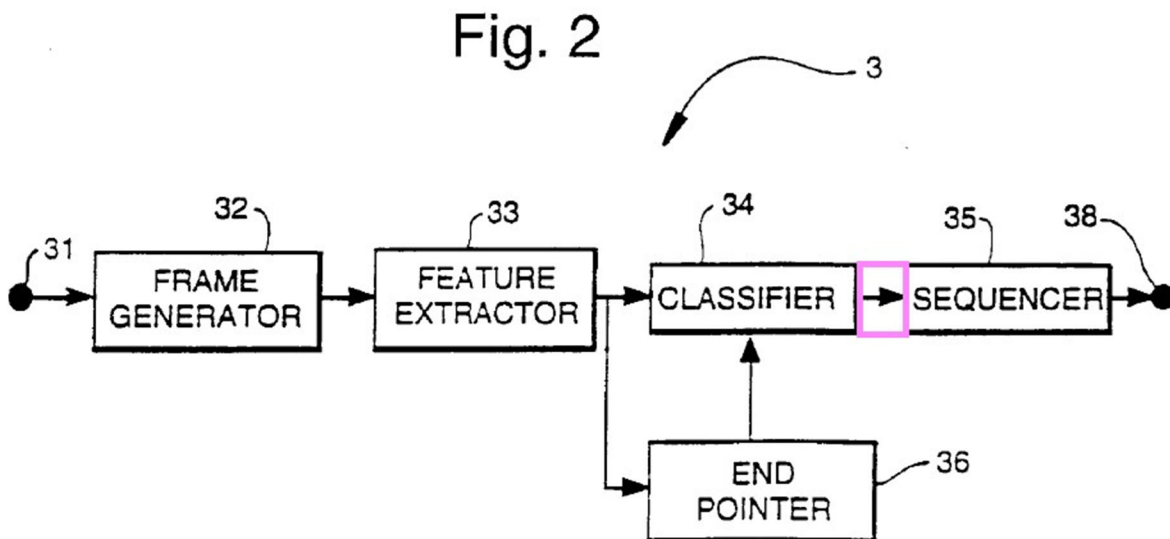
E. Claim 1(d)

Smyth alone or in combination with *Mozzer* teaches or renders obvious Claim 1(d).

1. *Smyth*’s Teachings

Smyth teaches a “second interface,” as claimed, namely the classifier 34 and at least one of the associated signal line or communication path between the classifier and sequencer 35 or the output at classifier. *See* Section V.B.1.b (discussing the “first interface”). *Smyth* teaches classifier processor 341 of classifier sends at least one calculated probability (i.e., distance) to sequencer 35: “a sequencer 35 which is

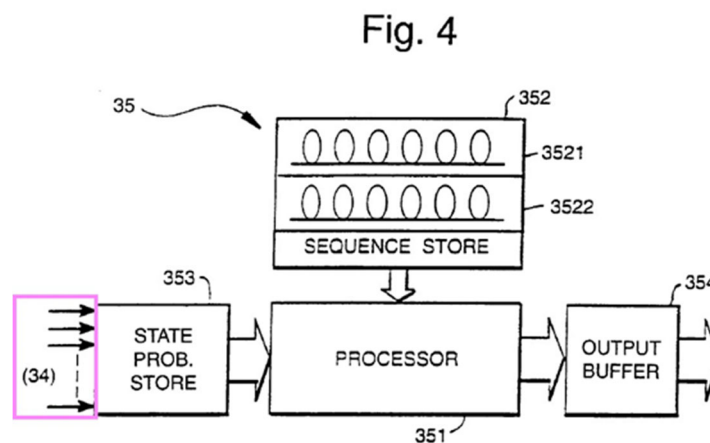
arranged to **receive the classification results** form [sic] **the classifier 34** and to determine the predetermined utterance to which the sequence of classifier output indicates the greatest similarity[.]” *Smyth*, 5:12-15. Figure 2 depicts classifier 34 transmitting classification results to sequencer 35:



Smyth, Fig. 2 (annotated). As mapped for Claim 1(c), the probability (i.e., distance) is “calculated by the calculating apparatus,” as claimed, namely the classifier including classifier processor.

The “classification results” discussed at *Smyth*, 5:7-9 that are sent from classifier 34 to sequencer 35 are the probabilities (i.e., distances) calculated by the classifier processor at *Smyth*, 5:63-67. *Smyth* discloses the sequencer includes a state probability memory 353 “which stores, for each frame processed, the state probabilities output by the classification processor,” i.e., the claimed “distances.” *Smyth*, 6:13-15; *Dec.*, 160. The sequencer (which includes parsing processor 351,

per *Smyth*, 6:10-12) accesses the stored probabilities to perform word identification: “The parsing processor 351 is arranged to read, at each frame, the state probabilities output by the classifier processor 341...and to calculate the most likely path of states to date over time” *Smyth*, 6:36-42. Figure 4 further illustrates state probability memory 353 receiving the probabilities from classifier 34:



Smyth, Fig. 4 (annotated). Therefore, a POSITA would have understood state probability memory 353 of sequencer 35 is receiving state probabilities (i.e., distances) as calculated by classifier processor 341. *Dec.*, 160.

Because *Smyth* teaches the collective classifier and signal line or communication path between classifier and sequencer (which includes state probability memory 353) transmits state probabilities from classifier to sequencer, *Smyth* teaches “sending at least one distance calculated by the calculating apparatus,” as claimed. *Dec.*, 160.

Smyth also teaches or renders obvious that the classifier 34 and at least one of the associated signal line or communication path between classifier and sequencer or the output at classifier is a “second interface,” as claimed. *Dec.*, 161. Because *Smyth* discloses sequencer 35 includes components, such as state probability memory 353 and parsing processor 351, for receiving the calculated probabilities, a POSITA would have understood the collective classifier and signal line or communication path between classifier and sequencer is a “second interface,” as claimed. *Dec.*, 162 (discussing the ’789 Patent’s usage of an “interface”), 163 (opining that classifier 34 including processor 341, which may be a DSP, and communication path between classifier and sequencer renders obvious an interface); *Smyth*, 5:12-15, 5:63-67 (discussing calculating probability based on input feature vector), 6:13-15 (discussing classifier processor 341 outputs probabilities), Fig. 2 (showing output from classifier 34 to sequencer 35), Fig. 3 (showing output of probabilities from classifier processor 341 to sequencer 35).

As also discussed for Claim 1(a) and the “first interface” (the mapping for which is incorporated by reference), *Smyth* teaches a DSP model for the classifier including I/O ports, further confirming to a POSITA that *Smyth*’s classifier include an output for serving as an interface for sending data. *Dec.*, 136; *Smyth*, 6:5-8, 5:49-53; *Kloker*, pp. 3-4, 14-15.

Because classifier both receives and outputs data from other electronic components (feature extractor 33 and sequencer 35), a POSITA would have understood the collective classifier and signal line or communication path between classifier and sequencer is or includes an *interface* or that such is obvious. *Dec.*, 161.

2. *Mozer's Teachings*

Similar to the mapping for the claimed “first interface,” *Mozer* also teaches a “second interface” for sending a calculated distance. Figure 2 of *Mozer* discloses an “audio recognition peripheral” configured to “run alongside a processor” (i.e., to be a *coprocessor*). *Mozer*, Fig. 2, 5:60-67. Figure 3 discloses an embodiment of the audio recognition peripheral 300 on a single IC. *Mozer*, 6:1-3. Audio recognition peripheral includes vector processor for calculating a “vector distance.” *Mozer*, 5:47-53 (discussing the Fig. 2 embodiment), 7:1-13 and 7:38-39 (discussing the similar Fig. 3 embodiment).

Mozer describes, for FIG. 3 and the audio recognition peripheral 300, that a “vector operation” is performed, where the “vector operation carried out by the vector processor may be a vector distance calculation.” *Mozer*, 7:7-13, 7:38-39. *Mozer* teaches that after the “vector operation” (i.e., the vector distance calculation) is performed, the vector processor may “transmit a result back to the external processor for use in the audio recognition algorithm.” *Mozer*, 7:10-13.

When the *Mozer* system of Figure 2 includes the audio recognition peripheral of Figure 3, the distance calculation is sent from the vector processor to the external CPU via bus 230. *Dec.*, 164. Thus, bus 230, at the least, is a “second interface” for sending a distance calculated by the vector processor 223,323. *Mozer*, Fig. 4.

Petitioners incorporate the motivation to combine *Mozer* with *Smyth* provided for Claim 1(a).

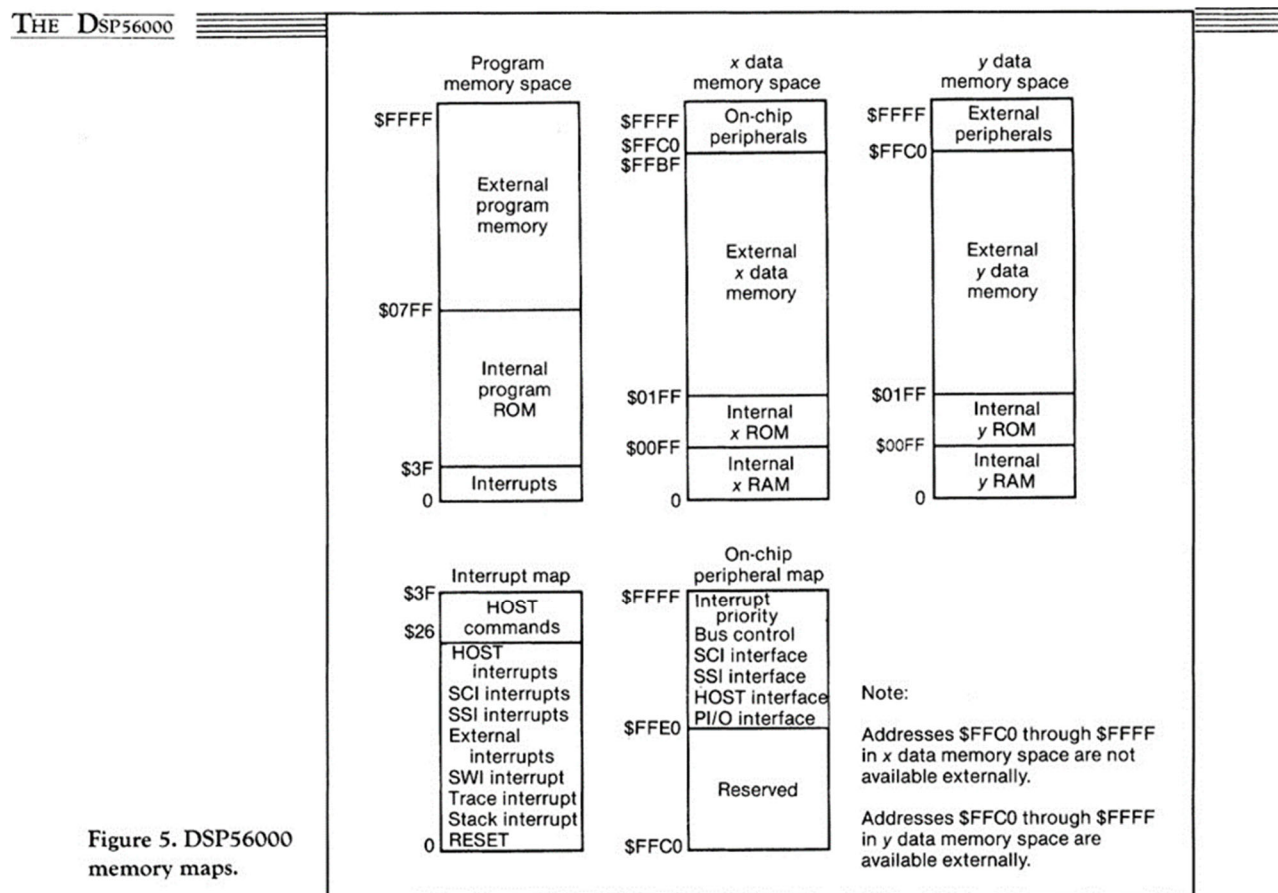
F. Claim 1(e)

Smyth alone or in combination with *Mozer* teaches or renders obvious Claim 1(e).

1. *Smyth’s Teachings*

As mapped for Claim 1(c) (Section V.A.4), *Smyth’s* classifier processor 341 teaches or renders obvious the claimed “calculating apparatus.” As mapped for Claim 1(b) (Section V.A.b), *Smyth* also teaches the claimed “acoustic model memory,” namely state memory 342. *Smyth*, 5:1-15, 5:55-57. *Smyth* teaches the classifier processor 341 may be a DSP, such as the Motorola DSP56000. *Smyth*, 5:49-53, 6:5-8. Because the Motorola DSP56000 provides for both a processor and memory to be fabricated thereon, *Smyth* teaches or renders obvious a single integrated circuit (IC) on which is fabricated both the classifier processor 341 and state memory 342.

The Motorola DSP56000 is discussed in detail in *Kloker* (EX1009), which discusses the Motorola DSP56000 combines “a core processor, RAM, ROM, peripheral interfaces, and a memory expansion interface on a single, 88-pin chip[.]” *Kloker* (EX1009), 1. Thus, a POSITA would have understood and found it obvious that a DSP such as the Motorola DSP56000 would have had both a calculating apparatus (e.g., processor) and memory (e.g., RAM, ROM) on a single chip (i.e., integrated circuit). *Dec.*, 165.



Kloker, 6 (“Figure 5. DSP56000 memory maps”).

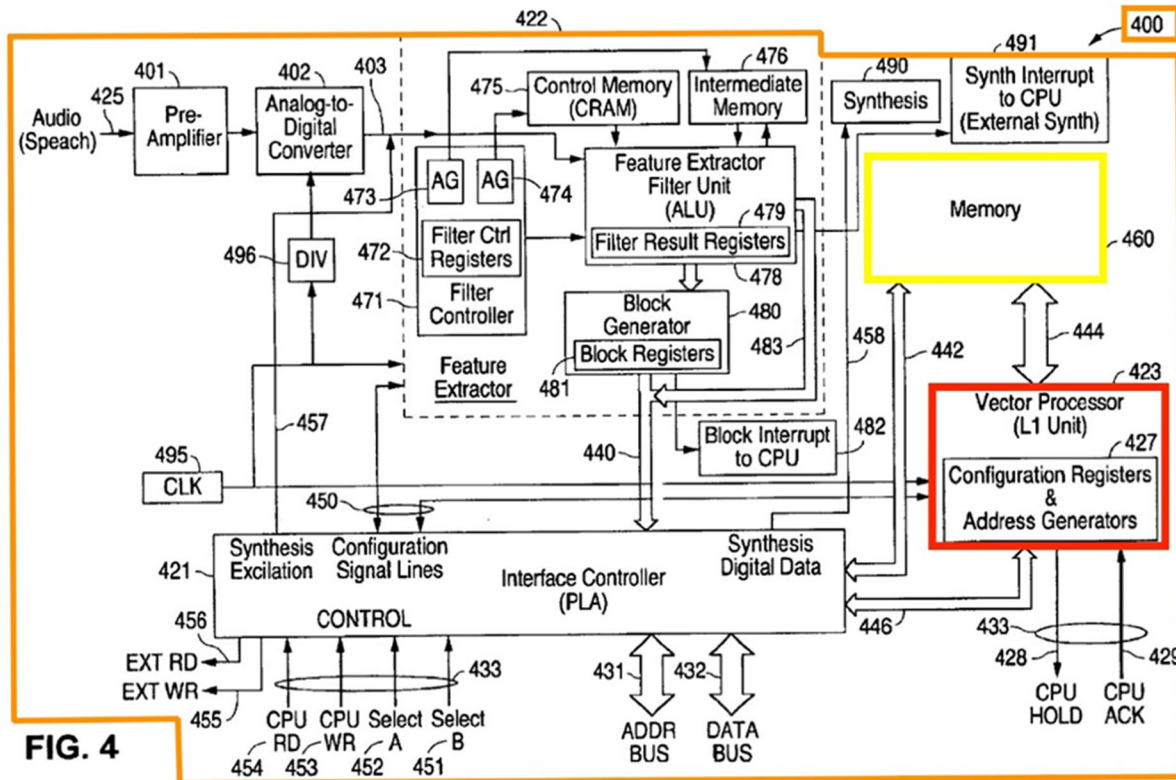
Because (1) *Smyth* teaches an exemplary DSP is the Motorola DSP56000 (*Smyth*, 5:49-53); (2) the classifier processor 341 may be implemented as the same model/type DSP as the feature extractor (*see Dec.*, 119-125); and (3) the DSP56000 provides for a single IC with a processor and a memory, *Smyth* teaches or renders obvious Claim 1(e).

2. *Mozer's Teachings*

To the extent Patent Owner contends *Smyth* does not teach or render obvious Claim 1(e), *Mozer* teaches a single integrated circuit on which a vector processor 423 (i.e., calculating apparatus) and memory 460 (i.e., acoustic model memory) are fabricated. *Mozer*, 7:53-58 (“Fig. 4 shows an example of an audio recognition peripheral integrated circuit 400 according to another embodiment of the present invention. Audio recognition peripheral 400 includes a pre-amplifier 401, an analog-to-digital converter 402, a feature extractor 422, an interface controller 421, a vector processor 423, and a memory 460.”).

Mozer discloses an “audio recognition peripheral 400 of FIG. 4 may also include[] vector processor 423 and/or memory 460.” *Mozer*, 9:22-24. Audio recognition peripheral integrated circuit 400 comprises a vector processor 423 that calculates the distance between a vector representation of an audio signal 425 (i.e., feature vector) and a second vector representing a template. *Mozer*, 9:22-26, 10:7-12, 7:38-39. *Mozer* teaches loading both first and second vectors into memory 460

and accessing both vectors to perform a distance calculation. *Mozer*, 10:17-21, 10:24-26, 10:1-16; *Dec.*, 166. *Mozer's* vector processor 423 (annotated red) and memory 460 (annotated yellow) are depicted in Fig. 4 below as fabricated on audio recognition peripheral integrated circuit 400 (annotated orange).



Mozer, Fig. 4 (annotated). Therefore, *Mozer* teaches a calculating apparatus for calculating distances and a memory for storing acoustic models fabricated on a single IC.

3. Motivation to Combine *Smyth* in View of *Mozer*

A POSITA would have found it obvious and been motivated to modify *Smyth* to fabricate classifier processor 341 and state memory 342 of classifier 34 on a single

integrated circuit according to *Mozer's* teachings. *Dec.*, 167-168. *Mozer* teaches providing a recognition peripheral system on one integrated circuit, including providing advantages. *Mozer*, 9:16-18 (“However, the embodiments described herein are illustrative of the ***advantages of providing a recognition peripheral system on one integrated circuit[.]***”). For example, *Mozer* teaches “embodiments of the present invention have the advantage of allowing the system designer to offload burdensome vector processing operations associated with audio recognition from the programmable processing unit to the audio recognition peripheral during the execution of the audio recognition algorithm.” *Mozer*, 10:46-52. *Smyth* already teaches the classifier performing the distance calculations is performed by a processor distinct from the processor for performing feature vector calculation and the processor for performing word identification. *Smyth*; 6:5-8; *Dec.*, 168. Therefore, a POSITA would have understood and found obvious that when modifying *Smyth* according to *Mozer's* teachings to fabricate a calculating apparatus and acoustic model memory on a single integrated circuit, the modified *Smyth* would have likewise captured such advantages. *Dec.*, 168. Such a use of a known technique of fabricating a memory and processor on a single integrated circuit would have improved *Smyth's* speech recognition system in the same way as *Mozer's* speech recognition system by “offload[ing] burdensome” processing to the “audio recognition peripheral” (i.e., *Smyth's* classifier 34). *Id.*

Smyth also already teaches a DSP for distance calculation, where the DSP includes the processor for performing the distance calculation and a memory. This is because *Smyth* teaches the DSP for classifier processor may be the Motorola DSP56000. *Smyth*, 6:1-4. *Smyth* also teaches separate DSPs or processors for the feature extractor and the word identification, as discussed above. Therefore, to the extent Patent Owner argues *Smyth*'s DSP for distance calculation is not a single IC, it would have been obvious to modify *Smyth* to form the classifier processor and associated memory on a single IC given *Smyth*'s teaching of a single DSP for the same components. *Dec.*, 167.

Additionally, a POSITA would have understood that such a modification would have had a reasonable expectation of success. *Dec.*, 168. *Smyth* and *Mozier* teach similar speech recognition systems with similar hardware. *Id.* *Smyth* already teaches a calculating apparatus and an acoustic model memory as part of recognition processor 3. *Smyth*, 4:22-31, 4:40-42, 5:1-12, Fig. 1. *Smyth* itself contemplates various hardware arrangements and modifications, including providing three separate DSPs or processors to perform specific tasks. *Smyth*, 5:49-53, 6:5-8, 6:47-51, 10:51-58; *Dec.*, 119-125. Therefore, a POSITA would have understood that modifying *Smyth* according to *Mozier*'s teachings would have been straightforward and would not have required undue experimentation. *Dec.*, 168. Also, it was well known in the prior art to fabricate a single integrated circuit with a calculating

apparatus and acoustic model memory. *Id.* (discussing *Toyoda*). Further, it was well known in the prior art to fabricate a single integrated circuit with multiple DSPs or processors and memories. *Id.* (discussing *Boike* and *Ozcelik*). Thus, a POSITA would have been motivated and found obvious that when modifying *Smyth*, to the extent not already expressly taught in *Smyth*, to fabricate classifier processor 341 (i.e., calculating apparatus) and state memory 342 (i.e., acoustic model memory) on a single integrated circuit, such a modification would have been advantageous and had a reasonable expectation of success. *Dec.*, 167-168.

G. Claim 10(Pre)

See Claim 1(pre).

H. Claim 10(a)

See Claim 1(a).

I. Claim 10(b)

See Claims 1(b)-1(c).

J. Claim 10(c)

See Claim 1(d).

K. Claim 11

See Claims 1(e), Claim 10.

L. Claim 14

See Claim 10. *Smyth* teaches a second interface for sending distances. *See* Claim 1(d). *Smyth* does not expressly teach said second interface is a serial bus.

Mozer teaches sending information between a programmable processing unit 110 and an audio recognition peripheral 120 over bus 130 and wherein “a serial bus may also be employed” for bus 130. *Mozer*, 4:17-28; *Dec.*, 174.

As discussed above, a POSITA would have been motivated and found it obvious to modify *Smyth* according to *Mozer*’s teachings. See Claim 1(e); *Dec.*, 175. Like *Smyth*, *Mozer* teaches dividing processing tasks amongst separate processing units. *Mozer* expressly teaches that this advantageously “drastically reduc[es] the resource loading on the programmable processing unit 110.” *Mozer*, 4:44-49. A POSITA would have understood and found obvious that such separate processing units work together as part of a speech recognition system, and as such said units transmit data between each other. *Mozer* expressly teaches performing such transmission via a serial bus. *Dec.*, 175. Further, a POSITA would have understood and found obvious that transmitting data in serial format using a serial bus from one DSP/processor to another DSP/processor was well known in the art and advantageous because a serial bus reduces trace complexity of a printed circuit board (e.g., IC) by using a single line instead of eight or sixteen parallel lines (copper traces) and thus requires less space on the circuit board. *Dec.*, 175. Therefore, such a modification would have had a reasonable expectation of success. *Dec.*, 175.

M. Claim 17

See Claim 10. *Smyth* teaches a calculating circuit that is configured to autonomously calculate distances for every acoustic state defined by the acoustic model. See Claim 1(b); *Dec.*, 158-159. *Smyth* teaches classifier processor 341 performs a distance calculation for every state field (where each state field contains a speech state) within memory 342. The “classification processor 34 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.” *Smyth*, 5:63-67. *Smyth* further teaches “[a]ccordingly, 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*, 6:1-4. Therefore, *Smyth* teaches “calculating the distances ...,” as claimed.

The classification processor calculates the probabilities *in response* to receiving a feature vector. Specifically, *Smyth* teaches the “classifier 34 receiving the succession of feature vectors and operating on each with a plurality of model states, to generate recognition results.” *Smyth*, 5:7-9; *Dec.*, 176.

Finally, a POSITA would have understood or at least found obvious that *Smyth*’s classifier processor 341 (and thus classifier 34 (i.e., acoustic coprocessor)) is configured to “*autonomously* calculate distances” because classifier processor 341

of classifier 34 is “arranged” to read each state field (and therefore, state stored in state field) and calculate distances for *each* state in response to receiving the feature vectors. *Jiang*, 5:7-9, 5:63-67; *Dec*, 176.

VI. GROUND 2: CLAIMS 2, 4, 6-9, 12, 16, 18, AND 20-21 ARE OBVIOUS OVER *SMYTH*, *MOZER*, AND *NGUYEN*

A. Claim 2

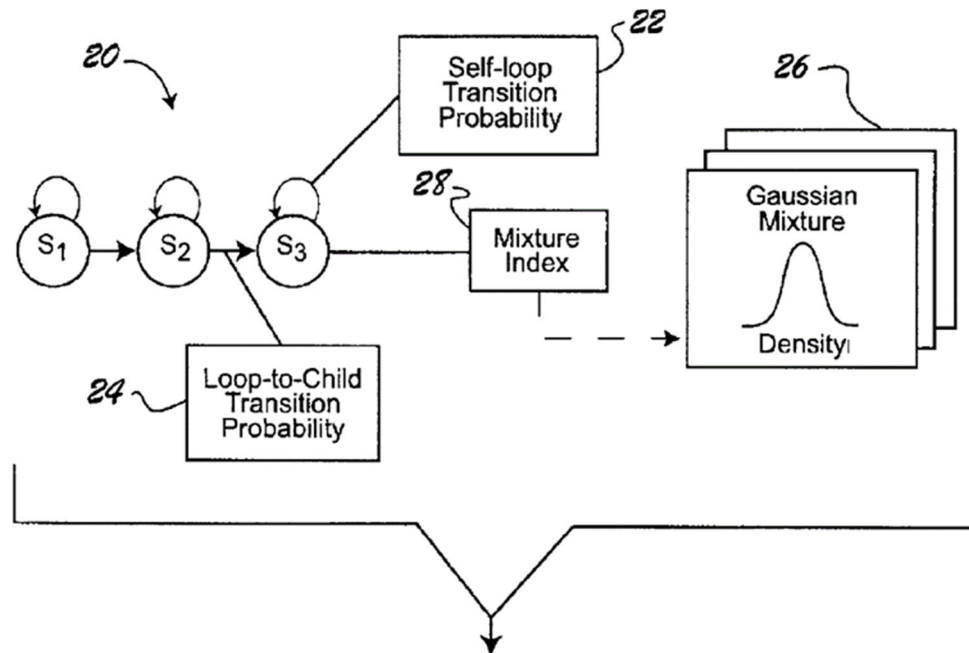
See Claim 1. *Smyth* discloses an acoustic model and distances that are probabilities. See Claims 1(b)-1(c). *Smyth* does not expressly teach the acoustic model contains Gaussian parameters. *Nguyen* teaches an acoustic model that contains Gaussian parameters and the distances are probabilities.

1. *Nguyen’s Teachings*

Nguyen teaches a speech recognition system a step where “***a similarity measure is computed*** between each frame of input speech and each reference pattern. The process defines ***a local measure of closeness*** between acoustic feature vectors[.]” *Nguyen*, 2:30-38. *Nguyen* teaches utilizing an acoustic model for which Gaussian data serves as the basis of said acoustic model. Specifically, *Nguyen* teaches performing a similarity measure for a group of vectors and in “the case of HMMs, the Gaussian distribution data or Gaussian mixture density data which serves as the basis for the acoustic model is loaded into memory.” *Nguyen*, 3:35-46. *Nguyen* teaches a “similarity measure can then be computed at step 36 for each acoustic feature vector in the first group of vectors. For example, a Gaussian

computation may be performed for each acoustic feature vector as is well known in the art.” *Nguyen*, 3:50-54.

Nguyen teaches each “Hidden Markov Model includes a collection of probabilities associated with the states themselves and transition amongst the states. Because probability values associated with each state may be more complex than a single value could represent, some systems will represent probability in terms of a Gaussian distribution.” *Nguyen*, 2:49-54. *Nguyen* further teaches that to “provide a more robust model, a mixture of Gaussian distributions may be used in a blended manner to represent probability values as shown diagrammatically at 26 and referenced by a mixture index pointer 28. Thus, associated with each state is a mixture index pointer which in turn identifies the Gaussian mixture density data for that state.” *Nguyen*, 2:54-60, 2:61-67. Fig. 1 depicts Gaussian distributions representing probability values “as shown diagrammatically at 26.” *Nguyen*, 2:55-58; *Dec.*, 177.



Nguyen, Fig. 1 (excerpt).

2. *Motivation to Combine Smyth and Nguyen*

A POSITA would have found it obvious and been motivated to modify *Smyth*'s acoustic model to represent probability in terms of a Gaussian distribution as taught by *Nguyen*. *Dec.*, 178. *Nguyen* provides express motivations, including “probability values associated with each state may be more complex than a single value could represent,” and to “provide a more robust model[.]” *Nguyen*, 2:49-67. A POSITA would have been motivated to modify *Smyth* to capture such advantages. *Dec.*, 178. Further, such a use of a known technique would have improved *Smyth*'s acoustic model in the same way as *Nguyen*'s. *Dec.*, 178. A POSITA would have had a reasonable expectation of success when modifying *Smyth*, as the modification

would have been straightforward and required only minor software or programming changes. *Dec.*, 178.

B. Claim 4

See Claim 2. *Smyth* teaches a second interface for sending distances. *See* Claim 1(d). *Smyth* does not expressly teach said second interface is a serial bus. *Mozer* teaches sending information between a programmable processing unit 110 and an audio recognition peripheral 120 over bus 130 and wherein “a serial bus may also be employed” for bus 130. *Mozer*, 4:17-28; *Dec.*, 179; *see* Claim 14.

C. Claim 6

See Claim 4. *Smyth* teaches a calculating circuit that is configured to autonomously calculate distances for every acoustic state defined by the acoustic model. *See* Claims 1(b)-1(c); *see also* Section V.M for Claim 17; *Dec.*, 180 (discussing classifier processor 341 is “arranged” to read and calculate distances).

D. Claim 7

See Claim 6. As discussed above, *Smyth*’s output probabilities meet the claimed distances. *See* Claims 1(c)-1(d). *Smyth* teaches an acoustic coprocessor and buffer memory (i.e., memory 353 of sequencer 35). *Smyth* does not expressly teach a result memory. *Nguyen* teaches a result memory. *Compare Nguyen* (EX1047), 4:1-7 (“resulting similarity measures may be stored in an output memory space” that is

“accessible to the processor performing the computations.”) with ’789 Patent, 6:9-23, 34:13-31, 32:32-35; *Dec.*, 181.

A POSITA would have been motivated and found it obvious to modify *Smyth-Mozer* to include one or more result memories per *Nguyen*’s teachings. *Dec.*, 182. Such a modification of adding a result memory to *Smyth*’s classifier 34 would have freed up the modified *Smyth*’s sequencer 35’s state probability memory 353 (i.e., buffer memory) so more distances would have been calculated and written to memory without having to wait for sequencer 35 to be ready to accept more results into state probability memory 353 (i.e., buffer memory). *Dec.*, 182. Such a use of a known technique would have advantageously improved *Smyth*’s speech recognition system’s processing efficiency in the same way as *Nguyen*’s. *Dec.*, 182. A POSITA would further have had a reasonable expectation of success when modifying *Smyth* because memories (including for storing distance calculation results) were well known in the art as taught by *Nguyen* and because modifying *Smyth* to include such a result memory would have been straightforward and would not have required undue experimentation. *Dec.*, 182.

E. Claim 8

Smyth-Mozer as modified by *Nguyen* teaches a result memory. *See* Claim 7. *Nguyen* (EX1047), 4:1-7 (“resulting similarity measures may be stored in an output memory space” that is “accessible to the processor performing the computations.”).

Nguyen teaches an embodiment where a similarity measure is computed at step 36 for *each* acoustic feature vector in a *first group* of vectors. *Nguyen*, 3:51-52, 3:61-4:7. After the similarity measure is computed, a search process is performed at step 44. *Nguyen*, 4:8-16. *Nguyen* teaches that “[c]ontemporaneous with the search process, **a subsequent group of acoustic feature vectors** may be retrieved into the memory workspace at step 48. **A similarity measure is computed for each acoustic feature vector in this subsequent group** as described above.” *Nguyen*, 4:17-22. Therefore, a POSITA would have understood that *Nguyen* teaches performing a similarity measure (i.e., distance calculation) while contemporaneously performing the search step. *Dec.*, 183.

A POSITA would have found it obvious and been motivated to modify *Smyth*’s speech recognition system to include a result memory. *See* Claim 7. Additionally, a POSITA would have been motivated and found it obvious to transmit results from said result memory to, for example, *Smyth*’s memory 353 of sequencer 35 via a second interface (*see* Claim 1(d)) for further speech recognition processing. *Dec.*, 183-184. Thus the modified *Smyth*’s result memory would have been configured to stored distances calculated from a second feature vector while distances calculated from a first feature vector would have been transmitted via said second interface. *Id.*

F. Claim 9(Pre)

See Claim 7. *Smyth* teaches a speech recognition system with an acoustic coprocessor and at least one processor (i.e., CPU). The '789 Patent specification describes using CPU and processor interchangeably. '789 *Patent* 30:1-2 (“We are also using the term ‘CPU’ and ‘processor’ interchangeably.”). *Smyth* teaches parsing processor 351 (i.e., processor). In one embodiment, *Smyth* expressly teaches “parsing processor 351 may, for example, be a microprocessor such as the Intel™ i-486™ microprocessor or the Motorola™ 68000 microprocessor, or may alternatively be a DSP device.” *Smyth*, 6:47-50. A POSITA would have understood or at least found obvious that parsing processor 351 implemented as a microprocessor would have met the claimed “CPU.” *Smyth*, 6:10-12, 6:36-51; *Dec.*, 185.

G. Claim 9(a)

Smyth teaches parsing³ processor 351 performs a speech recognition word search using the distances calculated by classifier 34 (comprising classifier processor

³ A POSITA would have understood *Smyth*’s disclosure of “parser processor 351” refers to “parsing processor 351.” *Dec.*, 186. Specifically, *Smyth* refers to element 351 as “parser” processor only once. *Smyth*, 6:54, 6:11-12, 6:45-51, 6:58-61, 6:65.

341 and state memory 342) in response to receiving a first feature vector. *Smyth* teaches parsing processor 351 calculates the most likely path of states to date over time, compares this path with stored state sequences, and ultimately identifies words and phrases. *Smyth*, 6:36-60. Specifically, *Smyth* teaches a word recognition process wherein parsing processor 351 reads, at each frame of speech, both the newest state probabilities (i.e., the claimed calculated distances) from classifier processor 341 and the historic state probabilities stored in state probability memory 353. *Smyth*, 6:36-39. Parsing processor 351 then calculates the most likely path of states to date and compares this most likely path with each of the known state sequences stored in state sequence memory 352. *Smyth*, 6:39-42. Each known state sequence in state sequence memory 352 corresponds to a word, phrase, or other speech sequence to be recognized. *Smyth*, 6:52-53. *Smyth* teaches parsing processor 351 outputs a probability score at each frame for each known state sequence. *Smyth*, 6:54-55. Then, when the end of an utterance is detected, *Smyth* teaches “a label signal indicating the most probable state sequence is output from the parsing processor 351 to the output port 38, to indicate that the corresponding name, word or phrase has been recognised.” *Smyth*, 6:56-61. Thus, *Smyth* teaches identifying words.

Therefore, a POSITA would have understood “parser processor 351” and “parsing processor 351” to be interchangeable with no difference in meaning. *Dec.*, 186.

Smyth teaches parsing processor 351 (i.e., CPU) executes software to perform or initiate a speech recognition word search. *Smyth* teaches parsing processor 351 is “arranged to” perform the word identification. *Smyth*, 5:12-15 (“a sequencer 35 which is arranged to receive the classification results form [sic] the classifier 34 and to determine the predetermined utterance to which the sequence of classifier output indicates the greatest similarity[.]”), 6:36-42 (“The parsing processor 351 is arranged to read, at each frame, the state probabilities output by the classifier processor 341, and the previous stored state probabilities in the state probability memory 353, and to calculate the most likely path of states over time, and to compare this with each of the state sequences stored in the state sequence memory 352.”). Further, in one embodiment, *Smyth* teaches “parsing processor 351...may be a DSP device” or a microprocessor. *Smyth*, 6:47-50. A POSITA would have understood that *Smyth*’s parsing processor 351 implemented as a DSP or microprocessor would have performed or initiated a speech recognition word search, as claimed. *Dec.*, 187.

H. Claim 9(b)

Smyth discloses a speech recognition system that identifies spoken words and calculates the claimed distances. *See* Claim 9(a). *Smyth* does not expressly teach the time period overlapping for distance calculation and word identification. As discussed above in Claims 1(c), 9(a), *Smyth* does teach calculating distances in classifier processor 341 and identifying spoken words in parsing processor 351.

1. *Nguyen's Teachings*

Nguyen teaches a speech recognition system where at step 12, “**a similarity measure is computed** between each frame of input speech and each reference pattern. The process defines **a local measure of closeness** between acoustic feature vectors[.]” *Nguyen*, 2:30-38. After a similarity measure is computed, a decoding step 16 occurs, wherein the “probability that a particular phoneme was spoken is provided by the acoustic models as part of the pattern matching process. A sequence of words can then be constructed by concatenating the phonemes observed during the pattern matching process.” *Nguyen*, 3:11-24. Fig. 1 depicts *Nguyen's* pattern classification (i.e., distance calculation) step 12 and decoding step 16.

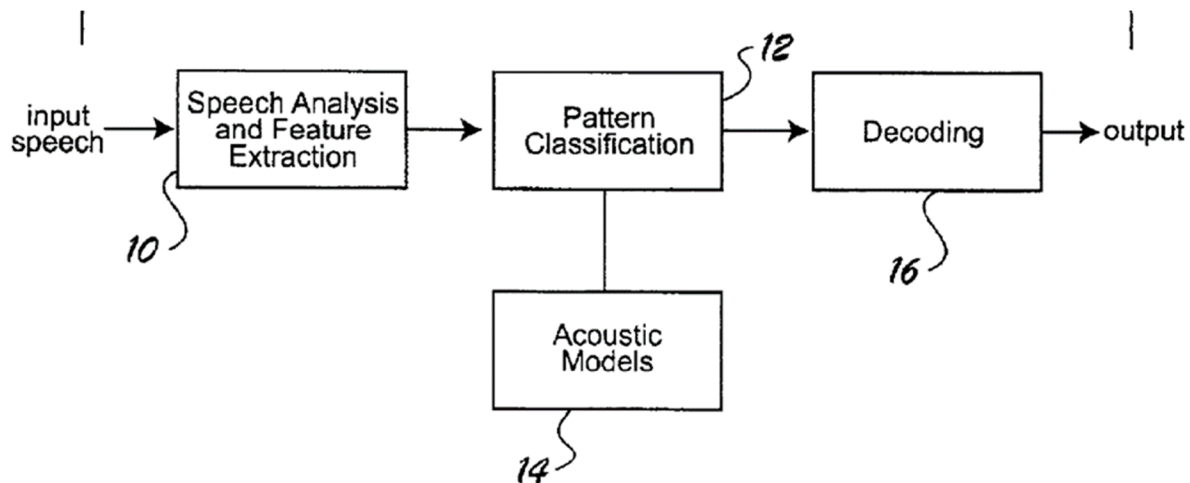


FIG. 1

Nguyen, Fig. 1 (excerpt).

Nguyen further teaches an embodiment where a similarity measure is computed at step 36 for *each* acoustic feature vector in a first group of vectors.

Nguyen, 3:51-52, 3:61-4:7. After the similarity measure is computed, a search process is performed at step 44. *Nguyen*, 4:8-16. *Nguyen* teaches that “[c]ontemporaneous with the search process, **a subsequent group of acoustic feature vectors** may be retrieved into the memory workspace at step 48. **A similarity measure is computed for each acoustic feature vector in this subsequent group** as described above.” *Nguyen*, 4:17-22. Therefore, a POSITA would have understood that *Nguyen* teaches performing a similarity measure (i.e., distance calculation) while contemporaneously performing the search step. *Dec.*, 189-193.

2. Motivation to Combine Smyth and Nguyen

A POSITA would have found it obvious and been motivated to modify the *Smyth-Mozer* speech recognition system to operate using the contemporaneous processing taught in *Nguyen*. *Dec.*, 193. *Smyth* itself teaches an embodiment with one DSP for performing distance calculations (i.e., classifier processor 341 implemented as a DSP) and another processor/DSP for performing speech recognition (i.e., parsing processor 351). See Claims 1(c), 9(Pre)-9(a). Modifying *Smyth* according to *Nguyen*’s teachings would have advantageously resulted in *Smyth*’s separate processors/DSPs processing contemporaneously, thus improving system performance. *Dec.*, 193. Indeed, a POSITA would have understood “contemporaneous” processing was well known in the prior art. *Dec.*, 194.

A POSITA would have been motivated and found it obvious to combine *Smyth*'s speech recognition architecture that uses multiple, separate DSPs with *Nguyen*'s group feature vector processing and *Nguyen*'s concurrent distance calculation and word parsing with a reasonable likelihood of success. *Dec.*, 195. A POSITA would have understood that *Smyth* already teaches utilizing multiple DSPs, such that modifying *Smyth* to utilize *Nguyen*'s contemporaneous processing functionality would have had a reasonable expectation of success and would have required only minor modifications. *Dec.*, 196. Further, because such contemporaneous processing was already well known in the prior art, as taught by *Nguyen*, such a combination would have had a reasonable expectation of success. *Dec.*, 195.

I. Claim 12

See Claims 2, 10.

J. Claim 16

1. Claim 16(Pre)

See Claims 9(Pre), 10.

2. Claim 16(a)

See Claim 9(a).

3. Claim 16(b)

See Claim 9(b).

K. Claim 18

See Claims 10, 17, 7. As discussed above, *Smyth*'s output probabilities meet the claimed distances. *See* Claims 1(c)-1(d). As discussed for Claim 7 above, *Smyth* as modified by *Nguyen* teaches a result memory. *See* Claim 7; *Dec.*, 201.

L. Claim 20

See Claims 2, 18.

M. Claim 21

Smyth in view of *Mozer* teaches an acoustic coprocessor using a neural network to calculate at least one of the distances. Specifically, *Mozer* teaches utilizing a neural network and that such was "well known by those skilled in the art." *Mozer*, 10:35-52, 5:53-60; *Dec.*, 203.

A POSITA would have been motivated and found it obvious to modify *Smyth* such that *Smyth*'s acoustic model would have used a neural network to calculate at least one of the distances, as taught by *Mozer*. *Dec.*, 204. Such a modification would have advantageously used a known technique (performing a distance calculation using a neural network) to improve *Smyth*'s similar speech recognition system in the same way as *Mozer*'s. A POSITA would have understood that such a modification would have had a reasonable expectation of success because using neural networks for distance calculations was well known in the art, including as taught by *Mozer*.

Further, using both neural networks and HMMs in a single speech recognition system was well known in the art. *Id.* (discussing *Jelinek*).

VII. GROUND 3: CLAIM 13 IS OBVIOUS OVER *SMYTH*, *MOZER*, AND *HOUSER*

A. Claim 13

See Claim 10. *Smyth* teaches an acoustic model memory. *See* Claim 1(b). *Houser* teaches a speech recognition system including “flash memory 307” for storing “phonemic or template vocabulary data[.]” *Houser*, 15:33, 15:46-48. A POSITA would have been motivated to modify *Smyth-Mozer* to utilize FLASH memory for state memory 342 (i.e., acoustic model memory) because FLASH memory was known to be advantageously non-volatile and fast. *Dec.*, 205. A POSITA would have had a reasonable expectation of success implementing *Smyth*’s state memory 342 using FLASH memory because such memory was well known and widely used in the prior art. *Dec.*, 205.

VIII. GROUND 4: CLAIMS 22-24 ARE OBVIOUS OVER *SMYTH*, *MOZER*, *NGUYEN*, AND *VENSKO*

A. Claim 22

See Claim 21. *Vensko* teaches template processors (i.e., acoustic coprocessors) 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*, 6:22-29. *Vensko* further teaches said template processors

“generate an interrupt to the master processor 160 informing it that the results are ready for transfer.” *Vensko*, 6:29-33, 10:10-19; *Dec.*, 206.

A POSITA would have been motivated and found obvious to modify *Smyth*’s classifier 34 (*see* Claim 1(Pre) for acoustic coprocessor mapping) to be capable of generating an interrupt in response to completion of calculation of a set of distances per *Vensko*’s teachings. *Dec.*, 207. Such a modification would have advantageously allowed *Smyth*’s acoustic coprocessor to be capable of generating an interrupt to inform sequencer 35 that calculation of a set of distances is completed and would have been a more efficient way to get data to sequencer 35. *Dec.*, 207. Furthermore, such a modification to *Smyth* would have been straightforward, requiring only minor software programming changes, and generating interrupts was well known in the art, as taught by *Vensko*. Therefore, such a modification would have had a reasonable expectation of success. *Dec.*, 207.

B. Claim 23

See Claim 22. *Mozer* teaches an acoustic model memory with a plurality of storage locations within an SRAM, as claimed. *Mozer* teaches loading both a “first vector representation of the audio signal 425” (i.e., feature vector) and a “second vector representing a template into memory 460.” *Mozer*, 10:17-20. *Mozer* expressly teaches memory 460 may be an SRAM. *Mozer*, 10:20-21 (“[T]he memory 460 may be an 4K SRAM, for example.”), 9:22-28. A POSITA would have understood

Mozer's 2K and 4K SRAMs would have each had a plurality of storage locations. *Dec.*, 208.

A POSITA would have been motivated and found it obvious to modify *Smyth* to implement state memory 342 using an SRAM. *Dec.*, 209. A POSITA would have understood that SRAM was fast and used low power; thus, SRAM would have advantageously allowed *Smyth's* acoustic model information to be accessed quickly and with low power consumption. *Id.* SRAM was also well known in the prior art. Modifying *Smyth's* acoustic model memory to utilize SRAM would have been a simple substitution of one known element for another to obtain predictable results. Further, such a use of a known technique would have improved *Smyth's* similar speech recognition system in the same way as *Mozer* by providing fast memory access and low power consumption. *Mozer*, 9:22-28, 10:17-21. Furthermore, a POSITA would have understood and found obvious that such a modification would have had a reasonable expectation of success because SRAM was well known in the prior art, as taught by *Mozer*, and because modifying *Smyth* to utilize SRAM memory would have been a straightforward well within the skillset of a POSITA and would not have required undue experimentation. *Dec.*, 209.

C. Claim 24

Smyth in view of *Mozer* teaches SRAM. *See* Claim 23. *Smyth* in view of *Mozer* teaches the SRAM memory 460 and vector processor 423 are fabricated on a single

integrated circuit, namely audio recognition peripheral integrated circuit 400. *See* Claim 1(e) (including a motivation to combine *Smyth* in view of *Mozer*); *Dec.*, 210.

IX. GROUND 5: CLAIMS 25-28 AND 37 ARE OBVIOUS OVER *SMYTH*, *MOZER*, *NGUYEN*, *VENSKO*, AND *COMEAU*

A. Claim 25

See Claim 22. The *Smyth* combination as modified by *Comeau* teaches or renders obvious an acoustic processor capable of setting one or more bits in a status register to values that indicate the completion of calculation of a set of distances. *Smyth* teaches an acoustic coprocessor. *See* Claim 1(Pre). *Smyth* in view of *Comeau* teaches a status register and setting bits that indicate the completion of a task. *Comeau* teaches a coprocessor mailbox/FIFO status register 102 that sets a bit to notify a host core processor 40 that the register 102 is empty, full, or not empty. *Comeau*, [0153]-[0161], [0151]-[0152], [0168], [0181]-[0182]. A POSITA would have understood and found obvious that status register 102 would have been a status register, as claimed. *Dec.*, 211-214.

TABLE 5

<u>Coprocessor Mailbox/FIFO Status Register</u>				
Bits	Name	Access	Reset State	Description
15:12	—	R	0	Reserved
11	c_hfifo_empty	R	1	Host FIFO 76 is empty

TABLE 5-continued

<u>Coprocessor Mailbox/FIFO Status Register</u>				
Bits	Name	Access	Reset State	Description
10	c_hfifo_full	R	0	Host FIFO 76 is full. A write while this bit is asserted causes an overflow error (i.e., keep writing FIFO until this bit asserts. Can also be configured for an interrupt).
9	c_hfifo_uflow	R/W	0	Asserted when the host performs a read on an empty host FIFO 76 (raises FIFO error interrupt) . Cleared by writing a 1 to this bit.
8	c_hfifo_oflow	R/W	0	Asserted when the coprocessor processor writes to the host FIFO 76 when the host FIFO 76 is full. Raises FIFO error interrupt. Cleared by a 1 written to this bit.

Comeau, [0342].

A POSITA would have been motivated and found it obvious to modify the *Smyth* combination to implement a status register per *Comeau*'s teachings. *Dec.*, 215. Such a modification would have advantageously allowed the modified *Smyth* to provide a "flexible mechanism for status information to be shared with the host [processor] 40." *Comeau*, [0201]; *Dec.*, 215. Such a use of a known technique of employing a status register would have advantageously improved the modified *Smyth*'s speech recognition system in the same way as *Comeau*'s system. A POSITA would have had a reasonable expectation of success when making such a modification because status registers were well known in the prior art and

implementing a status register in the *Smyth* combination would have been straightforward and would not have required undue experimentation. *Dec.*, 216. Specifically, *Comeau* teaches a “speech recognition manager to synchronize the management of the resources between the host processor and the coprocessor. Each of these resource manager/observables comprises a software component on the coprocessor and a peer component on the host.” *Comeau*, [0274].

B. Claim 26

See Claim 25. *Smyth-Nguyen* teaches a result memory. See Claim 7. The modified *Smyth* result memory would have stored distances to be copied to state probability memory 353 (i.e., buffer memory). *Dec.*, 217. A POSITA would have understood or at least found obvious that *Smyth*’s memory 353 was a buffer memory, as claimed, because memory 353 receives and stores state probabilities output by classifier processor 341 to use said probabilities to perform word identification. *Smyth*, 6:10-19, 6:36-42, Figs. 3-4; compare to ’789 Patent, 6:9-23, 34:13-31.

Smyth does not expressly teach a DMA controller. *Vensko* teaches a DMA controller. *Vensko* teaches a continuous speech recognition system with “a plurality of processors doing template comparisons of speech data. Each processor has an associated memory shared with the other processors by direct memory access (DMA) through a shared data bus.” *Vensko*, Abstract, 7:48-53. *Vensko* further teaches “DMA transfers may be used to transfer character distance vector data from

end processor 110 to each of the template processors 130, 140 ... 150; to transfer template data among the template processors 130, 140, ... 150; or to transfer template processor results data to the master processor.” *Vensko*, 10:3-9, 10:19-31, Figs. 4, 5, 7; *Dec.*, 218.

A POSITA would have been motivated and found it obvious to modify *Smyth* to include a DMA controller in *Smyth*’s recognition processor 3 for copying distances from a result memory to a buffer memory. Indeed, it was well known in the art that DMA controllers were used to access and copy data from one memory to another, with *Vensko* as one example. *Dec.*, 219. Such a use of a known technique would have improved the modified *Smyth*’s system by similarly including a DMA controller capable of transferring data from one memory to another, which would have reduced load on the processors/DSPs. *Dec.*, 219. A POSITA would further have had a reasonable expectation of success when modifying *Smyth* because DMA controllers were well known in the art and because modifying *Smyth* to include such a DMA controller would have been straightforward and would not have required undue experimentation. *Dec.*, 219.

C. Claim 27

See Claim 26. *Vensko* teaches a DMA controller, as mapped for Claim 26. A POSITA would have understood that a DMA controller would have been capable of copying distances in a plurality of sequential bursts. *Dec.*, 220. Mr. Schmandt opines

a POSITA would have understood and found obvious that doing DMA as big transfers or small bursts is entirely a question of software programming. A DMA controller is loaded with a source address, a destination address, and the number of bytes to transfer. *Dec.*, 221-223 (discussing *Carmon*, *Wilcox*, and *Pomerantz* and that a DMA controller can transfer data in various sizes depending on programming). The DMA controller then transfers those bytes. Based on the DMA controller's programming, the DMA controller could transfer, for example, 500 bytes at once or transfer 500 bytes in 10 transfers of 50 bytes. A POSITA would have understood and found obvious that choosing the sizes of the number of bytes to transfer is simply a design decision based on observed bus traffic. *Dec.*, 224. Therefore, *Smyth* as modified by *Vensko*'s DMA controller teaches or at least renders obvious that a DMA controller is capable of copying distances in a plurality of sequential bursts. Further, *Vensko* teaches DMA transfers. *Vensko*, 11:36-43, 9:56-10:31.

As discussed above, a POSITA would have been motivated and found it obvious to modify *Smyth* according to *Vensko*'s teachings and would have had a reasonable expectation of success when performing such modification. *See* Claim 26; *Dec.*, 224.

D. Claim 28(Pre)

See Claims 26, 1(Pre), 10(Pre).

E. Claim 28(a)

See Claims 16(Pre), 16(a).

F. Claim 28(b)

See Claim 16(b).

G. Claim 37

See Claim 27. *Smyth* teaches a speech recognition system with a microphone. *Smyth*, 1:24-32, 4:22-24 (“a telecommunications system including speech recognition generally comprises a microphone 1, typically forming part of a telephone handset[.]”), 4:38-40 (“In operation, a speaker speaks into the microphone 1 and an analog speech signal is transmitted form [sic] the microphone 1 into the network 2 to the recognition processor 3[.]”), 4:46-67, Fig. 1; *Dec.*, 228.

X. GROUND 6: CLAIM 29 IS OBVIOUS OVER *SMYTH*, *MOZER*, *NGUYEN*, *VENSKO*, *COMEAU*, AND *BOIKE*

A. Claim 29

See Claim 28. *Smyth* teaches classifier 34 (i.e., acoustic coprocessor) comprises classifier processor 341 and sequencer 35 comprises parsing processor 351 (i.e., at least one processor). *Smyth*, 5:55-57, 6:13-19. *Smyth* teaches classifier processor 341 may be a DSP and parsing processor 351 may be a DSP or microprocessor. *Smyth*, 6:5-8, 6:45-51. *Smyth* does not expressly teach the acoustic coprocessor and the at least one processor are fabricated on a single integrated circuit. *Boike* teaches a “system-on-a-chip (SoC) integrated circuit 10 containing

multiple digital signal processors (DSPs).” *Boike* (Exhibit 1006), 2:52-54. *Boike* depicts such DSP subsystems 18-21 as part of DSP/SoC 10 (i.e., integrated circuit 10). “FIG. 1 is a block diagram depicting a typical configuration of a DSP/SoC multiple digital signal processor integrated circuit.” *Boike*, 2:42-44. Annotated Fig. 1 below depicts a single integrated circuit 10 (annotated blue) and DSP subsystems 18-21 (annotated green).

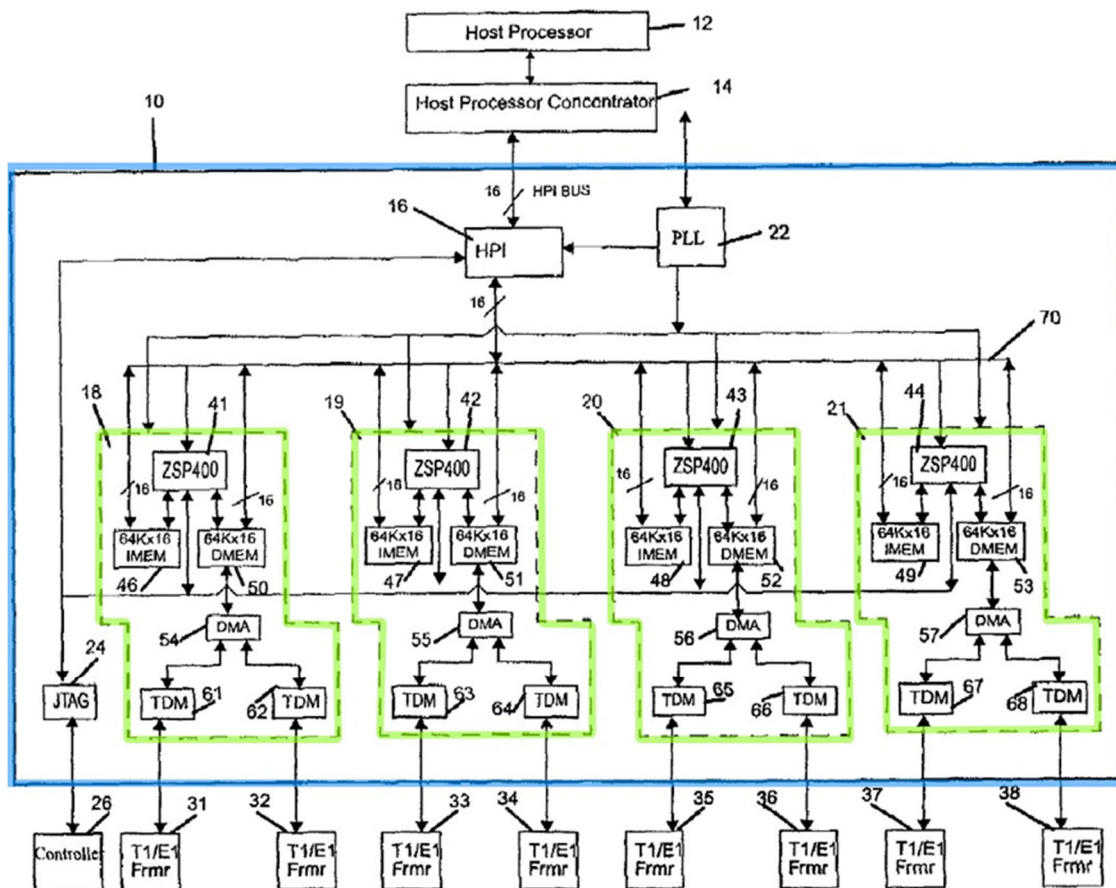


FIGURE 1

Boike, Fig. 1 (annotated). *Boike* teaches numerous advantages of an architecture placing multiple DSPs on a single IC, including reduced space and costs and conserved power. *Boike*, 1:19-42, 2:1-9, 2:21-35; *Dec.*, 230.

A POSITA would have found it obvious and been motivated to modify *Smyth* to include the three distinct processing circuits on a single IC, as taught by *Boike*. *Boike* alone expressly motivates the combination, informing a POSITA that it would have reduced space and costs and conserved power to form multiple DSPs on a single IC. *Dec.*, 231 (citing *Boike*, 1:19-42).

Smyth's speech recognition system already comprises three separate distinct DSPs or a combination of DSPs and a microprocessor for performing the respective functionality of feature vector extraction, distance calculation (i.e., computing probability scores of phonemes in an HMM), and word search identification. *Dec.*, 119-125. *Smyth* refers to one of its DSPs being the Motorola DSP56000. *Dec.*, 119. Motorola issued a spec and feature sheet extolling the benefits of using DSPs, including for speech recognition. *Dec.*, 126, 229, 72 (citing EX1009 *Kloker*, "DSP advantages"). *Kloker* further teaches the DSP56000 "can be used as a very fast microprocessor in any application that requires high-speed calculations or real-time response." *Kloker*, 1. Mr. Schmandt further notes using DSPs for speech recognition allowed for consistency, flexibility, and higher integration. *Dec.*, 229 (citing *Mariani* (EX1013), 9-10, which also refers to the Motorola DSP56000). Therefore, a

POSITA would have understood that DSPs for speech recognition were routinely employed and commonly understood to be beneficial for speech recognition. *Dec.*, 232. Thus, modifying *Smyth* according to *Boike's* teachings would have advantageously improved *Smyth's* speech recognition system. *Id.* A POSITA also would have had a reasonable expectation of making such a modification in light of *Boike's* teachings. *Dec.*, 233.

XI. GROUND 7: CLAIM 35 IS OBVIOUS OVER *SMYTH*, *MOZER*, *NGUYEN*, *VENSKO*, *COMEAU*, AND *WHITTAKER*

A. Claim 35

See Claim 26. The '789 Patent describes integers outside of Claim 35 only at 18:14-16 ("Each lexical tree comprises lexical tree elements, and each element has three fields. For now, the fields will be integers and will be converted to `std_logic_vector` later[.]"), 18:20-50; *Dec.*, 234. Mr. Schmandt opines *Whittaker* (EX1041) teaches converting an array of integers into a tree structure in the same way as described in the '789 Patent. *Dec.*, 235.

Whittaker teaches a "method compresses one or more ordered arrays of integer values. The integer values can represent a vocabulary of a language mode, in the form of an N-gram, of an automated speech recognition system." *Whittaker* (EX1041), Abstract, 2:20-28, 2:33-44, 3:8-15, 3:27-32 ("the method for compressing according to the invention converts each ordered array of integers into a tree structure[.]"); *Dec.*, 235.

A POSITA would have been motivated and found it obvious to modify the *Smyth* combination to utilize an acoustic model including parameters represented with integer data, as taught by *Whittaker, Dec.*, 236. Such a modification would have advantageously “obtain[ed] an 86% reduction in the size of the language model with no increase in the word error rate,” per *Whittaker’s* teachings. *Whittaker*, 2:30-32; *Dec.*, 236. Such a use of a known technique would have advantageously improved the modified *Smyth’s* speech recognition system in the same way as *Whittaker’s*. *Dec.*, 236. Further, a POSITA would have had a reasonable expectation of success making such a modification because representing acoustic model parameters with integer data was well known in the art. *Dec.*, 236. Such a modification would have been straightforward and would have required only minor changes in software or programming. *Dec.*, 236.

XII. GROUND 8: CLAIMS 38-39 ARE OBVIOUS OVER *SMYTH, MOZER, NGUYEN, VENSKO, COMEAU, AND ICHIKAWA*

A. Claim 38

See Claim 27. *Ichikawa* teaches a speech recognition system with a voice input microphone array constituted of a plurality of microphones. *Ichikawa*, 1:8-10 (“The present invention relates to a speech recognition system, especially a method for eliminating noise by using a microphone array.”), 1:28-29 (“The voice input part 181 is a microphone array constituted of a plurality of microphones.”), 1:40-42, 3:63-66, 8:66-67, 20:65-67, 22:11-22, Figs. 17, 18; *Dec.*, 237.

A POSITA would have been motivated and found it obvious to modify *Smyth* to use a plurality of microphones to advantageously eliminate noise, per *Ichikawa's* teachings. *Dec.*, 238. Use of such a known technique would have advantageously improved *Smyth's* speech recognition system in the same way as *Ichikawa's* system. *Dec.*, 238. Further, a POSITA would have had a reasonable expectation of success because using a plurality of microphones in an array was known in the prior art and would have required only minor hardware modifications. *Dec.*, 238.

B. Claim 39

See Claims 26, 38.

XIII. GROUND 9: CLAIMS 42-44 ARE OBVIOUS OVER *SMYTH, MOZER, NGUYEN, VENSKO, COMEAU, ICHIKAWA, AND SESHADRI*

A. Claim 42

See Claim 39. *Seshadri* teaches a “speech recognition system of the invention can be configured on at least but not limited to a laptop computer[.]” *Seshadri*, 3:60-65; *Dec.*, 241.

A POSITA would have been motivated and found it obvious to modify *Smyth* to implement *Smyth's* speech recognition system using a laptop computer as taught by *Seshadri*. *Dec.*, 240, 242. Such a use a known technique would have advantageously improved *Smyth's* speech recognition system in the same way as *Seshadri's* system. *Dec.*, 242. Further, a POSITA would have had a reasonable expectation of success because implementing speech recognition on a laptop

computer was known in the prior art and would have required only minor hardware modifications. *Dec.*, 242.

B. Claim 43

See Claim 39. *Smyth* does not expressly teach a mobile phone comprising the speech recognition system of claim 39. *Seshadri* teaches implementing speech recognition system on a mobile phone. *Seshadri*, 3:22-24 (“the user can access the mobile phone 100 and can activate the speech recognition system of the present invention.”), 3:28-31 (“the user can be alerted to commence speaking into the mobile phone 100. The user, for example, can speak and request to send an e-mail message[.]”), 3:34-36 (“The speech recognition system processes the audio signal input and can convert all part of the audio signal input into recognizable information[.]”), 3:55-57 (“wherein the mobile phone 100 can be configured with the speech recognition system of the present invention[.]”); *Dec.*, 244.

A POSITA would have been motivated and found it obvious to modify *Smyth* to implement *Smyth*’s speech recognition system using a mobile phone as taught by *Seshadri*. *Dec.*, 243, 245. Such a use a known technique would have advantageously improved *Smyth*’s speech recognition system in the same way as *Seshadri*’s system. *Dec.*, 243. Further, a POSITA would have had a reasonable expectation of success because implementing speech recognition on a mobile phone was known in the prior art and would have required only minor hardware modifications. *Dec.*, 245.

C. Claim 44

See Claim 39. *Smyth* does not expressly teach an electronic entertainment product comprising the speech recognition system of claim 39. *Seshadri* teaches “the speech recognition system using audio signals” “can be configured on a variety of electronic device[s], either mobile or stationary. For instance, the improved speech recognition system of the invention can be configured on at least but not limited to a laptop computer, a PDA, an audio/video recording device, a home computer, a game console, a remote controller, or other comparable device.” *Seshadri*, 3:54-65. A POSITA would have understood that a home computer, game console, audio/video recording device, and/or PDA would all have been an “electronic entertainment product,” as claimed. *Dec.*, 244-246.

A POSITA would have been motivated and found it obvious to modify *Smyth* to implement *Smyth*’s speech recognition system using an electronic entertainment product as taught by *Seshadri*. *Dec.*, 244-246. Such a use a known technique would have advantageously improved *Smyth*’s speech recognition system in the same way as *Seshadri*’s system. *Dec.*, 244-246. Further, a POSITA would have had a reasonable expectation of success because implementing speech recognition on an electronic entertainment product was known in the prior art and would have required only minor hardware modifications. *Dec.*, 244-246.

XIV. GROUND 10: CLAIM 45 IS OBVIOUS OVER SMYTH, MOZER, NGUYEN, VENSKO, COMEAU, ICHIKAWA, AND STEINBISS

A. Claim 45

See Claim 39. *Steinbiss* teaches an appliance comprising a speech recognition system:

Other areas of application include the control of electronic consumer appliances such as television sets, video recorders, hi-fi music systems, kitchen appliances or other appliances with speech control, in which the various appliance functions can be acted upon by means of speech, e.g. volume control, the choice of a television channel or radio station or the adjusting of the motor speed of a kitchen appliance.

Steinbiss, [0001]-[0002].

A POSITA would have been motivated and found it obvious to modify *Smyth* to implement *Smyth*'s speech recognition system using an appliance as taught by *Steinbiss*. *Dec.*, 247-250. Such a use a known technique would have advantageously improved *Smyth*'s speech recognition system in the same way as *Steinbiss*'s system. *Dec.*, 247-250. Further, a POSITA would have had a reasonable expectation of success because implementing speech recognition on an appliance was known in the prior art and would have required only minor hardware modifications. *Dec.*, 247-250.

XV. 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 ’789 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 (EX1068), 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. (IPR2023-0036, Paper 10; Factor 6).

§325(d). Petitioners seek to join already instituted IPR2023-00036, 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.

XVI. CONCLUSION

For the forgoing reasons, Petitioners respectfully request *inter partes* review of the Challenged Claims of the '789 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

XVII. 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 '789 Patent is currently the subject of *Apple Inc. v. Zentian Ltd.*, Inter Partes Review No. IPR2023-00036 (P.T.A.B.) (instituted Jun. 12, 2023). This Petition is substantively identical to the petition submitted in IPR2023-00036 and Petitioners have concurrently filed a motion to join that proceeding.

The '789 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., et al.*, No. 6:22-cv-00123 (W.D. Tex. Feb. 2, 2022).

C. Lead and Back-Up Counsel

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

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

Petition for *Inter Partes* Review of
U.S. Patent No. 10,839,789

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

D. Service Information

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

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

CLAIM LISTING APPENDIX

U.S. Patent No. 10,839,789 for Claims 1-2, 4, 6-14, 16-18, 20-29, 35, 37-39, 42-45

Claim Designation	Claim Language
Claim 1(Pre)	An acoustic coprocessor for processing data associated with an audio signal, comprising:
Claim 1(a)	first interface for receiving at least one feature vector, wherein the feature vector is determined from the audio signal;
Claim 1(b)	an acoustic model memory for storing an acoustic model defining a plurality of acoustic states;
Claim 1(c)	a calculating apparatus for calculating distances indicating a similarity between the at least one feature vector and respective acoustic states of the acoustic model read from the acoustic model memory; and
Claim 1(d)	a second interface for sending at least one distance calculated by the calculating apparatus;
Claim 1(e)	wherein the calculating apparatus and the acoustic model memory are fabricated on a single integrated circuit.
Claim 2	The acoustic coprocessor of claim 1, wherein the acoustic model contains Gaussian parameters and the distances are probabilities.
Claim 4	The acoustic coprocessor of claim 2, wherein the second interface for sending distances is a serial bus.
Claim 6	The acoustic coprocessor of claim 4, wherein the acoustic coprocessor, in response to receiving a feature 15 vector, autonomously calculates the distance for an associated acoustic state of the acoustic model.
Claim 7	The acoustic coprocessor of claim 6, further comprising a result memory, wherein the distance for an associated acoustic state of the acoustic model is stored in the result memory.

Claim Designation	Claim Language
Claim 8	The acoustic coprocessor of claim 7, wherein the result memory is configured so that distances calculated from a first feature vector are sent over the second interface for sending while at least one distance calculated from a second feature vector is being stored in the one or more result memories.
Claim 9 (Pre)	A speech recognition system comprising the acoustic coprocessor of claim 7, the speech recognition system further comprising at least one CPU, wherein:
Claim 9 (a)	the CPU executes software to perform or initiate a speech recognition word search using the distances calculated by the acoustic coprocessor in response to receiving the first feature vector, and
Claim 9 (b)	a time period for the speech recognition word search overlaps a time period for the acoustic coprocessor to calculate the distance in response to receiving the second feature vector.
Claim 10 (Pre)	An acoustic coprocessor, comprising:
Claim 10 (a)	a first interface for receiving at least one feature vector;
Claim 10 (b)	a calculating apparatus for calculating distances indicating a similarity between the at least one feature vector and at least one acoustic state of an acoustic model read from an acoustic model memory; and
Claim 10 (c)	a second interface for sending at least one distance calculated by the calculating apparatus.
Claim 11	The acoustic coprocessor of claim 10, wherein the calculating apparatus and the acoustic model memory are fabricated on a single integrated circuit.
Claim 12	The acoustic coprocessor of claim 10, wherein the acoustic model contains Gaussian parameters and the distances are probabilities.

Claim Designation	Claim Language
Claim 13	The acoustic coprocessor of claim 10, wherein the acoustic model memory is a FLASH memory.
Claim 14	The acoustic coprocessor of claim 10, wherein the second interface for sending distances is a serial bus.
Claim 16 (Pre)	A speech recognition system comprising the acoustic coprocessor of claim 10, the speech recognition system further comprising at least one CPU, wherein:
Claim 16 (a)	the CPU executes software to perform or initiate a speech recognition word search using one or more distances calculated by the acoustic coprocessor in response to receiving a first feature vector, and
Claim 16 (b)	a time period for the speech recognition word search overlaps with a time period for the acoustic coprocessor to calculate one or more distances in response to receiving a second feature vector.
Claim 17	The acoustic coprocessor of claim 10, wherein the acoustic coprocessor is capable of, in response to receiving a feature vector, autonomously calculating the distances for each acoustic state of the acoustic model.
Claim 18	The acoustic coprocessor of claim 17, further comprising one or more result memories configured to store the distances for each acoustic state of the acoustic model.
Claim 20	The acoustic coprocessor of claim 18, wherein the calculating apparatus is capable of computing distances using acoustic models representing Gaussian probability distributions.
Claim 21	The acoustic coprocessor of claim 20, wherein the acoustic model uses a neural network to calculate at least one of the distances.
Claim 22	The acoustic coprocessor of claim 21, wherein the acoustic coprocessor is capable of generating an interrupt in response to a completion of calculation of a set of distances.

Claim Designation	Claim Language
Claim 23	The acoustic coprocessor of claim 22, wherein the acoustic model memory includes a plurality of storage locations within a static random access memory (SRAM).
Claim 24	The acoustic coprocessor of claim 23, wherein the SRAM and the calculating apparatus are fabricated on a single integrated circuit.
Claim 25	The acoustic coprocessor of claim 22, further comprising a status register, wherein the acoustic coprocessor is capable of setting one or more bits in the status register to values that indicate the completion of calculation of the set of distances.
Claim 26	The acoustic coprocessor of claim 25, wherein the acoustic coprocessor further comprises a direct memory access (DMA) controller capable of copying distances from the one or more result memories to a buffer memory.
Claim 27	The acoustic coprocessor of claim 26, wherein the DMA controller is capable of copying distances in a plurality of sequential bursts.
Claim 28 (Pre)	A speech recognition system comprising the acoustic coprocessor of claim 26, wherein:
Claim 28 (a)	the speech recognition system further comprises at least one processor configured to execute software to perform or initiate a speech recognition word search using one or more distances calculated by the acoustic coprocessor in response to receiving a first feature vector, and
Claim 28 (b)	a time period for the speech recognition word search overlaps with a time period for the acoustic coprocessor to calculate one or more distances in response to receiving a second feature vector.
Claim 29	The speech recognition system of claim 28, wherein the acoustic coprocessor and the at least one processor are fabricated on a single integrated circuit.

Claim Designation	Claim Language
Claim 35	The acoustic coprocessor of claim 26, wherein the acoustic model includes parameters represented with integer data.
Claim 37	A speech recognition system comprising the acoustic coprocessor of claim 27, the speech recognition system further comprising at least one microphone.
Claim 38	A speech recognition system comprising the acoustic coprocessor of claim 27, the speech recognition system further comprising a plurality of microphones.
Claim 39	The speech recognition system of claim 38, further comprising the buffer memory.
Claim 42	A laptop or tablet personal computer comprising the speech recognition system of claim 39.
Claim 43	A mobile phone comprising the speech recognition system of claim 39.
Claim 44	An electronic entertainment product comprising the speech recognition system of claim 39.
Claim 45	An appliance comprising the speech recognition system of claim 39.

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 13,969 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: July 12, 2023

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

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

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

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

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

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

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

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

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

/J. David Hadden/

J. David Hadden
Reg. No. 40,629

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