

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

APPLE INC.,

Petitioner

v.

ZENTIAN LIMITED,

Case IPR2023-00033

U.S. Patent No. 7,587,319

PETITIONER'S REPLY TO PATENT OWNER'S RESPONSE

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	ZENTIAN PROVIDES NO EVIDENCE OF A LACK OF REASONABLE EXPECTATION OF SUCCESS.....	1
A.	ZENTIAN’S ARTIFICIAL COMPLEXITY REQUIREMENTS ARE UNCLAIMED AND UNDESCRIBED IMPLEMENTATION DETAILS	2
B.	LEVEL OF SKILL.....	5
1.	<i>Zentian Does Not Dispute the Level of Skill</i>	<i>5</i>
2.	<i>Mr. Schmandt’s Level of Skill</i>	<i>6</i>
3.	<i>Dr. Anderson’s Level of Skill</i>	<i>8</i>
4.	<i>Zentian’s “Specialized” Requirements Are Found in the Cited References</i>	<i>9</i>
C.	SKILLED PERSONS KNEW OF VARIOUS SPEECH RECOGNITION HARDWARE ARCHITECTURES	10
III.	ZENTIAN DOES NOT IDENTIFY REASONS FOR NON- OBVIOUSNESS.....	13
A.	MEMORY COLLISIONS AND SLOWER PROCESSING	14
B.	REPROGRAMMING EFFORT	20
C.	ZENTIAN’S MISCELLANEOUS COMPLAINTS.....	22
1.	<i>Zentian’s “Quantitative Analysis” Theory</i>	<i>22</i>
2.	<i>No Additional Processors in the Combination</i>	<i>24</i>
3.	<i>No Desire to Complicate the Recognizer.....</i>	<i>25</i>
IV.	CONCLUSION	26

TABLE OF AUTHORITIES

Cases

<i>Allied Erecting and Dismantling Co. v. Genesis Attachments, LLC</i> , 825 F.3d 1373 (Fed. Cir. 2016)	19
<i>Antor Media Corp.</i> , 689 F.3d 1282 (Fed. Cir. 2012)	13
<i>Corephotonics, LTD. v. Apple Inc., et al.</i> , No. 2020-1961, 2021 WL 4944471 (Fed. Cir. Oct. 25, 2021)	18, 20
<i>Elektta Limited v. ZAP Surgical Systems, Inc.</i> , 81 F.4th 1368 (Fed. Cir. 2023)	20, 23
<i>Facebook, Inc. v. Windy City Innovations, LLC</i> , 973 F.3d 1321 (Fed. Cir. 2020)	5
<i>In re Epstein</i> , 32 F.3d 1559 (Fed. Cir. 1994)	3
<i>In re Keller</i> , 642 F.2d 413 (CCPA 1981)	23
<i>In re Mouttet</i> , 686 F.3d 1322 (Fed. Cir. 2012)	22
<i>In re Nievelt</i> , 482 F.2d 965 (CCPA 1973)	23
<i>In re Publicover</i> , 813 F. App'x. 527 (Fed. Cir. 2020)	3
<i>In re Wands</i> , 858 F.2d 731 (Fed. Cir. 1988)	13
<i>Intel Corp. v. PACT XPP Schweiz AG</i> , 61 F.4th 1373 (Fed. Cir. 2023)	22
<i>Keep Truckin, Inc., v. Innovative Glob. Sys., LLC</i> , No. IPR2020-00694, Paper 33 (PTAB July 21, 2021)	7
<i>KSR Int'l Co. v. Teleflex Inc.</i> , 550 U.S. 398 (2007)	14
<i>Samsung Electronics Co., LTD v. Staton Techiya, LLC</i> , IPR2022-00324, Paper 33 (PTAB July 10, 2023)	3
<i>SEB S.A. v. Montgomery Ward & Co.</i> , 594 F.3d 1360 (Fed. Cir. 2010)	7
<i>Sundance, Inc. v. DeMonte Fabricating Ltd.</i> , 550 F.3d 1356 (Fed. Cir. 2008)	7
<i>TQ Delta, LLC v. CISCO Systems, Inc.</i> , 942 F.3d 1352 (Fed. Cir. 2019)	13, 17
<i>Uber Techs., Inc. v. X One, Inc.</i> , 957 F.3d 1334 (Fed. Cir. 2020)	3
<i>Walmart Inc., et al. v. Power Concepts, LLC</i> , No. IPR2022-00534, Paper 50 (PTAB Sept. 15, 2023)	3

I. INTRODUCTION

Zentian does not dispute the Petition's combination teaches each of the claim elements. Nor does Zentian dispute the Petition's level of skill for a POSITA or that Mr. Schmandt (Apple's expert) qualifies as a POSITA. Zentian instead argues lack of motivation to combine for two reasons. The first is that there is no reasonable expectation of success in the combination because a POSITA would not know how to address tasks stemming from the combination: "Such tasks are not within the skill level of the ordinary artisan in the field of speech recognition, even under the Petition's definition." (Paper 19, *Patent Owner Response*, 16). The second argument comprises assorted complaints regarding the efficacy or required work by a POSITA to make the combination, including alleged memory collisions that would cause the combined system to not operate in real time (*id.* at 20); the combination would not speed up the recognition process (*id.* at 23-25); and the combination would have required significant reprogramming effort (*id.* at 27-28). Zentian's critiques are contrary to legal requirements establishing obviousness and are unresponsive to the Petition's mapping and motivations.

II. ZENTIAN PROVIDES NO EVIDENCE OF A LACK OF REASONABLE EXPECTATION OF SUCCESS

Zentian incorrectly contends a POSITA could never make the proposed combinations and there would be no reasonable expectation of success in the combination because the '319 Patent combines speech recognition and computer

architecture technologies a speech recognition artisan allegedly “would have not been capable of addressing[.]” (Paper 19, 16, 15 (contending the combination “would have been a highly complex undertaking beyond the skill of an ordinary speech recognition artisan”), 6 (arguing a person meeting Petitioner’s level for a POSITA would be a skilled artisan in either speech recognition or “high performance computing and parallel processing” but “would not be both”)).

A. Zentian’s Artificial Complexity Requirements Are Unclaimed and Undescribed Implementation Details

Zentian identifies several issues a POSITA would allegedly need to address in the combination, such as memory collisions, controlling task sharing, etc. (Paper 19, 15). These alleged issues are notably absent from the ’319 Patent Specification and claims, indicating a POSITA would know how to address each.

The Federal Circuit and Board have consistently held a POSITA is assumed to have requisite knowledge to address complications arising from a combination or from how to achieve a claimed limitation, especially when the challenged patent’s specification does not correspondingly address such complications or how to achieve the claimed limitation. For example, the Federal Circuit held that a specification “entirely silent” on how to achieve a claimed step or function “suggest[s] that a [POSITA] was more than capable of selecting between the known methods of accomplishing” the claimed step or function, noting an argument that the prior art is “too sparse to adequately explain to a skilled artisan how to modify [the primary

reference's] system" to achieve the claimed limitation fails when the application is "just as sparse on how a system would" achieve the claimed limitation. *Uber Techs., Inc. v. X One, Inc.*, 957 F.3d 1334, 1339 (Fed. Cir. 2020), *cited by In re Publicover*, 813 F. App'x. 527, 531-32 (Fed. Cir. 2020); *In re Epstein*, 32 F.3d 1559, 1568 (Fed. Cir. 1994) ("Rather, the Board's observation that appellant did not provide the type of detail in his specification that he now argues is necessary in prior art references" supports a finding that a POSITA would have known how to implement the references and would have concluded they were enabling.); *Walmart Inc., et al. v. Power Concepts, LLC*, No. IPR2022-00534, Paper 50, at 31-32, 20-21, 55, 60-61, 115-16, 126 (PTAB Sept. 15, 2023) (rejecting patent owner's contention the combination "would require modifications that would introduce complications discouraging a [POSITA] from making the modifications[,]" noting the challenged patent "does not describe any techniques to address the complications" and finding "[i]n the absence of a specific description of such techniques in the [challenged] patent," addressing the complications was "within the knowledge of a person of ordinary skill"); *Samsung Electronics Co., LTD v. Staton Techiya, LLC*, IPR2022-00324, Paper 33, at 36-38, 29, 42, 63 (PTAB July 10, 2023) (noting prior art references are presumed enabling and include "no implementation details," undermining patent owner's argument the combination would have been beyond a POSITA).

Here, the '319 Patent does not discuss any issues with or techniques for “avoiding memory conflicts, controlling task sharing in an efficient manner, addressing communication bandwidth and latency issues among the various hardware components, and developing a messaging strategy to coordinate information sharing between and within clusters.” (Paper 19, 15). For example, the '319 Patent has a single sentence regarding a solution to potential “bandwidth” limitations, stating “[b]y providing a plurality of processors in a group with a common memory, flexibility in the processing is provided without being bandwidth limited by the interface to the memory that would occur if only a single memory were used for all processors.” *'319 Patent*, 3:50-54. Zentian notes this citation in the '319 Patent (Paper 19, 15), yet does not point to any other discussion alleging techniques to address alleged complexities in the claimed clustered processors. The 3:50-54 disclosure describes no more than what is expressly claimed. No details of how to overcome bandwidth complications, task sharing, or messaging strategies are described in the '319 Patent.

Dr. Anderson admitted specific details regarding memory structures and cache consistency are not required by Claim 1, stating “The implementation details are discussed in the specification. The claim...is not ever expected to have all the implementation details.” (Ex. 2017, *Anderson Depo. Tr.*, 23:6-25 (discussing alleged “difficulty” of coordinating multiple caches), 26:2–27:3 (describing implementation

details regarding memory and caches), 27:9-28:2 (admitting such implementation details are not recited)).

Zentian’s alleged complications are generic, not identifying specifically *how* or *why* memory conflicts, latency issues, or other impediments to the combination would arise. A petitioner is not required to predict and proactively explain every unrecited and generic implementation detail. *Facebook, Inc. v. Windy City Innovations, LLC*, 973 F.3d 1321, 1343 (Fed. Cir. 2020) (affirming Board’s rejection of patent owner’s argument “about the difficulty of physically creating the combination,” as the petitioner was not arguing bodily incorporation). Given the lack of any description in the ’319 Patent of challenges (or solutions thereto) for performing known speech recognition techniques on clustered processors, the Board should determine that Zentian’s generic complications would have been within a POSITA’s skill to address.

B. Level of Skill

1. Zentian Does Not Dispute the Level of Skill

The Parties agree to the Petition’s POSITA skill level. (Paper 10, 8, *citing* Paper 1, 5; Paper 19, 8-9). Zentian does not dispute the level of skill, nor provide an alternative level of skill. (Ex. 2017, 9:20–10:12). Zentian’s declarant, Dr. Anderson, did not dispute the Petition was correct:

Q. Okay. Do you dispute that this is a correct level of skill for the 140 and the 319 patents?

A. I do not.

Id. at 10:9-12. Dr. Anderson opined “the Petitioner put forth a person of ordinary skill in the art definition. **It was reasonable.** Somebody with experience in speech recognition. **I believe somebody with those particular backgrounds could implement the patents at issue, and so those are the ones that I used.**” *Id.* at 9:20-13:4 (pinpoint at 12:1-6), 10:5-8 (Dr. Anderson agreeing the level of skill includes education and experience), 11:12-17, 13:2-4 (opining the level is reasonable, he adopted it, and is “not proposing anything different”)).

The Board considered the level of a POSITA at Institution, determining the combination of *Jiang-Chen* did not exceed the qualifications of a POSITA. (Paper 10, 7-8, 16-17).

2. Mr. Schmandt’s Level of Skill

Though Zentian does not dispute that Mr. Schmandt meets the Petition’s level of skill, Zentian complains that Mr. Schmandt “lacks an adequate basis from his personal knowledge and experience to opine” on the proposed combination because he has not personally built (nor supervised the building of) the combination’s particular architecture for a speech recognition circuit. (Paper 19, 9-12). Whether Mr. Schmandt has built or supervised building the claimed invention does not determine whether Mr. Schmandt meets the level of skill of a POSITA, whether his opinions are credible, or whether he has personal knowledge.

The Board previously held “with respect to the issue of whether an expert has sufficient qualifications to offer admissible testimony, there is no requirement that the expert’s experience be a perfect match to the field of art at issue” or that the expert has experience working on specific projects providing first-hand knowledge of claimed details. *Keep Truckin, Inc., v. Innovative Glob. Sys., LLC*, No. IPR2020-00694, Paper 33, at 12 (PTAB July 21, 2021), citing *SEB S.A. v. Montgomery Ward & Co.*, 594 F.3d 1360, 1373 (Fed. Cir. 2010). “To testify as an expert, an individual does not need to be a person of ordinary skill in the art, but rather must be ‘qualified in the pertinent art.’” *Id.*, citing *Sundance, Inc. v. DeMonte Fabricating Ltd.*, 550 F.3d 1356, 1363-64 (Fed. Cir. 2008). It is sufficient that the expert has the requisite advanced degree and “considerable *technical* experience related to the “[t]echnical [f]ield” of the [] patent[.]” *Id.* at 13. Mr. Schmandt having not built the specifically-claimed speech recognizer does not determine his skill level analysis or qualification to testify.

In fact, Mr. Schmandt was educated in courses covering multiprocessor architectures, parallelism, concurrency, and memory contention during his time at MIT. (Ex. 1003, 109 (Appendix A – Education and Professional Experience at MIT)). Mr. Schmandt also co-authored a paper in 1993 on hardware architectures for speech recognition entitled *Speech Recognition Architectures for Multimedia Environments*. (Ex. 1003, 113 (Appendix A – Publications)). Mr. Schmandt also

authored a textbook including a chapter entitled “Architectures” discussing the various computer architectures for speech recognition. (Ex. 1012, 9).

Further, Zentian did not timely object to Mr. Schmandt’s Declaration and therefore cannot move to exclude the Declaration. *Keep Truckin*, No. IPR2020-00694, Paper 33, at 11.

3. Dr. Anderson’s Level of Skill

Zentian’s arguments recite Dr. Anderson’s experience, or lack thereof. For example, Zentian states Dr. Anderson has “never taught a course that would have equipped a masters-level student in speech recognition to apply a parallel processing architecture like Chen’s to known speech recognition techniques.” (Paper 19, 6). Zentian also provides an anecdotal story of Dr. Anderson’s failed attempts in constructing a parallel computing system for a computer vision project (unrelated to speech recognition) while employed at Texas Instruments. (Paper 19, 1-2, 16-17).

Notwithstanding that Zentian denigrates its own declarant, whether Dr. Anderson has taught a specific course or failed at a parallel processing project is irrelevant to whether a POSITA would have had a reasonable expectation of success. Perhaps Dr. Anderson simply did not have the necessary skills to lead a parallel processing project in an unrelated field. Though Zentian’s argument impugns Dr. Anderson’s credibility, it does not prove a POSITA would lack the knowledge to reasonably succeed in the combination.

4. *Zentian's "Specialized" Requirements Are Found in the Cited References*

Zentian also contends the skilled person would not “have specialized in parallel processing architectures and methods or [sic] high performance computing[.]” (Paper 19, 6). “Specializing” in parallel processing architectures or high performance computing is not required of the POSITA definition nor necessary for a reasonable expectation of success. It is unclear from Zentian’s argument what experience would encompass *specializing*, and Zentian purposefully provides no alternative POSITA definition. (Ex. 2017, 11:12-17 (Dr. Anderson opining he does not have an alternative skill level definition, that the Petition’s level of skill “was a reasonable place to start,” and he “adopted it”)). Regardless, the cited art establishes the skilled person would have been knowledgeable in both high performance computing and parallel processing, as discussed in Section II.C.

Chen expressly states its disclosed embodiments are “**equally applicable to any type of parallel processing computer system**, regardless of the performance levels of the individual processors and memory components that make up such a system.” *Chen*, 11:33-41, 10:14-35 (disclosing the *Chen* invention could be built with commercially available microprocessors and memory chips); Paper 1, 38-40; Ex. 1003, 82-84. *Chen* is thus not only directed to high performance computing. Because *Chen* is analogous art (which Zentian does not dispute), a skilled person seeking to utilize parallel processing techniques for speech recognition would have

been knowledgeable of *Chen* and its applicability to “any type of parallel processing computer system.” *Id.*

Chen, *Thelen* (see Title, Abstract), and *Bailey* (see Abstract) are also directed to parallel processing architectures. Given *Thelen* and *Bailey* are parallel processing architectures for speech recognition and are enabling for their teachings, a POSITA would have certainly had knowledge of parallel processing architectures for speech recognition, contrary to Zentian’s assertions. (Paper 19, 6).

C. Skilled Persons Knew of Various Speech Recognition Hardware Architectures

Zentian asserts a POSITA “would have not been capable of addressing the complex challenges of combining *Chen* with *Thelen* and *Bailey* in the manner the Petition has proposed[.]” (Paper 19, 8, 15-16). The evidence shows the exact issues raised by Zentian were being addressed by skilled artisans.

Evaluating hardware architectures for speech recognition was of significant interest to POSITAs as of the 1990s. Speech recognition researchers recognized issues and tradeoffs of speed, accuracy, power, scalability, monetary cost, ease of building, flexibility, expandability, miniaturization, etc. *See, e.g.*, Ex. 1006, *Hon*, 6 (¶ 2), 17-19; Ex. 2021, *Mathew*, 17 (discussing energy consumption and performance tradeoffs).

For example, Ex. 1006 (*Hon*) from 1991 is entitled *A Survey of Hardware Architectures Designed for Speech Recognition*. *Hon* discusses speech recognition

architectures, including pipelined and shared memory multiprocessor machines. (Ex. 1006, 6, ¶ 2). *Hon* notes “[o]f course, every architecture has its own pros and cons.” *Id.* *Hon*’s stated purpose is to assist future speech recognition researchers: “We would like this document to serve as a future reference for people who may contemplate building new speech architectures or for those who would like to choose among the architecture alternatives that are available.” *Id.*

In other examples, *Ravishankar* explains a “memory bottleneck” impedes speech recognition parallelization, which may be overcome through techniques such as load balancing and/or processing state selection. (Ex. 1007, 92, 111). Mr. Schmandt’s own textbook provides a brief discussion of several known hardware computing platforms for speech recognition, including the parallel-processing Sphinx system. (Ex. 1012, *Schmandt*, 151-152).

A POSITA would have been aware of survey articles like *Hon* discussing available hardware and programming tools for resolving complications such as memory conflicts in speech recognizers. The “hypothetical ‘person having ordinary skill in the art’ to which the claimed subject matter pertains would, of necessity have the capability of understanding the scientific and engineering principles applicable to the pertinent art[.]” *Ex parte Hiyamizu*, 10 USPQ2d 1393, 1394 (Bd. Pat. App. & Inter. 1988). A speech recognition POSITA would have knowledge to address “trivial” issues arising in the combination, at least because the references expressly

disclose skilled persons were building speech recognizers utilizing parallel processing architectures. (Ex. 1003, 56-60; Ex. 2017, 137:21–143:10 (Schmandt opining making the combination is “trivial” and within the POSITA’s knowledge); Ex. 2021, 11 (discussing parallelization techniques for the Sphinx recognizer)).

Zentian also does not respond to or directly dispute any of Mr. Schmandt’s motivations for the combination, instead generically arguing Mr. Schmandt did not identify any “tasks” or “challenges” in the combination or how the POSITA “would manage” the parallel operation of the combination. (Paper 19, 8-9, 21-22). Zentian’s assertions are so amorphous as to be meaningless.

Dr. Anderson also does not provide any supporting factual bases, instead simply repeating the attorney argument from the Patent Owner Response. *See, e.g.*, Ex. 2020, 19, 24, 29-30, 51-53. Dr. Anderson’s conclusory, verbatim opinions should be given no weight. *TQ Delta, LLC v. CISCO Systems, Inc.*, 942 F.3d 1352, 1358-59 (Fed. Cir. 2019) (“This court’s opinions have repeatedly recognized that conclusory expert testimony is inadequate to support an obviousness determination on substantial evidence review.”).

Finally, Zentian does not raise a non-enablement argument nor provide a *Wands* factors analysis. *In re Wands*, 858 F.2d 731, 737 (Fed. Cir. 1988). The references are presumed enabling. *Antor Media Corp.*, 689 F.3d 1282, 1288 (Fed. Cir. 2012) (discussing reference presumed enabling). *Chen* already teaches an

operable clustered processor architecture; the combination simply requires utilizing the architecture for speech recognition according to *Thelen's* acoustic model hierarchy and employing *Bailey's* parallel speech recognizers. Zentian's alleged complications are specific to a clustered processor architecture and would occur regardless of the type of information being processed (e.g., for speech recognition or other uses). *Chen* further directly envisions accounting for memory latency. *See, e.g., Chen*, 11:25-32, 12:1-13, 18:24-42. Therefore, addressing memory conflicts or bandwidth latency, for example, are issues *Chen* is already operable to address because *Chen* is enabling.

III. ZENTIAN DOES NOT IDENTIFY REASONS FOR NON-OBVIOUSNESS

Zentian does not argue that the combination is inoperable, could not be done, renders the art unsatisfactory for its intended purpose, or otherwise directly respond to or rebut the motivations in the Petition and Declaration (Ex. 1003). Instead, Zentian's complaints are essentially there would be issues the skilled person would need to work through in the combination. The Federal Circuit has held that a POSITA "is also a person of ordinary creativity, not an automaton." *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 421 (2007). "[I]n many cases a person of ordinary skill will be able to fit the teachings of multiple patents together like pieces of a puzzle." *Id.* at 420. Here, fitting the puzzle pieces together appropriately includes thoughtfully combining the grouped processor structure from *Chen* with the *Thelen-*

Bailey system “to achieve a compatible **improvement** in processing speed and power[.]” (Paper 1, 44-45 (emphasis added), *citing* Ex. 1003, 84-85).

A. Memory Collisions and Slower Processing

Zentian asserts the combination would “require multiple processors in one of Chen’s clusters to simultaneously request recognition model data from one cluster memory at the same time,” thereby creating “memory contention problems” or “memory collisions” that would slow down real-time processing. (Paper 19, 28, 17-18, 20-21, 23; Ex. 2020, 45; *cf.* Paper 1, 40-46 (providing numerous motivations to combine), *citing* Ex. 1003, 84-85, *Hon*, 19-20). Zentian never explains why such alleged problems would be present in the combination but not the clustered-processor speech recognition system of Claim 46, nor does Zentian acknowledge that the claims do not recite real-time recognition. (Paper 19, 2, 8, 20-21, 30). Regardless, Zentian incorrectly assumes that a POSITA would have thoughtlessly appended processors to the system, ignoring the proper skill, hardware and software, calibration, and testing that a POSITA would have applied in combining the systems as mapped. Zentian ignores that the Petition and Mr. Schmandt provided numerous reasons supporting a reasonable expectation of success for the combination, such as the well-known nature of the techniques and hardware involved. (Paper 1, 21-22, 37-38, 43-46, 49-50, 52-53; Ex. 1003, 74, 79, 81, 84, 89).

Zentian admits data access was a known problem. (Paper 19, 18). *Chen*, which pre-dates the '319 Patent by ten years, includes lengthy discussion of addressing memory access for inter-memory clusters. *Chen* is enabling for what it teaches, including a clustered computing system that balances the tradeoffs of parallel processing and memory access. *Chen*, 5:47-53, 14:39–18:42 (providing lengthy discussion of inter-cluster memory access, including variable delays for flexibility), 18:7-16 (teaching options for configuration “in a way best suited for the particular user environment, as well as the types of parallel programming task performed by the parallel processing computer system”). Zentian’s conclusory argument that memory collisions would occur and thereby slow down the processing does not account for the combination having *multiple* processors now performing the speech recognition. (Paper 1, 39-43). The *multiple* processors for each cluster are performing the word recognition across all the lexical trees at the same time (i.e., in parallel) for the lexical data associated with that cluster (i.e., the specific *Thelen* recognition model (e.g., sports) targeted to a specific part of the huge vocabulary). (Paper 1, 23-25, 37, 41, *citing Thelen*, 7:17-58 (teaching multiple recognizers each “independently recognize the same speech input at almost the same moment”). Per *Chen*, these *multiple* processors would more than account for any memory contention issues, which are well known to occur in parallel processing systems. *Chen*, 5:47-53 (“Thus, the present invention accommodates parallel processing

performance first and flexibility of the interconnection of processors and memory second, with the tradeoff between these two considerations balanced at the point at which the performance of inter-cluster memory accesses start to significantly degrade.”); Cf. Paper 1, 35, 44-45, *citing* Ex. 1003, 81, 84-85, pinpoint at 84 (opining regarding reducing duplication of lexical data).

For example, *Chen* is primarily directed to addressing memory latencies occurring *between* clusters (**inter**-cluster memory access), not *within* a cluster (**intra**-memory cluster access). *Chen*, 4:17-28, 4:44-49, 8:42-47. *Chen* states inter-cluster memory access will have higher latencies than intra-memory cluster access due to the physical connection and separation (i.e., distance) of the clusters. *Id.* at 4:58-63. Yet, even in addressing **inter**-cluster memory access, *Chen* teaches that its clustered architecture “**does not have to sacrifice performance** for the increased interconnection necessary to increase the size of the problem solution space provided by an extended shared memory model...[.]” *Id.* at 5:33-41, 2:24-29. That *Chen* “does not have to sacrifice performance” across *multiple clusters of processors* indicates *Chen* certainly does not have to sacrifice performance *within a single cluster*, per Dr. Anderson’s conclusory assertion. (Ex. 2020, 45). Because *Chen* is enabled for what it teaches, *Chen*’s system would have accounted for and resolved any potential memory collision or contention problems to not have to “sacrifice performance[.]” *Chen*, 5:33-41. Because the combination utilizes *Chen*’s multiple clusters of

processors with each cluster having its own memory, the combination is similarly operable and would not have sacrificed performance.

Dr. Anderson also does not evaluate any techniques known to ameliorate memory collisions. For example, a larger memory cache size is known to reduce memory contention. *See* Ex. 2021, 9-10 (employing cache optimizations to address memory latencies and noting that “processor to memory speed ratios increasing in the future”).

Instead, Dr. Anderson provides only conclusory opinions that should be given little weight. (Ex. 2020, 44-45; *TQ Delta*, 942 F.3d at 1358-59). For example, Dr. Anderson’s only basis for his opinion references a single *Mathew* sentence about a multi-threaded HMM stage having high synchronization overhead. (Ex. 2021, 11; Ex. 2020, 46). *Mathew*’s teaching here is unhelpful and inapplicable to whether memory collisions would slow down a clustered processor architecture per the combination at least because *Mathew* proposes multiple processors for the HMM stage only. In contrast, the present combination employs clustered processors all being collectively used for all three stages of the speech recognition (feature extraction, distance calculation, and search stage). (Paper 1, 2-5). Additionally, *Mathew*’s single-sentence disclosure does not explain what is meant by high synchronization overhead, if techniques were employed to address, or other advantages gained such that tradeoffs were made. Given *Chen*’s lengthy analysis of

memory conflicts arising in a clustered processor architecture, as claimed, *Mathew's* extremely limited discussion of its dissimilar architecture would not serve to dissuade a POSITA from the combination.

Zentian's attack on obviousness because of alleged memory collisions also fails to recognize that design decisions often entail tradeoffs between simultaneous advantages and disadvantages, which does not obviate a motivation to combine references. *Corephotonics, LTD. v. Apple Inc., et al.*, No. 2020-1961, 2021 WL 4944471, Slip. Op., at 6 (Fed. Cir. Oct. 25, 2021) ("it is a commonplace fact that design decisions entail making tradeoffs among multiple objectives") (not reported in Fed. Rptr.), *citing Allied Erecting and Dismantling Co. v. Genesis Attachments, LLC*, 825 F.3d 1373, 1381 (Fed. Cir. 2016) ("[a] given course of action often has simultaneous advantages and disadvantages, and this does not necessarily obviate motivation to combine"). Such a tradeoff analysis between the cost and speed of a speech recognition system is laid out plainly in the Petition. (Paper 1, 41-46). Mr. Schmandt's declaration testimony that *Chen's* clustered processors utilizing commercially-available processors strikes a balance between cost and processing power motivations known to a POSITA is sufficient to support a motivation of the combination. (Ex. 1003, 84). Mr. Schmandt identified a predictable result from the combination, and Zentian does not dispute the predictable result. *Id.*

Lastly, Zentian's assertion that memory problems would arise from *Bailey's* mention of "private memory" is unresponsive to the Petition's mapping relying on *Chen's* memory structure and should be dismissed. (Paper 19, 12-15; Paper 1, 2-5 (providing an overview of the combination "utiliz[ing] *Chen's* architecture"), 40 ("The proposed modification to *Thelen* is grouping *Thelen's* recognizers 331-333 in groups of recognizers (where each recognizer has a processor as already taught by *Thelen*) and connecting each group of recognizers (and associated processors) to a memory, **all of which is taught by *Chen*** and mapped for Claim 46(d)."); Ex. 1003, 83.

B. Reprogramming Effort

Zentian argues no motivation for the combination because of an alleged “significant reprogramming effort” of porting a software program from a shared memory model to a distributed or private memory model. (Paper 19, 26-28, *citing Chen*, 2:6-9). Zentian’s analysis misunderstands *Chen* and the mapped architecture model, misapplying *Chen*’s teachings. When considered in context, *Chen*’s discussion of a “significant reprogramming effort” explains that Zentian’s cited challenges regarding distributed memory models are intended to be **solved** by *Chen*’s architecture, not **caused** by *Chen*’s architecture.

A POSITA building a speech recognizer would have to undertake reasonable programming efforts no matter the recognizer’s architecture, so the POSITA would not need to perform additional **re**programming to achieve the claimed combination. Additionally, even if the combination included a reprogramming effort, such does not negate obviousness nor is finding the optimal approach required. *Elekta Limited v. ZAP Surgical Systems, Inc.*, 81 F.4th 1368, 1374 (Fed. Cir. 2023). Depending on the goals and tradeoffs of the POSITA, the clustered processor architecture provides increased benefits over a single-processor architecture, which Zentian does not rebut. *Corephotonics*, No. 2020-1961, Slip. Op., at 6.

Additionally, Zentian is technically incorrect in applying *Chen*’s teachings. *Chen* defines both shared and distributed memory models, stating: “In a distributed

memory model, data is stored in the private local memory of each processor and is communicated among processors by some type of message passing scheme. In a shared memory model, data is stored in a common shared memory that is equally accessible to all processors in the parallel processing system.” *Id.* at 1:56-62. At *Chen*, 2:6-9, a prior art architecture is discussed that is a parallel processing system with a *shared* memory model. *Chen*, 1:66–2:1. *Chen* explains shared memory models are easier to program than distributed memory models in some respects. *Id.* at 2:1-29. One reason is that in a distributed memory model, the program is divided into separate tasks that are small enough to fit on each processor’s local memory. *Id.*

Zentian states *Thelen* is a shared memory model, whereas the *Chen* embodiment mapped in the Petition “relies on multiple private [distributed] memories, with one memory per cluster, in which access to those private memories is extended also to directly adjacent clusters.” (Paper 19, 27). Here, Zentian is incorrect that for each cluster’s memory, access is extended to adjacent clusters (which would then form a distributed memory model among the clusters). In the mapped combination, each cluster’s memory stores a particular portion (context) of the entire lexicon according to *Thelen*’s hierarchy, e.g., cluster 1’s memory stores the sports lexicon, cluster 2’s memory stores the health lexicon, etc. (Ex. 1003, 83, 75-76). To perform speech recognition, each cluster does not need to continually

access the other clusters' memories; a core vocabulary replicated in each local memory eliminates such a need. *Jiang*, 7:40-43 (discussing “overlap” vocabulary).

Thus, in the mapped combination, there are multiple, clustered processors, with each cluster of processors sharing a single memory. This architecture is a shared memory model per *Chen's* definition and does not require any significant reprogramming effort. *Chen*, 5:20-25 (“[a] shared memory model can be used with programs that can be executed in the cluster shared memory of a single cluster”). Therefore, contrary to Zentian's assertions, the software program would not need to be divided into separate tasks across each cluster's memory and would not require significant reprogramming effort.

C. Zentian's Miscellaneous Complaints

1. Zentian's “Quantitative Analysis” Theory

Zentian argues that Mr. Schmandt did not perform a “quantitative analysis” of real processors and speech recognition systems available at the time of the '319 Patent. (Paper 19, 19-23). Obviousness does not require building the combination and testing its efficacy: “It is well-established that a determination of obviousness based on teachings from multiple references does not require an actual, physical substitution of elements.” *In re Mouttet*, 686 F.3d 1322, 1332 (Fed. Cir. 2012); *Intel Corp. v. PACT XPP Schweiz AG*, 61 F.4th 1373, 1381 (Fed. Cir. 2023) (finding that petitioner need not show an improvement but rather a “suitable option”).

Additionally, “it is not necessary that the inventions of the references be physically combinable to render obvious the invention under review.” *Id.*, citing *In re Keller*, 642 F.2d 413, 425 (CCPA 1981); see also *In re Nievelt*, 482 F.2d 965, 968 (CCPA 1973) (“Combining the teachings of references does not involve an ability to combine their specific structures.”). Nor is it necessary that the combination be the most optimal, desirable, or preferred. *Elekta*, 81 F.4th at 1374.

Mr. Schmandt explained a POSITA would have been motivated to utilize *Chen*’s clusters of simple (cheaper) commercially-available processors to improve (reduce) the cost of computing resources, thereby enabling a larger amount of computing resources for a given cost. (Paper 1, 39-40; Ex. 1003, 83). Any “quantitative analysis” of real processors and speech recognition systems available at the time of the ’319 Patent would not only have been arbitrary (which specific processors?), but also irrelevant to the Petition’s financially-focused motivation. Additionally, the Petition and supporting declaration provide ample support that a POSITA would have been motivated by the flexibility and scalability provided by *Hon* to utilize *Chen*’s processing clusters in *Thelen-Bailey*’s speech recognition system, regardless of respective speeds and/or costs of the systems. (Paper 1, 45; Ex. 1003, 58, 84, citing *Hon*, 19-20).

2. *No Additional Processors in the Combination*

Zentian also contends the combination would not add “processing power” beyond *Thelen* alone, citing Mr. Schmandt’s deposition questioning regarding whether the combination with *Chen* would add more processors than what *Thelen* alone teaches. (Paper 19, 23-25). This argument is not responsive to the Petition’s mapping and proffered motivations. As Mr. Schmandt reiterated in his deposition, the combination provides the advantage of “reducing the duplication of lexical data,” necessitating “less memory to operate, further reducing the overall financial cost of the system” and “increas[ing] the processing power[.]” (Ex. 1003, 84, 83 (discussing the combination “store[s] the overlapping lexical data in fewer unique locations[.]” thereby allowing transfer of the lexical data to/from the lexical memory “for processing more efficiently” and “reproduced fewer times”); Paper 1, 43-45; Ex. 2017 122:13–123:19 (opining *Thelen* is “not an optimally efficient processor”), 137:2-20, 138:2-24, 140:8–141:14). The combination’s motivation to combine is not contingent on increasing processing power via adding more processors beyond what is taught in *Thelen*, which was Zentian’s line of questions during Mr. Schmandt’s deposition. (Ex. 2017, 121:3–136:6, pinpoint at 132:6-17 (Mr. Schmandt opining that “distributing among multiple processors would be an increase in performance and speed” in multi-threaded single-processor systems). **None** of Zentian’s arguments in the Patent Owner Response address or rebut the advantage noted by

Mr. Schmandt of sharing of lexical data and reducing duplication of lexical data, though Zentian’s counsel asked Mr. Schmandt about it during deposition. (Paper 1, 40-46; Ex. 1003, 83-84; Ex. 2017, 122:13–125:7).

3. *No Desire to Complicate the Recognizer*

Zentian also uses an out-of-context excerpt from an answer by Mr. Schmandt to assert incorrectly that Mr. Schmandt testified “it is never desirable to complicate a speech recognition system with additional hardware or software when that system can already perform adequately.” *Id.*, citing Ex. 2017, 163:4-12. Mr. Schmandt’s *entire* answer includes an initial discussion regarding the undesirability of overcomplicating systems for a **small vocabulary recognizer**, such as a voice dialer with a limited address book. (Ex. 2017, 163:4–164:22). Mr. Schmandt goes on to explain that in **large vocabulary systems** (such as *Thelen’s*, see Paper 1, 24-25, citing *Thelen*, 7:30-35), one would employ “multiple processors to get real time performance because there’s a heck of a lot of names that I’d have to search for to compare against the acoustic data....” *Id.*

The Petition established a known problem in the art regarding the need for high-speed processing at an affordable cost, and the combination achieves a low-cost large vocabulary speech recognition system that is higher-speed than *Thelen* alone. (Paper 1, 41-46). Thus, the Board should maintain its finding at Institution that “a person of skill in the art would have considered *Chen’s* architecture desirable

as a substitution for that of the *Thelen-Bailey* combination because, at the very least, it leverages the power of multiple processors available for use within each of *Thelen's* recognition contexts.” (Paper 10, 22-23, *citing* Ex. 1003, 84).

IV. CONCLUSION

For the reasons provided above and in the Petition and supporting evidence, Petitioner Apple Inc. requests cancellation of the Challenged Claims.

Respectfully submitted,

ERISE IP, P.A.

BY: /s/ Jennifer C. Bailey

Jennifer C. Bailey Reg. No. 52,583

Adam P. Seitz, Reg. No. 52,206

7015 College Blvd., Suite 700

Overland Park, KS 66211

P: (913) 777-5600

F: (913) 777-5601

jennifer.bailey@eriseip.com

adam.seitz@eriseip.com

COUNSEL FOR PETITIONER
APPLE INC.

APPENDIX OF EXHIBITS

Exhibit 1001	U.S. Patent No. 7,587,319 (“ <i>319 Patent</i> ”)
Exhibit 1002	Prosecution History for the 7,587,319 Patent (“ <i>319 File History</i> ”)
Exhibit 1003	Declaration of Christopher Schmandt
Exhibit 1004	U.S. Patent No. 6,374,219 to Jiang (“ <i>Jiang</i> ”)
Exhibit 1005	U.S. Patent No. 5,428,803 to Chen et al. (“ <i>Chen</i> ”)
Exhibit 1006	Hsiao-Wuen Hon, <i>A survey of hardware architectures designed for speech recognition</i> , Carnegie Mellon University, 1991 (“ <i>Hon</i> ”)
Exhibit 1007	Ph.D. Thesis of Mosur Ravishankar (“ <i>Ravishankar</i> ”)
Exhibit 1008	<i>Left Intentionally Blank</i>
Exhibit 1009	U.S. Patent No. 5,983,180 to Robinson (“ <i>Robinson</i> ”)
Exhibit 1010	U.S. Patent No. 5,036,539 to Wrench Jr., et al. (“ <i>Wrench</i> ”)
Exhibit 1011	Frederick Jelinek, <i>Statistical Methods for Speech Recognition</i> , The MIT Press, 1997 (“ <i>Jelinek</i> ”)
Exhibit 1012	Christopher Schmandt, <i>Voice Communication with Computers</i> , Van Nostrand Reinhold, 1994 (“ <i>Schmandt</i> ”)
Exhibit 1013	Lawrence Rabiner and Biing-Hwang Juang, <i>Fundamentals of Speech Recognition</i> , Prentice Hall PTR, 1993 (“ <i>Rabiner</i> ”)
Exhibit 1014	Richard Klevans and Robert Rodman, <i>Voice Recognition</i> , Artech House, 1997 (“ <i>Klevans</i> ”)
Exhibit 1015	U.S. Patent No. 7,120,582 to Young et al. (“ <i>Young</i> ”)
Exhibit 1016	John Holmes and Wendy Holmes, <i>Speech Synthesis and Recognition</i> , 2nd Edition, Taylor & Francis, 2001 (“ <i>Holmes</i> ”)
Exhibit 1017	U.S. Patent No. 4,926,488 to Nadas et al. (“ <i>Nadas</i> ”)
Exhibit 1018	U.S. Patent No. 6,182,037 to Maes (“ <i>Maes</i> ”)
Exhibit 1019	U.S. Patent Application Publication No. 2003/0220796 to Aoyama et al. (“ <i>Aoyama</i> ”)
Exhibit 1020	U.S. Patent No. 6,092,045 to Stubbley et al. (“ <i>Stubbley</i> ”)
Exhibit 1021	U.S. Patent No. 6,151,574 to Lee et al. (“ <i>Lee</i> ”)
Exhibit 1022	U.S. Patent No. 6,580,814 to Ittycheriah et al. (“ <i>Ittycheriah</i> ”)
Exhibit 1023	U.S. Patent No. 6,452,348 to Toyoda (“ <i>Toyoda</i> ”)
Exhibit 1024	William A. Wulf and C.G. Bell, <i>C.mmp—A multi-mini-processor</i> , Carnegie-Mellon University, 1972 (“ <i>Wulf</i> ”)
Exhibit 1025	Lee D. Erman, Richard D. Fennell, Victor R. Lesser, and D. Raj Reddy, <i>System Organizations for Speech Understanding: Implications of Network and Multiprocessor Computer Architectures for AI</i> , IEEE Transactions on Computers, Vol. C-25, No. 4, April 1976 (“ <i>Erman</i> ”)

Exhibit 1026	U.S. Patent No. 6,393,481 to Deo et al. (“ <i>Deo</i> ”)
Exhibit 1027	U.S. Patent No. 6,615,338 to Tremblay et al. (“ <i>Tremblay</i> ”)
Exhibit 1028	U.S. Patent No. 5,922,076 to Garde (“ <i>Garde</i> ”)
Exhibit 1029	Lawrence R. Rabiner, <i>A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition</i> , Proceedings of the IEEE, Vol. 77, No. 2, February 1989 (“ <i>Rabiner 89</i> ”)
Exhibit 1030	U.S. Patent No. 6,526,380 to Thelen et al. (“ <i>Thelen</i> ”)
Exhibit 1031	U.S. Patent No. 5,459,798 to Bailey et al. (“ <i>Bailey</i> ”)
Exhibit 1032	Zentian, Ltd. v Apple Inc. WDTX Waco (6:22-cv-122) Scheduling Order “ <i>Scheduling Order</i> ”
Exhibit 1033	Declaration of June Ann Munford (“ <i>Munford</i> ”)
Exhibit 1034	U.S. Patent No. 6,076,056 to Huang et al. (“ <i>Huang</i> ”)
Exhibit 1035	U.S. Patent No. 4,831,550 to Katz (“ <i>Katz</i> ”)
Exhibit 1036	U.S. Patent No. 4,567,606 to Vensko et al. (“ <i>Vensko</i> ”)
Exhibit 1037	U.S. Patent No. 4,907,278 to Cecinati et al. (“ <i>Cecinati</i> ”)
Exhibit 1038	B. I. (Raj) Pawate and Peter D. Robinson, “ <i>Implementation of an HMM-Based, Speaker-Independent Speech Recognition System on the TMS320C2x and TMS320C5x</i> ”, 1996 (“ <i>Pawate</i> ”)
Exhibit 1039	U.S. Patent No. 4,386,416 to Giltner et al. (“ <i>Giltner</i> ”)
Exhibit 1040	U.S. Patent No. 4,571,700 to Emry, Jr. et al. (“ <i>Emry</i> ”)
Exhibit 1041	U.S. Patent No. 4,646,261 to Ng (“ <i>Ng</i> ”)
Exhibit 1042	U.S. Patent No. 5,150,399 to Yasuda (“ <i>Yasuda</i> ”)
Exhibit 1043	Declaration of Christina Canino
Exhibit 1044	Deposition Transcript of David V. Anderson, Ph.D. of November 17, 2023

CERTIFICATION OF WORD COUNT

The undersigned certifies pursuant to 37 C.F.R. § 42.24 that the foregoing Petitioner's Reply to Zentian's Response, 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 5,554 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: January 9, 2024

/s/ Jennifer C. Bailey

Jennifer C. Bailey, Reg. No. 52,583

ATTORNEY FOR PETITIONER

CERTIFICATE OF SERVICE ON ZENTIAN
UNDER 37 C.F.R. § 42.6

Pursuant to 37 C.F.R. § 42.6(e), the undersigned certifies that on January 9, 2024, the foregoing *Petitioner Apple Inc.'s Reply to Patent Owner's Response* was served via electronic filing with the Board and via Electronic Mail on the following practitioners of record for Zentian:

Peter C. Knops (peter@noroozipc.com)
Kayvan B. Noroozi (kayvan@noroozipc.com)
Katherine Rhoades (katherine.rhoades@bartlitbeck.com)
Nevin Gewertz (nevin.gewertz@bartlitbeck.com)

/s/ Jennifer C. Bailey

Jennifer C. Bailey, Reg. No. 52,583

ATTORNEY FOR PETITIONER