

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

APPLE INC., AMAZON WEB SERVICES, INC., and
AMAZON.COM SERVICES LLC
Petitioner,

v.

ZENTIAN LIMITED,
Patent Owner.

IPR2023-00035¹
Patent 10,062,377

Before KEVIN F. TURNER, JEFFREY S. SMITH, and
CHRISTOPHER L. OGDEN, *Administrative Patent Judges*.

TURNER, *Administrative Patent Judge*.

JUDGMENT
Final Written Decision
Determining All Challenged Claims Unpatentable
35 U.S.C. § 318(a)

¹ IPR2023-01194 has been joined with this proceeding.

I. INTRODUCTION

A. Background and Summary

Apple Inc. (“Petitioner”) filed a Petition (Paper 1, “Pet.”) requesting institution of *inter partes* review of claims 1–6 of U.S. Patent No. 10,062,377 B2 (Ex. 1001, “the ’377 Patent”), to which Zentian Limited (“Patent Owner”) filed a Preliminary Response. We issued an Institution Decision (Paper 10, “Inst. Dec.”) instituting the petitioned review on all challenged claims. Thereafter, Patent Owner filed a Patent Owner Response (Paper 19, “PO Resp.”) to the Petition, Petitioner filed a Reply (Paper 21, “Pet. Reply”) to the Patent Owner Response, and Patent Owner filed a Sur-reply (Paper 27, “PO Sur-reply”) to the Reply. An oral hearing was held on March 12, 2024, for which the transcript was entered into the record (Paper 33, “Tr.”).

We have jurisdiction under 35 U.S.C. § 6(b)(4) and § 318(a). This Decision is a final written decision under 35 U.S.C. § 318(a) and 37 C.F.R. § 42.73 as to the patentability of claims 1–6 of the ’140 patent. We determine Petitioner has shown by a preponderance of evidence that claims 1–6 are unpatentable.

B. Related Matters

Both parties identify the following judicial or administrative matters that would affect, or be affected by, a decision in this proceeding: *Zentian Ltd. v. Apple Inc.*, Case No. 6:22-cv-00122 (W.D. Tex.); *Zentian Ltd. v. Amazon.com, Inc.*, Case No. 6:22-cv-00123 (W.D. Tex.). Pet. 75; Paper 3, 1. Both parties also identify related *inter partes* reviews: IPR2023-00033, IPR2023-00034, IPR2023-00036, and IPR2023-00037, involving the instant parties. *Id.*

C. The '377 Patent

The '377 Patent is titled “Distributed Pipelined Parallel Speech Recognition System,” and is directed to a speech recognition system using multiple programmable devices to perform different steps of the recognition process. Ex. 1001, code (54), 5:63–6:6, 6:20–33. Figure 16, reproduced below, illustrates a speech recognition system comprising a front end 103, distance calculation engine 104, and search stage 106:

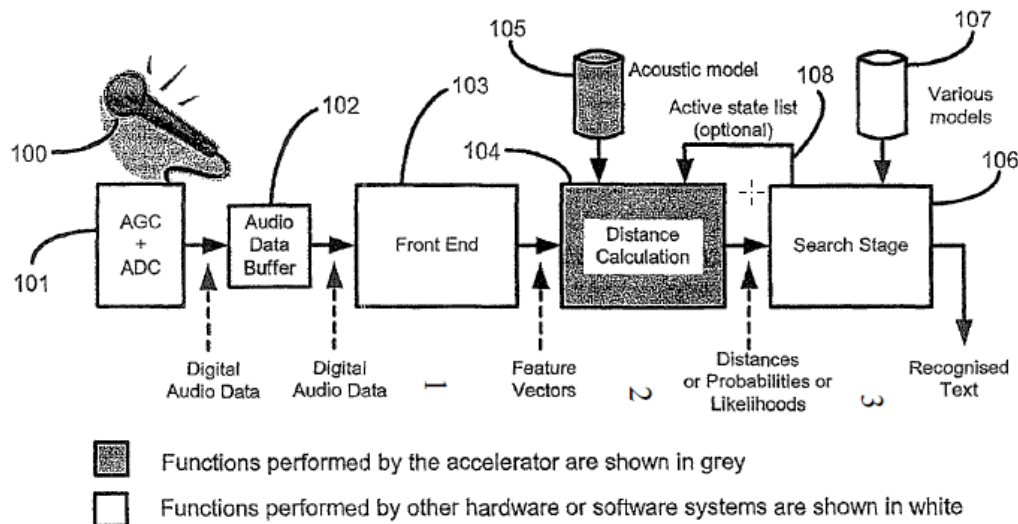


Figure 16

Figure 16 illustrates a block diagram showing an embodiment of a speech recognition system, illustrating data flow between parts thereof

The '377 patent teaches that an “audio input for speech recognition” may be input to the front end (for example at Front End 103) in the form of digital audio or analog audio that is converted to digital audio using an analog to digital converter. Ex. 1001, 12:50–53. “The audio input is divided into time frames, each time frame typically being on the order of 10 ms.” *Id.* at 12:53–55. “For each audio input time frame, the audio signal is converted into a feature vector. This may be done by splitting the audio signal into

spectral components,” such as, for instance, 13 components plus their first and second derivatives, creating a total of 39 components. *Id.* at 12:56–58. The feature vector thus “represents a point in an N-dimensional space,” where N is generally in the range of 20 to 39. *Id.* at 13:19–23.

Each feature vector is then passed to the calculating circuit, or distance calculation engine (for example element 104), which calculates a distance indicating the similarity between a feature vector and one or more predetermined acoustic states of an acoustic model. Ex. 1001, 5:63–6:2, 25:33–35 (“Each feature vector is transferred to a distance calculation engine circuit 204, to obtain distances for each state of the acoustic model.”). “The distance calculator stage of the recognition process computes a probability or likelihood that a feature vector corresponds to a particular state.” *Id.* at 13:24–26. “The likelihood of each state is determined by the distance between the feature vector and each state.” *Id.* at 13:1–2. The distance calculation may be a Mahalanobis distance using Gaussian distributions. *Id.* at 4:20–21. “The MHD (Mahalanobis Distance) is a distance between two N-dimensional points, scaled by the statistical variation in each component.” *Id.* at 13:13–15. The distance calculation engine or calculating circuit “may be included within an accelerator” (Ex. 1001, 3:59–61), which may be a “loosely bound co-processor for a CPU running speech recognition software,” and which “has the advantage of reducing computational load on the CPU, and reducing memory bandwidth load for the CPU.” *Id.* at 24:17–20; *see* Figs. 17–23.

The distances calculated by the distance calculation engine are then transferred to search stage 106 of the speech processing circuit, which uses models, such as one or more word models and/or language models, to generate and output recognized text. Ex. 1001, 24:5–11. Search stage 106

may use the distance calculations using a Hidden Markov Model (HMM) or a neural network. *Id.* at 4:40–44.

Thus, “the DSP [element 103] provides a feature vector to the Accelerator, [element 104] and the accelerator provides a set of distance results to the CPU [element 105].” Ex. 1001, 27:45–50. The system, thereby, provides a speech recognition circuit with three programmable devices per independent claim 1.

D. Illustrative Claim

Claim 1 is the sole independent claim challenged in this proceeding, with each of challenged claims 2–6 dependent on claim 1, directly or indirectly. Independent claim 1 is considered to be representative and is reproduced below:

1. [1(Pre)] A speech recognition system comprising:

[1(a)] a first programmable device programmed to calculate a feature vector from a digital audio stream, [1(b)] wherein the feature vector comprises a plurality of extracted and/or derived quantities from said digital audio stream during a defined audio time frame;

[1(c)] a second programmable device programmed to calculate distances indicating the similarity between a feature vector and a plurality of acoustic states of an acoustic model [1(d)] wherein said feature vector is received by the second programmable device after it is calculated by the first programmable device; and

[1(e)] a third programmable device programmed to identify spoken words in said digital audio stream using Hidden Markov Models and/or Neural Networks [1(f)] wherein said word identification uses one or more distances that were calculated by the second programmable device, [1(g)] wherein said identification of spoken words uses one or more distances calculated from a first feature vector; and

[1(h)] a search stage for using the calculated distances to identify words within a lexical tree, the lexical tree comprising a model of words.

Ex. 1001, 38:53–39:30 (with annotations provided by Petitioner, Pet. 77–78).

E. Instituted Grounds of Unpatentability

Petitioner asserts the following grounds of unpatentability (Pet. 5), supported by the declaration of Mr. Christopher Schmandt (Ex. 1003), upon which institution was granted:

Claims Challenged	35 U.S.C. §	Reference(s)/Basis
1	103(a) ²	Jiang, ³ Smyth ⁴
2, 3	103(a)	Jiang, Smyth, Nguyen ⁵
4	103(a)	Jiang, Smyth, Nguyen, Boike ⁶
5	103(a)	Jiang, Smyth, Nguyen, Baumgartner ⁷
6	103(a)	Jiang, Smyth, Nguyen, Boike, Baumgartner

² The Leahy-Smith America Invents Act (“AIA”), Pub. L. No. 112-29, 125 Stat. 284, 285–88 (2011), revised 35 U.S.C. § 103 effective March 16, 2013. Because the challenged patent claims priority to an application filed before March 16, 2013, we refer to the pre-AIA version of § 103.

³ U.S. Patent No. 6,374,219 B1, filed Feb. 20, 1998, issued Apr. 16, 2002 (Ex. 1004, “Jiang”).

⁴ U.S. Patent No. 5,819,222, issued Oct. 6, 1998 (Ex. 1005, “Smyth”).

⁵ U.S. Patent No. 6,879,954 B2, filed Apr. 22, 2002, issued Apr. 12, 2005 (Ex. 1047, “Nguyen”).

⁶ U.S. Patent No. 6,959,376 B1, filed Oct. 11, 2001, issued Oct. 25, 2005 (Ex. 1006, “Boike”).

⁷ U.S. Patent Publication 2002/0049582 A1, published Apr. 25, 2002 (Ex. 1007, “Baumgartner”).

II. ANALYSIS

A. *Legal Standards*

“In an [*inter partes* review], the petitioner has the burden from the onset to show with particularity why the patent it challenges is unpatentable.” *Harmonic Inc. v. Avid Tech., Inc.*, 815 F.3d 1356, 1363 (Fed. Cir. 2016) (citing 35 U.S.C. § 312(a)(3) (requiring *inter partes* review petitions to identify “with particularity . . . the evidence that supports the grounds for the challenge to each claim”)); *see also* 37 C.F.R. § 42.104(b) (requiring a petition for *inter partes* review to identify how the challenged claim is to be construed and where each element of the claim is found in the prior art patents or printed publications relied upon).

A claim is unpatentable under 35 U.S.C. § 103(a) if “the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains.” *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 406 (2007). The question of obviousness is resolved on the basis of underlying factual determinations, including: (1) the scope and content of the prior art; (2) any differences between the claimed subject matter and the prior art; (3) the level of skill in the art; and (4) when in evidence, objective evidence of obviousness or nonobviousness, i.e., secondary considerations. *See Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966). An obviousness analysis “need not seek out precise teachings directed to the specific subject matter of the challenged claim, for a court can take account of the inferences and creative steps that a person of ordinary skill in the art would employ.” *KSR*, 550 U.S. at 418.

Additionally, the obviousness inquiry typically requires an analysis of “whether there was an apparent reason to combine the known elements in the fashion claimed by the patent at issue.” *KSR*, 550 U.S. at 418 (citing *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2016) (requiring the adjudicator to provide “some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness”)). Furthermore, Petitioner does not satisfy its burden of proving obviousness by employing “mere conclusory statements,” but “must instead articulate specific reasoning, based on evidence of record, to support the legal conclusion of obviousness.” *In re Magnum Oil Tools Int’l, Ltd.*, 829 F.3d 1364, 1380 (Fed. Cir. 2016).

B. Level of Ordinary Skill in the Art

Petitioner, supported by Mr. Schmandt’s testimony, proposes that a person of ordinary skill in the art at the time of the invention “would have had a master’s degree in computer engineering, computer science, electrical engineering, or a related field,” and would have had “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.” Pet. 4 (citing Ex. 1003 ¶ 24). Patent Owner indicates that it does not “dispute the Petition’s proposed level of skill for a person of skill in the art,” but also emphasizes that such artisans who possess “*ordinary skill* in the field of speech recognition, [would] not [be] one of extraordinary creativity.” PO Resp. 4.

We continue to determine, as we did in the Institution Decision, that Petitioner’s proposal consistent with the level of ordinary skill in the art reflected by the prior art of record, *see* Inst. Dec. 7; *Okajima v. Bourdeau*, 261 F.3d 1350, 1355 (Fed. Cir. 2001); *In re GPAC Inc.*, 57 F.3d 1573, 1579

(Fed. Cir. 1995); *In re Oelrich*, 579 F.2d 86, 91 (CCPA 1978), and, we continue to adopt Petitioner’s unopposed position as to the level of ordinary skill in the art for purposes of this decision.

C. Claim Construction

We construe the challenged claims

using the same claim construction standard that would be used to construe the claim in a civil action under 35 U.S.C. 282(b), including construing the claim in accordance with the ordinary and customary meaning of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.

37 C.F.R. § 42.100(b) (2022). “In determining the meaning of [a] disputed claim limitation, we look principally to the intrinsic evidence of record, examining the claim language itself, the written description, and the prosecution history, if in evidence.” *DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.*, 469 F.3d 1005, 1014 (Fed. Cir. 2006). Claim terms are given their ordinary and customary meaning as would be understood by a person of ordinary skill in the art at the time of the invention and in the context of the entire patent disclosure. *Phillips v. AWH Corp.*, 415 F.3d 1303, 1313 (Fed. Cir. 2005) (en banc). “There are only two exceptions to this general rule: 1) when a patentee sets out a definition and acts as his own lexicographer, or 2) when the patentee disavows the full scope of a claim term either in the specification or during prosecution.” *Thorner v. Sony Comput. Entm’t Am. LLC*, 669 F.3d 1362, 1365 (Fed. Cir. 2012).

We noted at Institution that “[n]either party provides explicit claim constructions for claim features.” Inst. Dec. 8. However, at numerous points during the oral hearing, the panel indicated that many of the arguments put forth in the parties’ papers raised implicit claim constructions,

and that attempts to formulate explicit claim constructions now would lack the benefit of the parties' briefings. Tr. 27:8–14, 29:20–24, 32:18–33:26. Nonetheless, although we adopt an explicit claim construction below, we note that this construction comes from the implications of the parties' arguments in the briefs, as well as the panel's discussion with both parties at the oral hearing, with both sides providing input into how the construction should be formulated.

Element 1(b) of independent claim 1 recites: “wherein the feature vector comprises a plurality of extracted and/or derived quantities from said digital audio stream during a defined audio time frame.” This portion of the claim provides a definition for what is a feature vector (“FV”). One of Patent Owner's central issues in its Response is the difference between the use of “codewords,” discussed in Jiang⁸, from the use of “feature vectors” in making distance calculations. *See* PO Resp. 5–13; PO Sur-reply 1–6. Petitioner responds that Patent Owner is incorrect in suggesting that “a codeword is not an FV” (Pet. Reply 3), and asserting that codewords are themselves FVs and also direct representations of a FV. *Id.* at 3–5. As such, the definition of “feature vectors,” or FVs, which comes from the language of independent claim 1, is important to resolving the issues raised by the parties.

The explicit language of claim 1 provides that “the feature vector comprises a plurality of extracted and/or derived quantities from said digital audio stream.” As discussed at oral hearing:

⁸ Jiang also discloses the use of “feature vectors,” as discussed below, which both parties acknowledge to be “feature vectors” within both parties' implicit claim constructions.

JUDGE SMITH: . . . given that extracted and derived are different words, and we want to give these words different meanings, if we give the meaning extracted to mean, you know, literally taken from the audio stream, as you're proposing, that means derived means something other than literally taking it from the audio stream.

Tr. 26: 23–27:6. In response, Patent Owner's advocate indicated that FVs "themselves are not literally extracted out of the audio stream," and they are "derived from the audio stream, because they're not literally, you know, sort of clicked out of the sound frequencies." *Id.* at 28:14–20. Patent Owner also asserted that the words "extracted" and "derived" are used in claim 1 "to show that we mean a similar concept," referencing the use of "bought and/or acquired," with "acquired encompass[ing] a broader scope, . . . than bought, but in both instances we've acquired something." *Id.* at 29:13–19. We do not agree with Patent Owner.

"All words in a claim must be considered in judging the patentability of that claim against the prior art." *In re Wilson*, 424 F.2d 1382, 1385 (CCPA 1970). The recitation of claim 1 that FVs are extracted *and/or derived* means that FVs may not only be extracted from the digital audio stream, thus obtaining a FV based on the scale of the FV ("representing a point in an N-dimensional space," depending on the value of N (Ex. 1001 13:19–23)). Such FVs, per claim 1, may also be *representations* of the digital audio stream, i.e., quantities *derived* therefrom. Any such derived quantities would perhaps even be lossy. *See* Ex. 1001, 13:13–18. As such, FVs, according to independent claim 1, are not limited to *pure* feature vectors, which are quantities extracted from the digital audio stream, but also include quantities derived from the stream, which are representations

thereof. *See* Pet. 22 (citing Ex. 1003 ¶¶ 140–141; Ex. 1001, 35:36–37, 5:6–8).

Given such a definition, vector quantization, which reduces the plurality of quantities in a FV to a single codeword that is associated with a codebook entry, representing multiple values, would also comprise “derived” representations of the “quantities from said digital audio stream” as recited in the claim. Although Patent Owner’s declarant, Dr. Anderson, provides that “[c]odewords are the result of vector quantization (VQ) of the original feature vectors, and are thus not the feature vectors themselves” (Ex. 2020 ¶ 39), that comports with one alternative of what encompasses a FV as recited in claim 1, namely, “extracted quantities,” but not the more expansive definition that includes “derived quantities” provided in independent claim 1. The vector quantization of a feature vector results in a vector quantized feature vector, or codeword, that still represents a multitude of values “derived . . . from said digital audio stream,” but is more computationally expedient. *See* Ex. 1003 ¶¶ 140–141. We agree with Dr. Anderson that persons of ordinary skill in the art would have understood that calculating distances from a feature vector comprising extracted quantities was distinct from computing probabilities from a codeword (Ex. 2020 ¶ 43), but that, again, does not take into consideration the broader definition of “feature vector” confirmed by the claims. Rather, we agree with Petitioner and Mr. Schmandt, that a person of ordinary skill in the art would have understood that a codeword is a representative feature vector. Pet. 22 (citing Ex. 1003 ¶¶ 140–141). A codeword, or representative feature vector, “comprises a plurality of . . . derived,” or vector quantized, “quantities from said digital audio stream” within the meaning of claim 1.

Thus, we determine that a “feature vector,” as defined by independent claim 1, comprises a plurality of extracted qualities from the digital audio stream during the defined audio time frame, and/or derived quantities that are representative of the digital audio stream during the defined audio time frame, such as the derived values of a codeword.

D. Obviousness over Jiang and Smyth

Petitioner asserts that the combination of Jiang and Smyth would have rendered the subject matter of claim 1 obvious to one of ordinary skill in the art at the time of the invention. Pet. 5–59. Patent Owner argues that the cited references do not teach or suggest at least limitation 1(g) and that the Petition has failed to establish the requisite motivation to combine the cited references. PO Resp. 5–19. We begin with brief discussions of the cited references, and then consider Petitioner’s arguments with respect to the references’ teachings applied to the instant claim as well as Patent Owner’s arguments asserting deficiencies in this ground of unpatentability.

1. Jiang

Jiang is directed to “computer speech recognition performed by conducting a prefix tree search of a silence bracketed lexicon.” Ex. 1004, 1:15–18. Possible words represented by the input data stream are provided as a prefix tree including a plurality of phoneme branches connected at nodes. *Id.* at 4:8–12. Speech is input into system 60 in the form of audible voice signal provided by the user to microphone 62, which converts the audible speech into an analog electric signal, which is converted by A/D converter 64 into a sequence of digital signals that are then provided to feature extraction module 66. *Id.* at 6:45–52. Feature extraction module 66

divides the digital signal into frames, each approximately 10 ms in duration.
Id. at 6:62–65.

The frames are then preferably encoded by feature extraction module 66 into a feature vector reflecting the spectral characteristics for a plurality of frequency bands. Ex. 1004, 6:65–7:1. “In the case of discrete and semi-continuous hidden Markov modeling, feature extraction module 66 also preferably encodes the feature vectors into one or more codewords using vector quantization techniques and a codebook derived from training data.”
Id. at 7:1–5.

Upon receiving the codewords from feature extraction module 66, and the boundary detection signal provided by silence detection module 68, tree search engine 74 accesses information stored in the phonetic speech unit model memory 72. Ex. 1004, 7:30–34. Based upon the HMMs stored in memory 72, tree search engine 74 determines a most likely phoneme represented by the codeword received from feature extraction module 66, and hence representative of the utterance received by the user of the system.
Id. at 7:37–42.

An example system of Jiang with feature extraction module 66 functionally in pipeline with tree search engine 74 is illustrated in Figure 2, reproduced below:

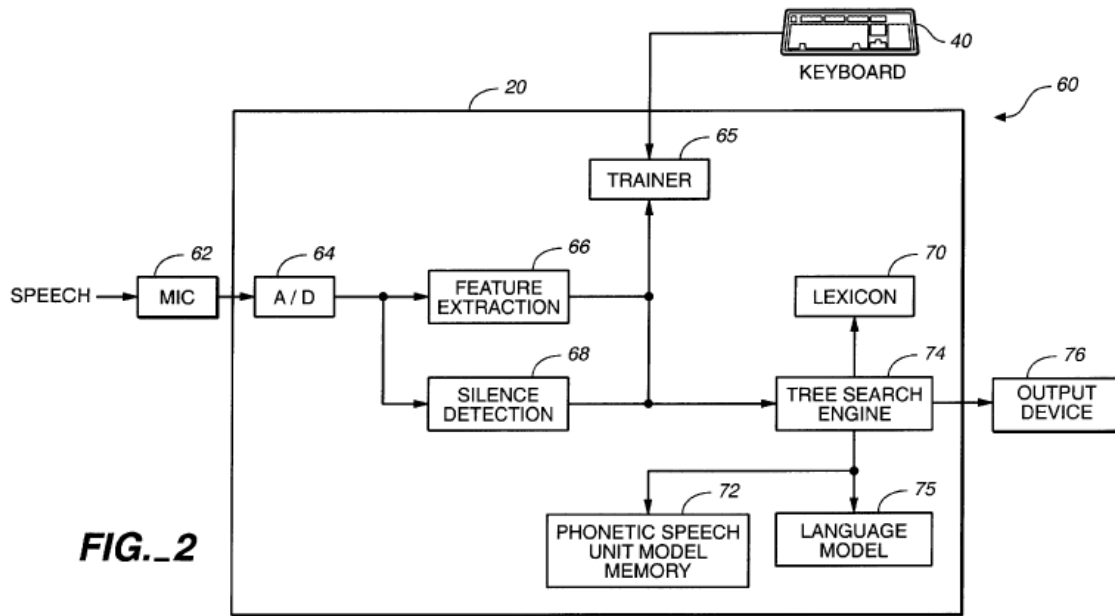


Figure 2 of Jiang illustrates a detailed block diagram of a portion of its system

“In a preferred embodiment, feature extraction module 66 is a conventional array processor which performs spectral analysis on the digital signals.” Ex. 1004, 6:53–55. “Further, tree search engine 74 is preferably implemented in CPU 21 (which may include one or more processors) or may be performed by a dedicated speech recognition processor employed by personal computer 20.” *Id.* at 6:39–42.

2. Smyth

Smyth is directed to “[a] speech recognition system [which] recognizes connected speech using a plurality of vocabulary nodes.” Ex. 1005, Abs. Smyth is of particular interest in the area of task-constrained connected word recognition where the task, for example, might be to recognize one of a set of account numbers or product codes. *Id.* at 1:18–21.

Smyth teaches a system including feature extractor 33 for generating from a frame of samples a corresponding feature vector and classifier 34

receiving the succession of feature vectors and operating on each with a plurality of model states, to generate recognition results. Ex. 1005, 5:5–9. Smyth further includes parsing processor 351, part of sequencer 35, arranged to read, at each frame, the state probabilities output by classifier processor 341, which is part of classifier 34. *Id.* at 6:35–39. Parsing processor 351 is specifically configured to recognize certain phrases or words. *Id.* at 6:62. This system is illustrated in Figure 2, reproduced below:

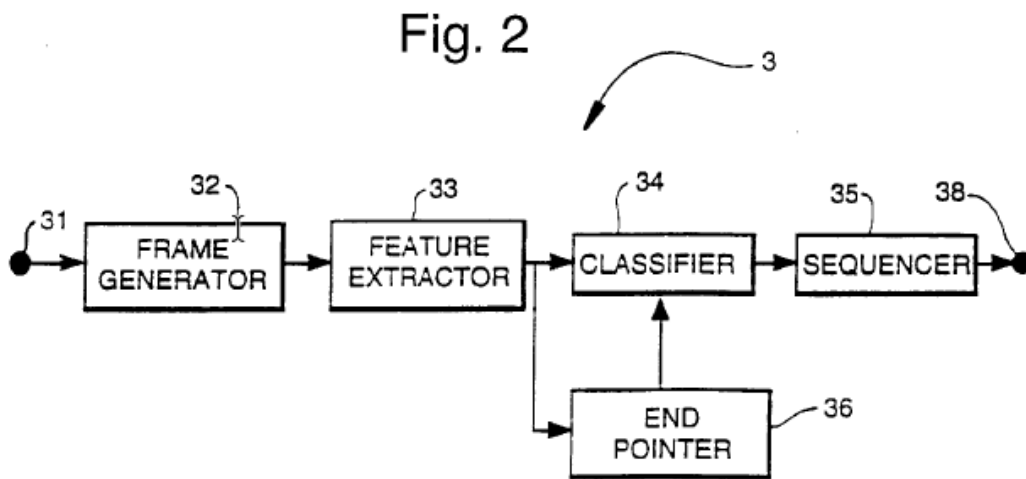


Figure 2 of Smyth illustrates a block diagram showing the functional elements of the recognition processor

The frame generator 32 and feature extractor 33 are, in this embodiment, provided by a single suitably programmed digital signal processor (DSP) device “(such as the Motorola DSP 56000, or the Texas Instruments TMC C 320) or similar device.” Ex. 1005, 5:49–53. The classifying processor 341 may be a suitably programmed digital signal processing (DSP) device, may in particular be the same digital signal processing device as the feature extractor 33. Ex. 1005, 6:5–9. The parsing processor 351 may, for example, be a microprocessor “such as the Intel^(TM) i-486^(TM) microprocessor or the Motorola^(TM) 68000 microprocessor, or may

alternatively be a DSP device (for example, the same DSP device as is employed for any of the preceding processors).” *Id.* at 6:49–51.

Frame generator 32 is arranged to receive speech samples, and feature extractor 33 receives frames from the frame generator 32 and generates, in each case, a set or vector of features, where each is a corresponding feature vector. Ex. 1005, 5:7, 20, 31–33.

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. Ex. 1005, 5:63–67. “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.” *Id.* at 6:1–4.

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 to date over time, and to compare this with each of the state sequences stored in the state sequence memory 352. Ex. 1005, 6:36–43. “The calculation employs the well-known Hidden Markov Model method.” *Id.* at 6:43–45. Accordingly, for each state sequence (corresponding to a word, phrase or other speech sequence to be recognized) a probability score is output by the parser processor 351 at each frame of input speech. For example, the state sequences may comprise the names in a telephone directory. *Id.* at 6:52–56.

3. *Claim 1*

a. *Element 1(Pre): Preamble, “A speech recognition system”*

With respect to the preamble of claim 1, Petitioner asserts that “Jiang teaches a speech recognition system: ‘A speech recognition system recognizes speech based on an input data stream indicative of the speech.’” Pet. 10 (citing Ex. 1004, 4:8–9, 4:30–32, 1:15–19, 6:18–20, code (57), Figs. 1–2).

Patent Owner does not contest that Jiang is directed to “a speech recognition system.” See PO Resp. We find Petitioner’s arguments persuasive for the reasons outlined in the Petition, and determine, without deciding whether the preamble is limiting, that Petitioner has established that Jiang meets the preamble of claim 1 for the reasons explained by Petitioner.

b. *Element 1(a): “a first programmable device programmed to calculate a feature vector from a digital audio stream”*

With respect to a first programmable device calculating a feature vector from audio data, Petitioner asserts that Jiang teaches a feature extraction module 66 for calculating a feature vector, as claimed. Jiang’s feature extraction module 66 is a programmable device, as the module 66 may be a hardware or software module in computer 20 having CPU 21. Pet. 10 (citing Ex. 1004, 6:31–36, 6:46–7:10).

Petitioner further states that to any extent Patent Owner argues that Jiang fails to teach the element, “Smyth teaches a ‘first programmable device’ programmed to calculate a feature vector, as claimed, namely ‘a single suitably programmed digital signal processor (DSP) device . . . or similar device’ providing frame generator 32 and feature extractor 33.” Pet. 14 (citing Ex. 1005, 5:49–53). Petitioner continues with the assertion that Smyth also teaches that its programmable device is *programmed to*

calculate a feature vector of a digital audio stream. Pet. 14–15 (citing Ex. 1005, 5:2–3).

Patent Owner does not offer counterarguments with respect to this element of claim 1. *See* PO Resp. We find Petitioner’s arguments persuasive for the reasons outlined in the Petition, and determine that Petitioner has established that both Jiang and Smyth meet this limitation of claim 1 for the reasons explained by Petitioner.

- c. Element 1(b): “wherein the feature vector comprises a plurality of extracted and/or derived quantities from said digital audio stream during a defined audio time frame”*

Petitioner asserts that Jiang teaches the calculated feature vector comprises extracted and/or derived quantities in the form of “spectral characteristics.” The frames are preferably encoded by feature extraction module 66 into a feature vector reflecting the spectral characteristics for a plurality of frequency bands. Pet. 17 (citing Ex. 1004, 6:62–7:10, 6:53–57). Petitioner further asserts that Jiang teaches the feature vector includes extracted and/or derived quantities of the digital audio stream “during a defined audio time frame,” as claimed. Jiang teaches the digital signal is divided into frames of approximately 10 milliseconds in duration, where the frames are then encoded by the feature extraction module 66 into a feature vector, where the feature vector “reflect[s] the spectral characteristics.” Pet. 18 (citing Ex. 1004, 6:62–7:1).

Patent Owner does not offer counterarguments with respect to this element of claim 1. *See* PO Resp. With respect to this limitation, we take into account the claim construction of “feature vector,” discussed above. *See* II.C. We find Petitioner’s arguments persuasive for the reasons outlined

in the Petition, and determine that Petitioner has established that Jiang meets this limitation of claim 1 for the reasons explained by Petitioner.

- d. Element 1(c): “a second programmable device programmed to calculate distances indicating the similarity between a feature vector and a plurality of acoustic states of an acoustic model”*

Petitioner asserts that Jiang in combination with Smyth teaches this limitation, with Smyth being relied on in the combination for teaching the “second programmable device.” Pet. 19. Petitioner asserts that both Jiang and Smyth teach the claimed functionality of calculating distances. *Id.* at 18.

In regard to Jiang, Petitioner demonstrates that Jiang uses HMMs to identify speech units, thereby teaching an acoustic model with a plurality of acoustic states. Pet. 21. Petitioner further maps the HMM states to the recited distance calculations, asserting a distance is a probability or likelihood of a feature vector compared to a particular state of an HMM, thereby equating distances with probabilities or likelihoods. *Id.* at 20–21.

Petitioner notes that Jiang teaches encoding feature vectors into codewords using vector quantitation techniques, which will be lossy. Pet. 22. Jiang teaches these codewords are applied to the HMMs to identify utterances. *Id.* at 23. Petitioner argues that since a probability is associated with a corresponding state of an HMM, this represents a likelihood that a codeword corresponds to a phoneme utterance, which Petitioner further argues is the recited “distance” of this limitation. *Id.* at 23–25.

In regard to Smyth, Petitioner asserts Smyth teaches “three distinct DSPs for performing the three general speech recognition steps” of feature vector calculation, distance calculation, and word identification. Pet. 28. Petitioner further asserts that Smyth discloses the second programmable device is programmed to calculate distances as recited. *Id.* at 34. In support

of this proposition, Petitioner reasons Smyth teaches calculating a plurality of probabilities, and that this operation is calculating distances as recited because processor 341 receives the feature vector and calculates the probability of correspondence. *Id.* at 37.

Patent Owner does not offer counterarguments with respect to this element of claim 1. *See* PO Resp. We find Petitioner's arguments persuasive for the reasons outlined in the Petition, and determine that Petitioner has established that both Jiang and Smyth meet this limitation of claim 1 for the reasons explained by Petitioner.

- e. Element 1(d): "wherein said feature vector is received by the second programmable device after it is calculated by the first programmable device"*

Petitioner asserts that Jiang teaches the recited element where it discloses that tree search engine 74 receives codewords from feature extraction module 66, where the codeword is a representation of a feature vector; based upon HMMs stored in memory, tree search engine 74 determines a most likely phoneme represented by the codeword received from feature extraction module 66. Pet. 41–42 (citing Ex. 1004, 7:30–42). Therefore, Petitioner seems to assert that because codewords represent feature vectors, receiving the codewords at tree search engine 74 after the feature vectors are calculated by feature extraction module 66 and converted to codewords teaches the recited.

Petitioner further asserts that Smyth teaches the recited element because Smyth teaches feature extractor 33 which calculates feature vectors, and classifier 34 which uses the feature vectors calculated by feature extractor 33 to determine probability the likelihood that a feature vector corresponds to a state. Pet. 42–43 (citing Ex. 1005, 5:7–9, 63–67).

Patent Owner does not offer counterarguments with respect to this element of claim 1. *See* PO Resp. We find Petitioner's arguments persuasive for the reasons outlined in the Petition, and determine that Petitioner has established that both Jiang and Smyth teach the calculation of feature vectors at a device, and then passing the feature vectors or a representation thereof to a further device for further computation.

f. Element 1(e): "a third programmable device programmed to identify spoken words in said digital audio stream using Hidden Markov Models and/or Neural Networks"

Petitioner asserts that Jiang alone or in combination with Smyth teaches this claim element of claim 1. Pet. 44.

First, in regard to Jiang, Petitioner asserts that Jiang teaches the recited third programmable device. Pet. 44. Petitioner also argues it would have been obvious to use a third programmable device because it would have been obvious to use single function devices for each functional stage of the three-stage pipeline of feature vector calculation, similarity determination, and word identification. *Id.* at 44–45 (citing Ex. 1003 ¶¶ 71–72, 111–120, 156, 187–188). Petitioner additionally cites to Smyth as teaching three programmable devices when applied to Jiang. *Id.*

Further in regard to Jiang, Petitioner asserts Jiang teaches the recited features of identifying spoken words using HMM. Pet. 45. Petitioner states that the HMMs are used to determine a score for a node and associated phoneme branch of the prefix tree, such that the score indicates the likelihood the phoneme being examined for the phoneme branch is the actual phoneme indicated by the codeword. *Id.* (citing Ex. 1004, 8:16–24, 7:30–43).

Second, in regard to Smyth, Petitioner asserts Smyth teaches the claimed third programmable device which identifies words, citing to the parsing processor 351 of Smyth. Pet. 47. Petitioner states that parsing processor 351 calculates the most likely path of states, compares this path with stored state sequences, and identifies words and phrases, where parsing processor 351 maybe a microprocessor, or may alternatively be a DSP device. *Id.* (citing Ex. 1005, 6:36–60).

Further in regard to Smyth, Petitioner asserts Smyth teaches identifying spoken words using Hidden Markov Models, as claimed. Pet. 50. Petitioner points out Smyth also teaches parsing processor 351’s word identification calculation employs the well-known Hidden Markov Model method. *Id.* (citing Ex. 1005, 6:43–44). Patent Owner does not offer counterarguments with respect to this element of claim 1. *See* PO Resp.

We determine that Petitioner has sufficiently established that an ordinarily skilled artisan would have known to use HMMs to identify spoken words in a final stage of speech recognition at least from Smyth. We determine that use of HMMs in speech recognition was well known and commonly used, as exemplified by both Jiang and Smyth. Furthermore, Smyth teaches to use HMMs in a DSP device for the final word recognition determination. When this teaching of Smyth is applied to Jiang, one of skill in the art would understand that the sequential functions of Jiang may be affected by distinct devices, and that the final function in the word recognition pipeline may be a HMM calculation effected by a third programmable device.

- g. *Element 1(f): “wherein said word identification uses one or more distances that were calculated by the second programmable device”*

As in regard to claim 1 element 1(c) discussed above, Petitioner equates the recited distances of claim 1 with a probability or likelihood score. Pet. 56. Petitioner asserts Jiang teaches calculating a score indicating a likelihood that the phoneme being examined in the prefix tree is the actual phoneme indicated by the codeword (i.e., the feature vector) received from the feature extraction module, where this calculation of the score utilizes the HMM states of Jiang. *Id.* (citing Ex. 1004, 7:34–43, 8:16–24).

Patent Owner does not offer counterarguments with respect to this element of claim 1. *See* PO Resp. We find Petitioner’s arguments persuasive for the reasons outlined in the Petition, and determine that Petitioner has established that that Jiang meets this limitation of claim 1 for the reasons explained by Petitioner.

h. Element 1(g): “wherein said identification of spoken words uses one or more distances calculated from a first feature vector”

Petitioner references the above discussion of claim 1 element 1(f), and the equating of probability with distances. Pet. 58. In regard to mapping the recited feature of “distances calculated from a first feature vector,” Petitioner states that Jiang teaches word identification which “uses one or more distances calculated from a first feature vector because the word identification uses a phoneme determined from a first feature vector.” *Id.* at 59 (citing Ex. 1003 ¶ 214).

Patent Owner asserts that this limitation requires that “said identification of spoken words uses one or more distances calculated from a first feature vector,” and that such a limitation “is not met by distances calculated from feature vector representatives or any other proxy or substitute for feature vectors.” PO Resp. 1. Patent Owner continues that the

instant limitation requires calculations made from the feature vectors themselves, and the limitation would have no purpose “if it were read to be satisfied by distance calculations made from feature vector representatives.” *Id.* Patent Owner emphasizes that limitation 1(c), discussed above, could be interpreted as not requiring a direct comparison between a feature vector and an acoustic state, and may not proscribe use of codewords to represent feature vectors, but that limitation 1(g) specifically requires using one or more distances calculated from a first feature vector for word recognition. *Id.* at 5–6. Patent Owner also argues that interpreting limitation 1(g) to not require distances calculated from the feature vectors themselves would render limitation 1(g) pointless and meaningless in view of limitations 1(f) and 1(c). *Id.* at 7 (citing Ex. 1020 ¶ 37).

Patent Owner also argues that vector-quantized codewords are not the original underlying feature vectors recited in limitation 1(g), and thus computations “*from* codewords” are not computations “*from* feature vectors.” PO Resp. 7 (citing Ex. 1020 ¶ 39). Patent Owner additionally argues that at the time of the invention it was widely understood that calculating distances from a feature vector was a specific computational technique that was distinct from computing probabilities from codewords. *Id.* at 8 (citing Ex. 1020 ¶ 43).

In response, Petitioner asserts that Jiang does not mandate codewords, and that Patent Owner “ignores the optional language and improperly focuses only on Jiang’s teaching of performing word recognition using codewords.” Pet. Reply 1–2. Petitioner also argues that “[c]odewords are, in and of themselves, FVs,” as well as direct representations of an input FV. *Id.* at 3, 5. Petitioner also argues that Dr. Anderson, Patent Owner’s declarant, acknowledges that a codeword is a vector. *Id.* at 5–6 (citing Ex.

1073, 53:16–21 (“[v]ector quantization is the representation of a vector”), 55:2–10 (“The vector is not the same at the one that you are replacing, ... [but it] does represent or indicate a vector.”). Petitioner also argues that limitation 1(g) remains distinct from claim 1’s other limitations, because whether employing an input of codeword FV, the distinct features of claim 1 still requires a distance calculation from a FV that indicates a similarity to a plurality of states and is used to identify speech. *Id.* at 10.

Patent Owner responds that the requirements for a FV are recited in limitation 1(b), and require that a single FV must comprise a plurality of quantities and those quantities must be extracted or derived from the digital audio stream. PO Sur-Reply 1. Patent Owner argues that a vector quantized codeword instead represents a single value (*id.* at 1–3) and that codeword values are created from “sample data,” and not from quantities derived or extracted from the digital audio signal (*id.* at 4–5). Patent Owner also argues that the Petition “never alleged that Jiang’s codewords themselves meet the definitional requirements of “feature vector” as set forth in limitation 1(b),” instead relying on Jiang teaching of feature extraction module 66. *Id.* at 5 (citing Pet. 18; Ex. 1004 6:65–7:1). As such, Patent Owner argues that Petitioner’s Reply argument “is entirely new, inconsistent with the Petition, and improper. *Id.* at 5–6 (citing *Wasica Fin. GmbH v. Cont’l Auto. Sys.*, 853 F.3d 1272, 1286–87 (Fed. Cir. 2017); 35 U.S.C. § 312(a)(3); 37 C.F.R. § 42.104(b)(4)-(5); 37 C.F.R. §42.23(b)).

We do not agree with Patent Owner’s arguments. As discussed above (Section II.C), we determine that a “feature vector,” as defined by independent claim 1, “comprises a plurality of extracted . . . quantities from said digital audio stream during a defined audio time frame,” and/or “derived quantities” that are representative of the digital audio stream

“during [the] defined audio time frame.” As such, vector quantized feature vectors, or codewords, are representative of the digital audio stream and are feature vectors under the instant claim construction. Although a vector quantized codeword may be represented by a single value, it still represents multiple quantities derived from the audio stream, accessible through that single value. *See* Ex. 1003 ¶¶ 140–141. Additionally, although codewords may be determined in advance, they are still used in Jiang to derive representative, vector quantized versions of the computed feature vectors. *See* Ex. 1004, 6:62–7:7; Ex. 1003 ¶¶ 140–141.

With respect to Patent Owner assertions regarding Petitioner’s “new arguments” in the Reply, we also disagree. Claim limitation 1(b) describes the aspects of feature vectors. The Petition details how Jiang discloses this limitation through its use of feature vectors. *See* Pet. 17–18. There was no need for Petitioner to assert that codewords also meet this limitation because Jiang meets the limitation through its own disclosure of feature vectors, which Patent Owner does not appear to contest, at least with respect to 1(b). *See* PO Resp.; PO Sur-reply. Later in the Petition, when the calculations in Jiang are discussed, the Petition details that:

Jiang describes the speech recognition as determining the phoneme(s) represented by the *codewords*. *Jiang*, 7:38-43. A POSITA would have understood that because the codewords are vector quantized feature vectors, then *Jiang*’s teaching of determining the phonemes of the *codewords* is likewise determining the phonemes of the *feature vectors*.

Pet. 22 (emphasis in original) (citing Ex. 1003 ¶¶ 140–141). The arguments provided in Petitioner’s Reply are not “new,” and harken back to representations made in the Petition. As such, we are not persuaded the arguments are new, nor are they inconsistent with the Petition.

As such, we find Petitioner's arguments persuasive for the reasons outlined in the Petition, and determine that Petitioner has established that that Jiang meets this limitation of claim 1 for the reasons explained by Petitioner.

- i. Element 1(h): "a search stage for using the calculated distances to identify words within a lexical tree, the lexical tree comprising a model of words"*

Petitioner asserts that Jiang teaches this limitation because Jiang teaches a search stage involving traversing a prefix tree for word identification. Pet. 60. Petitioner asserts the prefix tree of Jiang is an example of the claimed lexical tree, and comprises a model of words as recited. *Id.*

Patent Owner does not offer counterarguments with respect to this element of claim 1. *See* PO Resp. We find Petitioner's arguments persuasive for the reasons outlined in the Petition, and determine that Petitioner has established that that Jiang meets this limitation of claim 1 for the reasons explained by Petitioner.

- j. Proffered Motivation to Combine Jiang and Smyth*

Petitioner asserts that persons of ordinary skill in the art would have been motivated to combine Jiang and Smyth to arrive at the claimed invention. Pet. 16. Petitioner asserts Jiang teaches its speech recognition system may be "practiced with other computer system configurations," including a "multi-processor system" and "programmable consumer electronics." *Id.* (citing Ex. 1004, 4:60–65). Petitioner also asserts that one of skill in the art would have understood that Jiang contemplated modifications to its hardware architecture. *Id.* Petitioner asserts it would be obvious to one of ordinary skill in the art to decrease the immediate

computational load of the system via a distinct processor for distance calculation to provide discrete processors for feature extraction, distance calculation, and searching thereby allowing for segmentation of higher load processing, resulting in a faster, more efficient, and more accurate speech recognition process, which is a goal of speech recognition. *Id.* at 39, 40 (citing Ex. 1003 ¶¶ 178, 179). Further, Petitioner asserts that the addition of an additional processor for distance calculation to the system of Jiang would have had a reasonable expectation of success because Jiang contemplates system modification as discussed above. *Id.* at 40.

Petitioner also asserts that Jiang and Smyth are further amenable to combination because both references use HMMs for phoneme identification. Pet. 41. Building upon this assertion, Petitioner asserts that employing Smyth's HMMs for word identification in the prefix tree of Jiang would have achieved a more accurate threshold scoring, which in turn would realize a more precise pruning of the lexical tree than Jiang currently provides. *Id.* at 51 (citing Ex. 1004, 11:6–15). Petitioner argues that, because in Smyth the HMMs includes transition probabilities between phonemes, the HMM is used for word identification with increased accuracy, and thus it would have been obvious to use HMMs per Smyth in the system of Jiang. *Id.* at 52. Petitioner asserts there would have been a reasonable expectation of success in modifying the HMMs of Jiang per the technique of Smyth because “[g]iven the well-known prevalence at the time of utilizing HMMs for speech recognition, applying *Smyth*’s word HMM techniques to *Jiang* would have been seen as continuing the optimization of speech recognition.” *Id.* at 55.

Patent Owner responds that there would have been no motivation to combine Jiang and Smyth, and that such a combination would not have been

obvious. PO Resp. 13–19. First, Patent Owner reiterates that Jiang uses codewords to identify likelihood scores, and asserts the Petitioner provides no reason why one of ordinary skill in the art would have been motivated to replace the codewords of Jiang with the HMM calculations of Smyth. *Id.* at 13. Patent Owner asserts that the Petition’s sole support for such a specific combination comes from a single paragraph of Mr. Schmandt’s declaration. *Id.* at 13–14 (citing Ex. 1003 ¶ 183). Patent Owner continues that the proffered motivation is insufficient and conclusory. *Id.* at 14–15.

Petitioner responds that the Petition provides “a myriad of support with citations to *numerous* paragraphs. Pet Reply 12–13 (Pet. 40–41; Ex. 1003 ¶ 170–183). Patent Owner responds that while the cited sections of Mr. Schmandt’s declaration could be argued to support the overall combination of Jiang with Smyth, paragraph 183 in the only paragraph that discusses modifying Jiang from a codeword-based approach to a feature vector-based approach. PO Sur-reply 13–16.

We are not persuaded by Patent Owner’s arguments. As discussed above, we determine that Jiang uses codewords, which are representative feature vectors within the scope of claim 1, and those codewords are used to calculate distances. The Petition discusses a motivation to modify Jiang to use Smyth’s DSP device to perform distance calculations (Pet. 40-41), but it does so in the context of “[t]o the extent Patent Owner argues that *Jiang* does not teach calculating distances.” *Id.* at 40. Given that we determine that Jiang does teach performing distancing calculations that meet the claim limitations, Patent Owner’s contention is not persuasive. Rather, we are persuaded by Petitioner’s motivation for using Smyth’s dedicated DSP to perform Jiang’s distance calculations, namely, for the benefit of allowing for segmentation of higher-load processing, resulting in a faster, more efficient,

and more accurate speech recognition process. *Id.* at 39–40 (citing Ex. 1003 ¶¶ 176–180; Ex. 1004, 6:39–42). We agree with Mr. Schmandt, that “[i]t would have been obvious to modify Jiang’s tree search engine, which performs both distance calculation and word determination, to include Smyth’s distinct processors for performing each respective functionality,” because such a modification “results in a more efficient, faster, and more accurate speech recognition process by segmenting higher-load processing tasks.” Ex. 1003 ¶ 180 (citing Ex. 1012, 19, Ex. 1005, 5:49–53, 6:5–8, 6:43–51; Ex. 1004, 4:60–5:3).

Additionally, Patent Owner argues that using Smyth’s probability calculation in the system of Jiang will dramatically increase the computation burden of the calculations, and this is contrary to the purpose of Jiang to reduce computational requirements, as evinced by the use of lossy codewords. PO Resp. 16–19 (citing Ex. 1002 ¶¶ 54–59). Petitioner asserts, with respect to this latter argument, that “the Board need not consider this argument if the Board determines that Jiang teaches limitation 1(g).” Pet. Reply 14. We agree with Petitioner. Given the discussion with respect to limitation 1(g) above, and that Jiang is sufficient to teach or suggest element 1(g), we need not further consider the combination of Smyth and Jiang in the claimed calculation of distances.

Therefore, we determine that the motivation proffered by the Petition is sufficient to combine the teachings of Jiang and Smyth, as disclosed in the Petition.

k. Conclusion

Based on the record presented and for the foregoing reasons, we determine that Petitioner has demonstrated that the combination of Jiang, and Smyth teaches or suggests of all the limitations of independent claim 1

and conclude that a preponderance of the evidence establishes that claim 1 is unpatentable as being obvious.

E. Obviousness of Dependent Claims 2 and 3

Dependent claim 2 recites: *The speech recognition system of claim 1 wherein the time period taken for said identification of spoken words overlaps the time period taken for said second programmable device to calculate one or more distances from a second feature vector.*

Dependent claim 3 recites: *The speech recognition system of claim 2 wherein the first programmable device, the second programmable device, and third programmable device are physically distinct processing circuits.*

Petitioner asserts that the combination of Jiang, Smyth, and Nguyen would have rendered the subject matter of claims 2 and 3 obvious to one of ordinary skill in the art at the time of the invention. Pet. 61. We begin with a discussion of the further Nguyen reference, and then consider arguments with respect to the reference's teachings applied to claims 2 and 3.

Nguyen is directed to a speech recognition system. Ex. 1047, code (57). With respect to speech recognition, Nguyen teaches a similarity measure is computed at step 36 for each acoustic feature vector in a first group of vectors. *Id.* at 3:51–52, 3:61–4:7. After the similarity measure is computed, a search process is performed at step 44. *Id.* at 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.” *Id.* at 4:17–22.

Petitioner asserts that because Nguyen teaches processing feature vectors contemporaneously with the subsequent search process, a POSITA would have understood that Nguyen teaches performing a similarity measure

(i.e., distance calculation) while contemporaneously performing the search step. Pet. 62 (citing Ex. 1003 ¶¶ 219–227). Petitioner also asserts that it would have been obvious to apply the contemporaneous processing of Nguyen to the multiple processing devices of the system of Jiang and Smyth in order to improve system performance to arrive at the features of claim 2. *Id.* at 63.

On the present record, we determine that Petitioner has established that an ordinarily skilled artisan would have understood Nguyen to teach the overlapping processing of different calculations for word identification per claim 2 because Nguyen teaches that during or together with an ongoing acoustic search process, feature vectors are loaded and processed. The combination of Jiang and Smyth has already been argued as teaching three distinct processing circuits above. Petitioner asserts that the combination of Jiang and Smyth teaches the three physically distinct processing circuits of claim 3. Pet. 63.

Additionally, on the present record, we determine that Petitioner has established that an ordinarily skilled artisan would have understood the combination of Jiang and Smyth teaches the three physically distinct processing circuits of claim 3, at least because Smyth asserts each of the feature extractor, classifying processor, and parsing processor may be instantiated with a DSP, which, when separately instantiated teaches three distinct processing circuits per claim 3.

Patent Owner argues that the proposed combination of Jiang and Smyth with Nguyen “runs into a fatal technical obstacle.” PO Resp. 20. Patent Owner argues that pipelining approach of Nguyen, is fundamentally incompatible with feedback-based pruning, where Jiang’s pruning is

necessarily feedback-based pruning. *Id.* at 20–32. Patent Owner argues that:

The ordinary artisan would have known from contemporaneous references that under standard pruning techniques, a pruning determination at the *search stage* for a current frame would be fed back to the *calculation stage* for the next frame to eliminate unnecessary calculations relating only to pruned nodes and branches.

PO Resp. 23 (citing Ex. 2020 ¶¶ 64–67). Patent Owner also argues that ordinarily skilled artisans would have understood that Jiang’s pruning technique requires feedback from the search stage to the calculation stage, and that pruned branches and nodes would no longer be considered in further processing. *Id.* at 25 (citing Ex. 1004, 8:62–64; Ex. 2020 ¶¶ 68–69).

Patent Owner continues that feedback-based pruning cannot be implemented in combination with Nguyen’s alleged teaching of “performing a similarity measure (i.e., distance calculation) while contemporaneously performing the search step,” whereby the search stage for a current frame and the calculation stage for a next frame occur at the same time. PO Resp. 30–31 (citing Pet. 62; Ex. 2020 ¶ 83).

Petitioner responds that Patent Owner’s arguments rely on impermissible bodily incorporation and is based on false assumptions. Pet. Reply 15–27. Petitioner asserts that the combination does not rely on utilizing feedback-based pruning, nor does search-stage pruning, as Jiang teaches, require feeding back the pruned results to the distance calculation stage. *Id.* at 16. Petitioner also asserts that pruning is not required by the combination proposed in the Petition, and that Jiang does not disclose feedback-based pruning. *Id.* at 16–20. Petitioner also argues that the knowledge of those in the art does not proscribe what Jiang discloses, and

that the supporting references cited by Patent Owner disclose techniques for forgoing feedback-bases pruning in favor of parallelism/pipelining. *Id.* at 20–21. In addition, Petitioner asserts that forgoing feedback-based pruning in favor of pipelining is obvious based on the tradeoffs and benefits based on design objectives. *Id.* at 27–29.

Patent Owner responds that Petitioner’s arguments are irrelevant, that Petitioner’s new “one frame delay” approach is not supported by any prior art of record, and that Petitioner’s Reply fails to show that Nguyen could have been combined with feedback-based pruning, and thus concedes the issue. PO Sur-reply 17–19. Patent Owner also argues that the Reply fails to show that the combination does not include or require Jiang’s pruning, and that the pruning disclosed in Jiang is feedback-based pruning. *Id.* at 19–25.

Weighing the evidence proffered by both parties, we are not persuaded that Jiang discloses feedback-based pruning. Petitioner’s declarant disputes that Jiang teaches feedback-based pruning (Ex. 2016, 18:16–19:14), and Jiang itself, within its entire discussion of pruning (Ex. 1004, 8:52–64, 10:65–11:15), never discusses feedback of pruned results. We are left with Patent Owner’s assertions that such feedback-based pruning is implicitly taught by Jiang, and we determine that to be insufficient to require a disclosure that is not explicit. We acknowledge that Patent Owner’s declarant (Ex. 2020 ¶ 64–76) provides testimony, based on the teachings of Jiang and contemporaneous teachings in the art, that Jiang’s pruning requires feedback from the search stage to the calculation stage, but we are not persuaded that Jiang must operate with such feedback. We are also persuaded by Petitioner’s point that Jiang computes a score for each phoneme branch considered in the tree, that contradicts Patent Owner’s theory that scores are computed for only active, non-pruned phoneme

branches of the tree. Pet. Reply 18 (citing Ex. 1004, 8:29–36, 48–49). Lastly, Patent Owner seems to rely, at least partially, that “Jiang never states it does *not* use feedback-based pruning” (PO Sur-reply 25), but an absence of evidence is not evidence of absence. As such, given that we are not persuaded that Jiang must employ feedback-based pruning, we are not persuaded that the methods of Nguyen cannot be used in the combination of Jiang and Smyth, as proffered by the Petition with respect to claims 2 and 3.

In addition, we agree with Petitioner that Jiang discloses that “a pruning technique can be used,” and find that Jiang’s pruning technique is optional. Pet. Reply 16; Ex. 1004, 8:52–64. Although Patent Owner contends that Mr. Schmandt includes Jiang’s pruning as part of his combination theory (PO Resp. 20–22), we agree with Petitioner that the Petition does not rely on Jiang’s pruning technique in mapping Jiang’s process to the search stage of independent claim 1. Pet. Reply 16; *see* Pet. 44–61. We find that the Petition does not rely on the portions of Mr. Schmandt’s testimony that consider Jiang’s pruning.

Further, even under Mr. Schmandt’s combination theory, Mr. Schmandt testifies that the pruning “threshold value can be set at any desired level, based primarily on desired memory and computational savings, and a desired error rate increase caused by memory and computational savings.” Ex. 1003 ¶ 174 (quoting Ex. 1004, 2:25–46). Mr. Schmandt testifies that “if the speech recognition system is programmed to carry more hypotheses forward in the tree (because a relatively lower threshold value is set), the system has a higher recognition accuracy.” *Id.* ¶ 177. Mr. Schmandt testifies that modifying Jiang’s tree search engine 74 to comprise two digital signal processors, where a first processor performs distance calculations and the second performs other speech recognition tasks, yields the benefit of

increasing accuracy by lowering the pruning threshold. *Id.* ¶ 178. Thus, given that using two digital signal processors allows the first processor to increase accuracy by reducing the pruning threshold, and given that the pruning threshold can be set to any desired level based on a desired error rate, a person of ordinary skill in the art would have set the pruning threshold to zero for the benefit of increasing recognition accuracy and lowering the error rate as much as possible. With the pruning threshold set to zero, the combination does not perform pruning, even under Mr. Schmandt's combination theory.

Patent Owner contends that without pruning feedback from the search stage to the calculation stage, the computational load of the calculation stage would approximately double. PO Resp. 28–29 (citing Ex. 2020 ¶¶ 78–79). However, Patent Owner does not provide persuasive evidence or argument to show that the modified system, which uses two processors rather than one to perform the functions of Jiang's tree search engine, would have been unable to handle the increased computational load. *See* Ex. 1003 ¶ 178; *see id.* at ¶¶ 179–182.

Further, Smyth discloses a speech recognition system “specifically configured to recognize certain phrases or words for example a string of digits.” Ex. 1005, 6:61–63. Dr. Anderson testifies that “[i]f you had something that was very small, like a digits recognizer or something for account numbers like . . . [Smyth] was doing, then you might have an abbreviated model. And that's why [Smyth] could do the type of recognition he did.” Ex. 1073, 76:19–23. In addition, Jiang discloses that pruning reduces the computation needed when performing recognition for a large vocabulary. Ex. 1004, 2:19–29. Thus, Jiang itself suggests that pruning would not be necessary when performing recognition for a small

vocabulary. Patent Owner does not persuasively explain why pruning would be required when using Smyth's abbreviated model for recognizing certain phrases or words. Patent Owner further does not provide persuasive evidence or argument showing that when performing recognition using Smyth's abbreviated model without pruning, the computational load of the calculation stage would have exceeded the capability of the processor that performs the distance calculations.

Patent Owner contends that, according to Mr. Schmandt, reducing or eliminating feedback would create the risk of the system calculating distances increasingly behind in real time. PO Sur-Reply 20 (citing Ex. 1003 ¶ 177. However, in the next paragraph, Mr. Schmandt proposes modifying Jiang's tree search engine to comprise two digital signal processors, where one of the processors performs the computationally intensive distance calculations and the other processor performs other speech recognition tasks in order to reduce the pruning threshold. Ex. 1003 ¶ 178. As discussed above, with the increase in processing capability of Petitioner's combination, a person of ordinary skill would have had reason to reduce the pruning threshold to zero for the benefit of increasing recognition accuracy, especially when using Smyth's abbreviated model for certain phrases or words, such as a string of digits.

Based on the record presented and for the foregoing reasons, we determine that Petitioner has demonstrated that the combination of Jiang, Smyth, and Nguyen teaches or suggests of all the limitations of dependent claims 2 and 3 and conclude that a preponderance of the evidence establishes that claims 2 and 3 are unpatentable as being obvious.

F. Obviousness of Dependent Claim 4

Dependent claim 4 recites: *The speech recognition system of claim 3 wherein said physically distinct processing circuits are all formed on a single integrated circuit.* Petitioner asserts that the combination of Jiang, Smyth, Nguyen, and Boike would have rendered the subject matter of claim 4 obvious to one of ordinary skill in the art at the time of the invention. Pet. 65. Boike teaches a “system-on-a-chip (SoC) integrated circuit 10 containing multiple digital signal processors (DSPs).” Ex. 1006, 2:52–54. Boike depicts such DSP subsystems 18–21 as part of SoC 10. *Id.* at Fig. 1.

Petitioner asserts a POSITA would have found it obvious and been motivated to modify the combination of Jiang, Smyth, and Nguyen to include the three distinct processing circuits on a single IC, as taught by Boike because Boike alone expressly motivates the combination, informing a POSITA that there is reduced space and costs and conserved power in forming multiple DSPs on a single IC. Pet. 67 (citing Ex. 1006, 1:19–42; citing Ex. 1003 ¶¶ 231–234). Patent Owner did not separately address this ground in its Response, other than to argue that claim 4 depends from claim 2 and should be patentable based on its dependence on that latter claim. *See* PO Resp. 20.

Based on the record presented and for the foregoing reasons, we determine that Petitioner has demonstrated that the combination of Jiang, Smyth, Nguyen, and Boike teaches or suggests all the limitations of dependent claim 4 and conclude that a preponderance of the evidence establishes that claim 4 is unpatentable as being obvious.

G. Obviousness of Dependent Claim 5

Dependent claim 5 recites: *The speech recognition system of claim 2 wherein the first programmable device and the third programmable device*

comprise one physical uni-processor or multi-processor, and the second programmable device is a physically distinct processing circuit.

Petitioner asserts that the combination of Jiang, Smyth, Nguyen, and Baumgartner would have rendered the subject matter of claim 5 obvious to one of ordinary skill in the art at the time of the invention. Pet. 67. We begin with a discussion of the further Baumgartner reference, and then consider arguments with respect to the reference's teachings applied to claim 5.

Baumgartner is directed to techniques for speech recognition and in particular to techniques for reducing the computational load during speech recognition. Ex. 1007 ¶¶ 3, 6. To reduce computational load during speech recognition, Baumgartner teaches it is desirable for a distance calculation to be performed on a physically distinct device separate from feature vector calculation and word identification. Baumgartner teaches feature vector calculation and word identification performed on first processor core 1910. Pet. 68 (citing Ex. 1007 ¶¶ 31–33, 156–158, Fig. 19). In Baumgartner, distance calculation is performed on a separate, physically distinct processor. *Id.*

As a foundation for applying Baumgartner to the system of Jiang, Petitioner asserts that Jiang teaches a speech recognition system with multiple processors, citing to Figure 2. Pet. 67. Petitioner then asserts that because Jiang already teaches CPU 21 includes multiple processors, it would have been obvious to modify Jiang, to the extent any modification is required, to configure the multiple processors of the CPU to place the computationally intensive processing load of distance calculation on a physically distinct circuit or programmable device as taught by Baumgartner because Baumgartner expressly provides the motivation by teaching such

computing configuration reduces the computational load on the system processors. *Id.* at 70. Petitioner iterates that the combination of Jiang and Smyth is amenable to application of the taught multi-processor speech processing techniques of Baumgartner because the combination of Jiang and Smyth is a system of distinct computing devices. *Id.* at 71.

Patent Owner did not separately address this ground in its Response, other than to argue that claim 5 depends from claim 2 and should be patentable based on its dependence on that latter claim. *See* PO Resp. 20. Based on the record presented and for the foregoing reasons, we determine that Petitioner has demonstrated that the combination of Jiang, Smyth, Nguyen, and Baumgartner teaches or suggests all the limitations of dependent claim 5 and conclude that a preponderance of the evidence establishes that claim 5 is unpatentable as being obvious.

H. Obviousness of Dependent Claim 6

Dependent claim 6 recites: *The speech recognition system of claim 5 wherein said one physical uni-processor or multi-processor and said second programmable device are all formed on a single integrated circuit.*

Petitioner asserts that the combination of Jiang, Smyth, Nguyen, Baumgartner, and Boike would have rendered the subject matter of claim 6 obvious to one of ordinary skill in the art at the time of the invention. Pet. 72. We consider arguments with respect to the references' teachings applied to claim 6.

Petitioner reiterates Boike teaches multiple processors, such as DSPs, formed on a single IC, and Baumgartner expressly teaches a CPU core and speech label accelerator (SLA) reside on the same physical die, thereby teaching the recited. Pet. 72 (citing Ex. 1007 ¶ 180).

Patent Owner did not separately address this ground in its Response, other than to argue that claim 6 depends from claim 2 and should be patentable based on its dependence on that latter claim. *See* PO Resp. 20. Based on the record presented and for the foregoing reasons, we determine that Petitioner has demonstrated that the combination of Jiang, Smyth, Nguyen, and Baumgartner teaches or suggests all the limitations of dependent claim 6 and conclude that a preponderance of the evidence establishes that claim 6 is unpatentable as being obvious.

III. CONCLUSION

In summary, we determine a preponderance of the evidence establishes that claims 1–6 are unpatentable, as shown in the following table:⁹

Claim(s)	35 U.S.C. §	Reference(s)/Basis	Claims Shown Unpatentable	Claims Not Shown Unpatentable
1	103(a)	Jiang, Smyth	1	
2, 3	103(a)	Jiang, Smyth, Nguyen	2, 3	
4	103(a)	Jiang, Smyth, Nguyen, Boike	4	

⁹ Should Patent Owner wish to pursue amendment of the challenged claims in a reissue or reexamination proceeding subsequent to the issuance of this decision, we draw Patent Owner's attention to the April 2019 *Notice Regarding Options for Amendments by Patent Owner Through Reissue or Reexamination During a Pending AIA Trial Proceeding*. *See* 84 Fed. Reg. 16,654 (Apr. 22, 2019). If Patent Owner chooses to file a reissue application or a request for reexamination of the challenged patent, we remind Patent Owner of its continuing obligation to notify the Board of any such related matters in updated mandatory notices. *See* 37 C.F.R. § 42.8(a)(3), (b)(2).

5	103(a)	Jiang, Smyth, Nguyen, Baumgartner	5	
6	103(a)	Jiang in view of Smyth, Nguyen, Boike, Baumgartner	6	
Overall Outcome			1–6	

IV. ORDER

In consideration of the foregoing, it is hereby:

ORDERED that claims 1–6 of the '377 Patent have been shown by a preponderance of evidence to be unpatentable; and

FURTHER ORDERED that, because this is a final written decision, parties to this proceeding seeking judicial review of the Decision must comply with the notice and service requirements of 37 C.F.R. § 90.2.

IPR2023-00035

Patent 10,062,377 B2

For PETITIONER(Apple):

Jennifer C. Bailey,

Adam P. Seitz,

Kevin Rongish

ERISE IP, P.A.

jennifer.bailey@eriseIP.com

adam.seitz@eriseIP.com

kevin.rongish@eriseip.com

For PETITIONER(Amazon):

J. David Hadden

Dargaye Churnet

Saina Shamilov

FENWICK & WEST LLP

dhadden@fenwick.com

dchurnet@fenwick.com

sshamilov@fenwick.com

For PATENT OWNER:

Peter C. Knops

Kayvan B. Noroozi (*Pro Hac Vice*)

NOROOZI P.C.

peter@noroozipc.com

kayvan@noroozipc.com

Katherine Rhoades

Nevin Gewertz

BARTLITBECK L.L.P.

katherine.rhoades@bartlitbeck.com

nevin.gewertz@bartlitbeck.com