

Filed on Behalf of NETWORK-1 TECHNOLOGIES, INC.

By: Charles R. Macedo (Reg. No. 32,781)
Brian A. Comack (Reg. No. 45,343)
Amster, Rothstein & Ebenstein LLP
90 Park Avenue
New York, NY 10016
Telephone: (212) 336-8074
cmacedo@arelaw.com
N1-Google-IPR@arelaw.com

Gregory Dovel (admitted *pro hac vice*)
Dovel & Luner, LLP
201 Santa Monica Blvd., Suite 600
Santa Monica, CA 90401
Telephone: (310) 656-7066
greg@dovellaw.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE INC.
Petitioner

v.

NETWORK-1 TECHNOLOGIES
Patent Owner

Cases IPR2015-00343, IPR2015-00345, IPR2015-00347, and IPR2015-00348
Patents 8,640,179, 8,205,237, 8,010,988, and 8,656,441

DECLARATION OF DR. GEORGE KARYPIS

NETWORK-1 EXHIBIT A2005
Google Inc. v. Network-1 Technologies, Inc.
IPR2015-00348

Table of Contents

I.	Background to my opinions in this Declaration.....	1
A.	Expertise.....	1
B.	Assignment.....	3
C.	Approach.	4
D.	Understanding of the law.	5
II.	Summary of the IPR Patents and asserted art.....	8
A.	The IPR Patents.....	8
1.	‘237 patent (Ex. 1001 ‘237 IPR).....	8
2.	‘988 patent (Ex. 1001 ‘988 IPR).....	10
3.	‘179 patent (Ex. 1001 ‘179 IPR).....	11
4.	‘441 patent (Ex. 1001 ‘441 IPR).....	13
B.	The asserted art.....	16
1.	Overview of Ghias—Ex. 1010 (addressed in the ‘237, ‘988, ‘179, and ‘441 IPRs).	16
2.	Overview of Iwamura—Ex. 1012 (addressed in the ‘237 and ‘988 IPRs).	19
3.	Overview of Conwell—Ex. 1009 (addressed in the ‘179 and ‘441 IPRs).	22
4.	Overview of Philyaw—Ex. 1014 (addressed in the ‘179 and ‘441 IPRs as a secondary reference).....	24
5.	Overview of Chen—Ex. 1008 (addressed in the ‘237 IPR as a secondary reference).	25
III.	General Findings.....	27
IV.	General concerns with the IPR Petitions and Dr. Moulin’s Declarations (Exs. 1004 in each IPR).....	30
V.	Claim Constructions.	38
A.	sub-linear (‘237 patent).	40
1.	the words used in the construction: “size of the dataset”	41
2.	‘237 specification.....	43
3.	Petitioner’s Declarant.....	47
B.	non-exhaustive search (‘237, ‘988, ‘179, and ‘441 patents).....	54

Declaration of George Karypis

1.	The Board’s preliminary construction of “non-exhaustive search” is consistent with the understanding of one of ordinary skill in the art: “a search that locates a match without a comparison of all possible matches.”	54
2.	The Board properly rejected Petitioner’s assertion that a “non-exhaustive search” should be construed as “a search that locates a match without conducting a brute force comparison of all possible matches, and all data within all possible matches.” ...	56
C.	neighbor search / identifying a neighbor / neighbor / near neighbor (‘237, ‘988, ‘179, and ‘441 patents).....	58
D.	approximate nearest neighbor search (‘237 patent).	60
1.	“identifying a close match that is not necessarily the closest match”	61
2.	“sublinear”	62
VI.	‘237 patent.	65
A.	‘237 Ground 1: The instituted claims of the ‘237 patent are not anticipated by Iwamura.	65
1.	sub-linear time search (claims elements 1(b) and 5(b.2)).....	66
2.	approximate nearest neighbor search (claim elements 9(b) and 13(b.2)).....	85
3.	nonexhaustive search (claim element 25(b)).	95
4.	identify a neighbor / near neighbor (claims elements 1(b), 5(b), and 25(b)).....	126
5.	sublinear approximate nearest neighbor search (claim element 33(b)).....	132
B.	‘237 Ground 2: The instituted claims of the ‘237 patent are not anticipated by Ghias.	134
1.	sublinear time search (claim elements 1(b) and 5(b.2)).	134
2.	approximate nearest neighbor search (claim elements 9(b) and 13(b.2)).....	155
C.	‘237 Ground 3: The instituted claims of the ‘237 patent are not obvious over Iwamura and Chen.....	166
VII.	‘988 patent.	168

Declaration of George Karypis

A.	‘998 Ground 1: The instituted claims of the ‘988 Patent are not anticipated by Ghias.	168
1.	non-exhaustive search (claim element 15(b)).	169
2.	search identifying a neighbor (claim element 15(b)).	186
3.	determining an action based on the identification (claim element 15(c)).	190
B.	‘988 Ground 2: The instituted claims of the ‘998 patent are not obvious over Ghias.	193
C.	‘988 Ground 3: The instituted claims of the ‘998 patent are not anticipated by Iwamura.	194
1.	non-exhaustive search (claim 15(b)).	194
2.	identifying a neighbor (claim 15(b)).	203
VIII.	‘179 patent.	204
A.	‘179 Ground 1: The instituted claims of the ‘179 patent are not anticipated by Conwell.	204
1.	neighbor search (claims 1, 13, 25).	205
2.	non-exhaustive search (claims 1, 13, 25).	231
B.	‘179 Ground 2: The instituted claims of the ‘179 Patent are not obvious in view of Ghias and Philyaw.	244
1.	non-exhaustive search (claims 1, 13, 25).	245
2.	neighbor search (claims 1, 13, 25).	251
IX.	‘441 Patent.	256
A.	‘441 Ground 1: The instituted claims of the ‘441 Patent are not anticipated by Conwell.	256
1.	neighbor search (claims 1, 13, 25).	256
2.	non-exhaustive search (claims 1, 13, 25).	259
B.	‘441 Ground 2: The instituted claims of the ‘441 Patent are not obvious over Ghias and Philyaw.	260
1.	non-exhaustive search (claims 1, 13, 25).	261
2.	neighbor search (claims 1, 13, 25).	264

I, George Karypis, declare:

I am making this Declaration at the request of Patent Owner Network-1 Technologies, Inc. in the following *Inter Partes* Reviews of U.S. Patent Nos. 8,205,237 ('237 patent), 8,010,988 ('988 patent), 8,640,179 ('179 patent), and 8,656,441 ('441 patent) (collectively the "IPR Patents"):

- IPR2015-00345 ('237 patent),
- IPR2015-00347 ('988 patent),
- IPR2015-00343 ('179 patent), and
- IPR2015-00348 ('441 patent),

(collectively the "IPRs"), all initiated by petitioner Google Inc. ("Petitioner").

I. Background to my opinions in this Declaration.

A. Expertise.

1. I am a Professor in the Department of Computer Science and Engineering at the University of Minnesota. I hold a Ph.D. in Computer Science from the University of Minnesota, granted in 1996. I began my post-graduate school career as a Research Associate in my current department. I became an Assistant Professor in 1999, an Associate Professor in 2004, and a Professor in 2009. I teach courses in Algorithms and Data Structures, Parallel Programming, and Data Mining, among other subjects.

2. I am a member of the Editorial Board of a number of academic journals, and I have chaired a number of academic conferences.¹ I am a co-author of the books *Introduction to Parallel Computing*, and *Introduction to Parallel Computing: Design and Analysis of Algorithms*. I am an author of more than 80 published journal papers, and more than 115 published conference papers.²

¹ Representative academic conferences include:

- Program Committee co-Chair of the ACM Recommender Systems Conference (RecSys'13), Hong Kong, China (2013);
- Program Committee co-Chair of the 13th International Conference on Data Mining (ICDM), Dallas, Texas (December 2013); and
- Program Committee Co-Chair of the International Conference on Data Science and Advanced Analytics (DSAA 2014), Shanghai, China, (November 2014).

² Representative papers include:

- “L2Knng: Fast Exact K-Nearest Neighbor Graph Construction with L2-Norm Pruning” David C. Anastasiu and George Karypis, 24th ACM International Conference on Information and Knowledge Management (CIKM), Melbourne, Australia (2015).

3. I have also developed a number of software systems for a variety of functions, including software for analyzing high-dimensional data sets. A copy of my curriculum vitae is attached as Exhibit A to this Declaration. It contains a more complete listing of my professional activities and background.

B. Assignment.

4. I have been retained by Patent Owner Network-1 Technologies, Inc. as a technical consultant. I am being compensated for my time at my standard consulting rate of \$350 per hour. I am not receiving any compensation that depends on the outcome of the IPRs.

5. This declaration addresses the validity of:

-
- “*L2AP: Fast Cosine Similarity Search with Prefix L-2 Norm Bounds*” David Anastasiu and George Karypis, 30th IEEE International Conference on Data Engineering (ICDE), pp. 784—795 (2014).
 - “*Comparison of Descriptor Spaces for Chemical Compound Retrieval and Classification*” Nikil Wale and George Karypis, IEEE International Conference on Data Mining (ICDM), pp. 678—689 (2006).
 - “*Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering*” Ying Zhao and George Karypis, Machine Learning, 55, pp. 311-331 (2004).

- claims 1, 3–5, 7–9, 11–13, 15, 16, 21–27, 29, 30, 33, 34, 35, 37, and 38 of the ‘237 patent;
- claims 15–17, 21–28, 31–33, 51, and 52 of the ‘998 patent;
- claims 1-3, 6, 8–14, 18, 19, 21–27, 29–31, and 34–37 of the ‘179 patent; and
- claims 1–3, 6, 8–14, 18, 19, 21–27, 29, and 30 of the ‘441 patent.

C. Approach.

6. To develop my opinions, I have read:

- the four IPR Patents (the ‘237, ‘988, ‘179, and ‘441 patents);
- the four Petitions for *Inter Partes* Reviews;
- the exhibits accompanying the Petitions, including the four Declarations of Dr. Pierre Moulin (Exs. 1004 in each IPR);
- the four Decisions Instituting the IPRs; and
- the testimony of Dr. Pierre Moulin, dated August 19-20, 2015 (Ex. 2006).³

³ In this Declaration, I identify the specific Petition, Declaration, and Decision that I am citing by including the corresponding patent abbreviation in a parenthetical. For example, I refer to the Petition addressing the ‘237 patent as Pet. (‘237) at X; and the Moulin Declaration addressing the ‘179 patent as Moulin Decl. (‘179) ¶X. Because there is only one Dr. Moulin Deposition transcript for all

7. In addition, I relied on my personal knowledge and experience with both research and development in the technology underlying the IPR Patents and the art asserted against the IPR Patents.

D. Understanding of the law.

8. My understanding regarding the law as applicable to this Declaration is based on my discussions with counsel. I have included in the text of my Declaration quotations from or references to certain legal cases or statutes that were provided to me by counsel to provide me with an understanding of the relevant law.

E. Person of ordinary skill in the art.

9. Through my education, experience and training, in academia and industry, and my analysis of the IPR Patents, I am familiar with the knowledge of a person of ordinary skill in the field of the IPR Patents at the time of invention in 2000.

10. For the purposes of this Declaration, I am of the opinion that a person of ordinary skill in the art with respect to the IPR Patents is a person with a Bachelor's degree in computer science, mathematics, or a similar discipline and two to three years of relevant experience, or a graduate degree in the same area.

four IPRs (Ex. 2006), I simply refer to Dr. Moulin's deposition testimony as Moulin Depo. Z.

In determining what would be the level of ordinary skill in the field as of the 2000 time frame, I considered the following:

- (a) the educational level of the inventor, Ingemar J. Cox (it is my understanding that Dr. Cox has a bachelor's degree in electronics and computer science from University College London (1980) and a Ph.D. from Oxford (1983));
- (b) the type of problems encountered in the art—*i.e.*, how to identify a digital work without modifying the work (*see e.g.*, '237, 1:30-36);
- (c) the prior art solutions to those problems (*see e.g.*, '237, 1:37-4:4, and the prior art asserted by the Petitioner in the IPRs addressing related problems involving searching, matching, and identifying melodies, audio files, and other digital files within databases—Conwell, Ghias, Iwamura, Chen, and Philyaw);
- (d) the rapidity with which innovations are made (based on my observations over the past 20 plus years, major innovations in content identification occur about every 5 to 10 years);
- (e) the sophistication of the technology (developing content identification solutions is a moderately sophisticated technology); and

- (f) the educational level of workers in the field (workers in the field generally had have at least a bachelor's degree in computer science, mathematics or a similar discipline and at least two to three years of relevant experience).

11. Based on these factors, it is my conclusion that a person of ordinary skill in the art at the time would have been a person with a Bachelor's degree in computer science, mathematics, or a similar discipline and two to three years of relevant experience, or a graduate degree in the same or related area.

12. I note that Dr. Moulin suggests that the person of ordinary skill in the art "would have been highly skilled, and typically would have possessed at least an M.S. in computer science, electrical engineering, or mathematics; knowledge of video and audio processing techniques; and 1-2 years of experience in audio, video, or image processing." *See e.g.*, Moulin Decl. ('237) ¶7; Pet. ('237) at 4. Dr. Moulin's opinion as to the person of ordinary skill in the art is similar to mine with respect to the degrees and years of experience, but I note that: (1) Dr. Moulin does not provide any rational underpinnings for his opinion; and (2) the phrase "highly skilled" used by Dr. Moulin in his description is a relative term and Dr. Moulin does not provide the context for this phrase.

II. Summary of the IPR Patents and asserted art.

13. In this Declaration:

- I use the term “work” to mean the item (*e.g.*, a digital audio or image file) to be identified using the search (*see e.g.*, ‘237, 6:51-56; ‘988, 7:17-20; ‘179, 6:18-21; ‘441, 6:49-52);
- I use the term “record” to mean one of the units in the reference database that the extracted features of the work may be compared to (*see e.g.*, ‘237, 6:16-20; ‘988, 6:46-50; ‘179, 6:21-24; ‘441, 6:15-18); and
- I use the term “database,” “data set,” or “library” to mean the collection of all records to be searched (*see e.g.*, ‘237, 6:23-30; ‘988, 6:50-60; ‘179, 6:30-36; ‘441, 6:24-30).

A. The IPR Patents.

14. Each IPR Patent (the ‘237, ‘179, ‘988, and ‘441 patents) involves a search that compares features from a given work to records in a reference database of potential matches to identify an action to be taken.

1. ‘237 patent (Ex. 1001 ‘237 IPR).

15. The independent claims of the ‘237 patent include the following elements:

[1] receiving or obtaining features extracted from a work;

[2] identifying the work using the extracted features to perform a search of the database, where the search is:

- a sub-linear time search to identify a neighbor (claims 1 and 5);
- an approximate nearest neighbor search (claims 9 and 13);
- a non-exhaustive search ... to identify a near neighbor (claim 25); or
- a sublinear approximate nearest neighbor search (claim 33); and

[3] either (i) transmitting information about the identified work to the client device, or (ii) determining an action based on the identity of the work.

16. The invention claimed in the '237 patent includes two key features:

17. Feature 1: Although the language varies among the claims, each claim requires that the “identifying” be performed based on a search that has two properties:

(1) a sub-linear or non-exhaustive property (reflected in the underlined language):

- sub-linear time search ... to identify a neighbor (claims 1 and 5);
- approximate nearest neighbor search (claims 9 and 13);
- non-exhaustive search ... to identify a near neighbor (claim 25); and
- sublinear approximate nearest neighbor search (claim 33).

(2) a neighbor property (reflected in the underlined language):

- identify a neighbor (claims 1 and 5);
- approximate nearest neighbor search (claims 9 and 13);
- non-exhaustive search ... to identify a near neighbor (claim 25); and
- sublinear approximate nearest neighbor search (claim 33).

18. Feature 2: The system must either determine an “action” based on the identification (claims 25 and 33); or transmit information about the identified media work to a “client device” (claims 1, 5, 9, and 13). It is not sufficient to simply identify a match. Rather, an action must also be identified or information about the identified work must be transmitted to the client device.

2. ‘988 patent (Ex. 1001 ‘988 IPR).

19. The independent claims of the ‘988 patent include the following elements:

- [1] extracting features from a work;
- [2] identifying the work based on the extracted features by performing “a non-exhaustive search identifying a neighbor;”
- [3] determining an action based on the identity of the work; and
- [4] performing the action.

20. The invention claimed in the ‘988 patent includes two relevant distinguishing features:

- (1) the “identifying” must be performed using a “non-exhaustive search identifying a neighbor;” and
- (2) the system must “determin[e] an action” and “perform[] the action” based on the identity of the work. It is not sufficient to identify a match. Rather, “an action” associated with the match must be “determin[ed]” and “perform[ed].” ‘988, claim 15.

21. I note that the Board did not institute trial for independent claim 1 of the ‘988 patent and any claims dependent on claim 1. Accordingly, I do not address these claims in this Declaration.

3. ‘179 patent (Ex. 1001 ‘179 IPR).

22. The independent claims of the ‘179 patent (claims 1, 13, and 25) include the following five elements for identifying a work and performing a corresponding action:

- [1] a database comprising: (a) electronic representations of works; and (b) electronic data related to an action corresponding to works;
- [2] obtaining extracted features of an unknown work;
- [3] identifying the unknown work by comparing the extracted features and electronic representations using a “non-exhaustive neighbor search;”

[4] determining an appropriate action based on the electronic data related to
an action; and

[5] associating the determined action with the identified work.

‘179, claims 1, 13, and 25.

23. The claimed steps are illustrated in Figure 1:

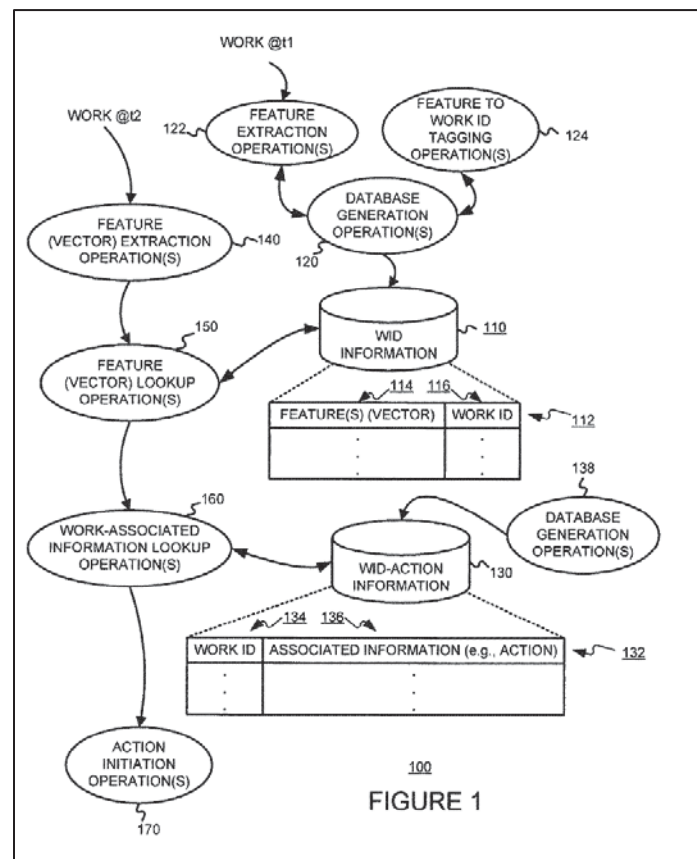


Figure 1 illustrates (“for work @t2”):

- “feature (vector) extraction operation(s)” (140) that extract features from the work (‘179, 6:45-47);

- “feature (vector) lookup operation(s)” (150) that identify the work by searching for a matching feature vector (‘179, 6:50-52);
- “work-associated information lookup operation(s)” (160) that retrieve(s) associated information, such as an action (‘179, 6:55-58); and
- “action initiation operation(s)” (170) that perform(s) some action based on the associated information (‘179, 6:58-60).

24. The invention claimed in the ‘179 patent includes two relevant distinguishing features:

- (1) the “identifying” must be performed by comparing the extracted features to the electronic representations using a “non-exhaustive neighbor search;” and
- (2) the system must determine or associate an “action” based on the identified work. It is not sufficient to simply identify a match. Rather, “an action” associated with the match must be “determined” or “associated.”

‘179, claims 1, 13, and 25.

4. ‘441 patent (Ex. 1001 ‘441 IPR).

25. The independent claims of the ‘441 patent (claims 1, 13, and 25) include the following five elements for identifying a work and performing a

corresponding action:

- [1] a database with (a) first data related to records, and (b) second data related to action information corresponding to the records;
- [2] extracting features from a work;
- [3] identifying the work by comparing the extracted features and the data related to the records using “a non-exhaustive neighbor search;”
- [4] determining an action based on the identity of the electronic work; and
- [5] performing the action.

‘441, claims 1, 13, and 25.

26. The claimed steps are illustrated in Figure 1:

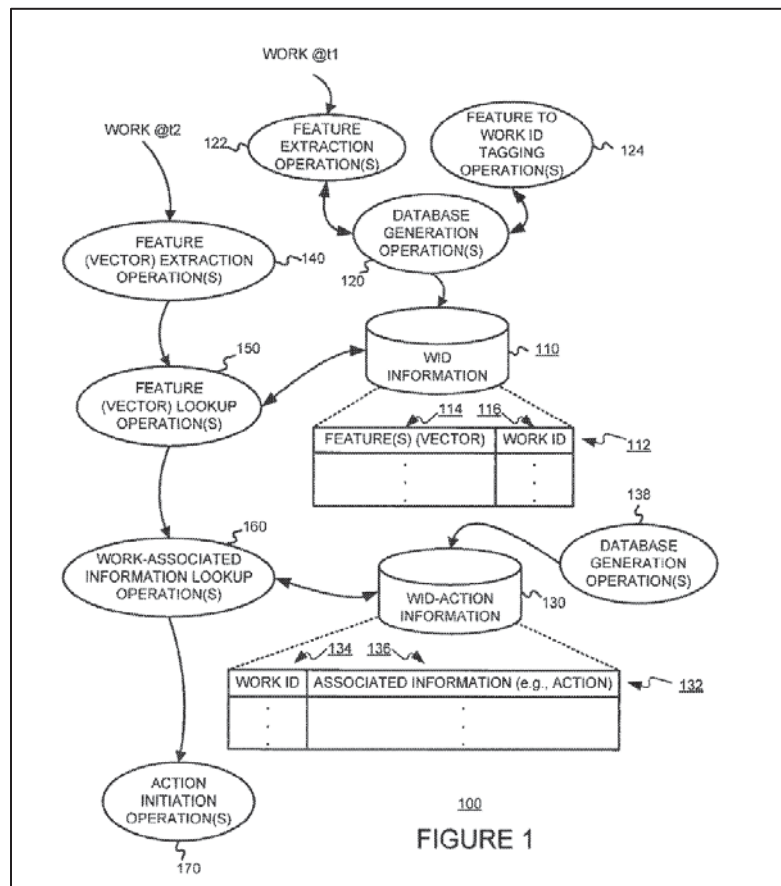


Figure 1 illustrates (“for work @t2”):

- “feature (vector) extraction operation(s)” (140) that extract(s) features from the work (‘441, 6:39-41);
- “feature (vector) lookup operation(s)” (150) that identify the work by searching for a matching feature vector (‘441, 6:44-48);
- “work-associated information lookup operation(s)” (160) that retrieve(s) associated information, such as an action (‘441, 6:49-51); and
- “action initiation operation(s)” (170) that perform(s) some action based on the associated information (‘441, 6:52-54).

27. The invention claimed in the ‘441 patent includes two relevant distinguishing features:

- (1) the “identifying” must be performed by comparing the extracted features to the electronic representations using a “non-exhaustive neighbor search;” and
- (2) the system must determine or associate an “action” based on the identified work. It is not sufficient to simply identify a match. Rather, “an action” associated with the match must be “determined” or “associated.”

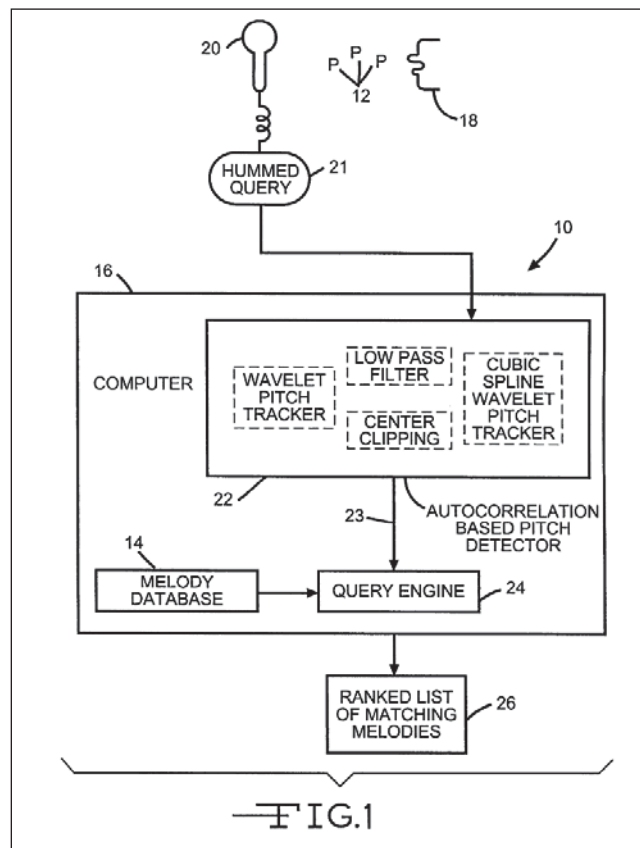
‘441, claims 1, 13, and 25.

B. The asserted art.

28. The four IPRs address three primary references and two secondary references. I address each reference in turn, starting with the primary references and then turning to the secondary references.

1. Overview of Ghias—Ex. 1010 (addressed in the ‘237, ‘988, ‘179, and ‘441 IPRs).

29. Ghias (Patent No. 5,874,686) discloses “an apparatus [for] searching melodies.” Ghias, Abstract. As illustrated in Figure 1 of Ghias, a “tune 12 is hummed by a person 18 into a microphone 20.” Ghias, 2:41-42.



The data from the microphone is fed into “a pitch tracking module 22 in computer 16” which extracts “a contour representation” of the melody (23). Ghias, 2:41-50. The computer uses a “query engine 24” which “searches the melody database 14.” Ghias, 2:50-52. The disclosed search can produce a ranked list of matching melodies—“ranked by how well they matched the query” (Ghias, 6:60-63) as illustrated at 26.

30. As I explain below in detail, all searches disclosed in Ghias are linear (not sub-linear) with respect to the size of the data set being searched. In addressing “the problem of approximate string matching,” Ghias identifies the running times of several algorithms:

Several Algorithms have been developed that address the problem of approximate string matching. Running times have ranged from $O(mn)$ for the brute force algorithm to $O(kn)$ or $O(n \log(m))$, where “ O ” means “on the order of,” m is the number of pitch differences in the query, and n is the size of the string (song). See Ricardo Baeza-Yates and G. H.

Ghias, 6:23-28. In each identified instance, the running time of the search is not sub-linear with respect to the data set. As clarified in this passage from Ghias (and as I address in detail below):

- “ m is the number of pitch differences in the query” corresponding to the length of the query (highlighted in green in the passage above); and
- “ n is the size of the string (song)” corresponding to the size of a record being searched (highlighted in orange in the passage above).

31. The disclosed searches may be sub-linear with respect to the length of the query being searched “ m ... the number of pitch differences in the query.” Specifically, the referenced search with a running time of $O(n \log(m))$ is sublinear with respect to “ m ” because it is a function of $\log(m)$). The disclosed searches, however, are never sub-linear with respect to “ n ...the size of the string (song)” or the size of the data set (N) (*i.e.*, the number of songs to be compared). Rather, the search time will grow linearly with each additional song to be searched and the length of the song.

32. Also as I describe in detail below, the searches disclosed in Ghias are exhaustive rather than “nonexhaustive.” The “query engine 24” compares the

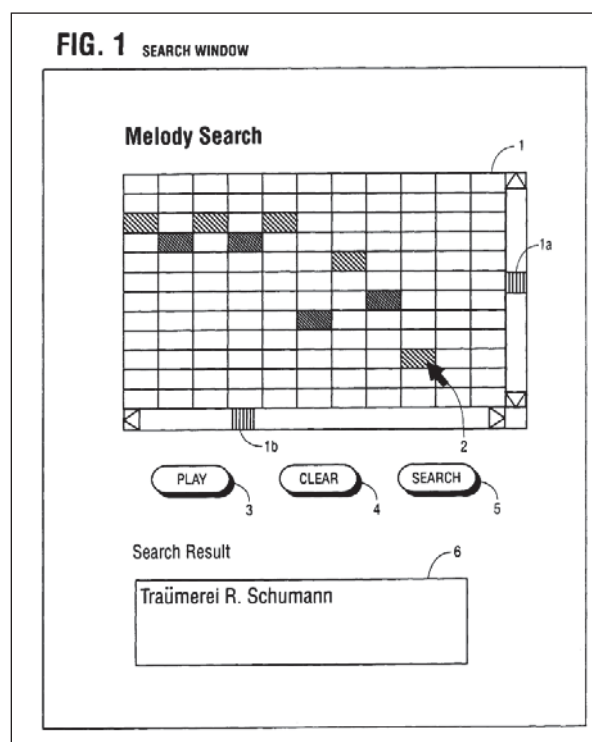
work (user input 23) to “all the songs” in the melody database 14 (the library).

Ghias, 5:66-6:2. After searching all possible matches, the system “output[s] a ranked list of approximately matching melodies.” Ghias, 2:50-53.

33. Finally, as I describe in detail below, the searches disclosed in Ghias are not “neighbor” searches because the searches always necessarily identify the exact or closest match—they are guaranteed to identify an exact match or the closest match. Ghias does not identify any search in which an exact or the closest match is not guaranteed to be identified.

2. Overview of Iwamura—Ex. 1012 (addressed in the ‘237 and ‘988 IPRs).

34. Iwamura (Patent No. 6,188,010) discloses a “method to enable one to search for a song title when only its melody is known.” Iwamura, Abstract. “A remote music database with melody information is searched for the melody entered by the user, using for example, a peak or differential matching algorithm.” Iwamura, Abstract. Figure 1 illustrates “an example of a search interface” (Iwamura, 2:45-46):



35. Iwamura discloses a searching algorithm that is designed to be more efficient than alternatives by matching up peak notes from the work to be identified with the peak notes of the records in the database when comparing the notes from the work to be identified with the notes in the records. “Peak notes are also detected and marked when the data base is built.” Iwamura, 6:59-60. “A fast search is performed by using a peak or differential matching algorithm.” Iwamura, 12:1-2.

36. As I explain in detail below, the search disclosed in Iwamura is exhaustive rather than the claimed “non-exhaustive,” “sublinear,” or “approximate nearest neighbor” search. While the individual comparisons of a work and a record in the library can be more efficient using the “peak note” approach

disclosed in Iwamura (search speed can be increased), in doing so each record in the library is searched as part of the disclosed algorithm and “[t]he reference melody that gives the least difference is returned as a search result.” Iwamura, 7:53-55.

37. Moreover, the Boyer-Moore algorithm referenced in Iwamura searches “word by word from the beginning of the database to the end.” Iwamura, 9:51-55. As a result, as I explain in detail below, while the Boyer-Moore algorithm may be sublinear with respect to the length of the query (the work to be identified),⁴ it is not sub-linear with respect to the relevant size of the dataset being searched.

⁴ If the query pre-processing step of the Boyer-Moore algorithm is included as part of the execution time, then the algorithm may be linear in terms of the length of the query. If the query is used repeatedly, however, the pre-processing execution time will only be incurred once. One can think of concatenating all the database strings to give as an aggregate length of n . If “ m ” is the query length, then the worse-case complexity is $\Theta(m) O(n)$, which is linear with respect to both the database n and the query length m .

3. Overview of Conwell—Ex. 1009 (addressed in the ‘179 and ‘441 IPRs).

38. Conwell (Patent No. 6,970,866) discloses associating media content, such as MP3 files, with identifiers and URLs. Conwell, Abstract. As illustrated in Figure 3, the identifiers (*e.g.*, “034”) are associated with corresponding URLs (*e.g.*, “www.sonymusic.com/catalog/05634.html”):

034	www.sonymusic.com/catalog/05634.html
112	www.sonymusic.com/catalog/00014.html
198	www.supertracks.com/index/artists/taylor.htm
376	www.emusic.com/0555353x.pdf
597	www.cdw.com/music/featured_CDs/index.html
612	www.sonymusic.com/catalog/00231.html
850	www.polygram.com/franklin/adf_234.htm
921	www.loudeye.com/rap/1999/46755646.html

FIG. 3

39. Conwell discloses two approaches to identifying a work: (1) assigning identifiers, or (2) implicitly generating identifiers derived from the data using a hashing algorithm. Conwell, Abstract. The implicit approach (rather than explicitly assigning identifiers) is the approach relevant to the IPR Petitions. *See* Pet. (‘179) at 22-23.

40. Conwell relies on hashing algorithms to extract features from the work to be identified:

“The identifiers can be assigned, or can be implicit (*e.g.*, derived from other data in the content object, as by hashing).”

Conwell, Abstract; 1:65-67 (“some or all of the content data is processed by a hashing algorithm to yield a 128 bit identifier corresponding to that content.”) As I explain in detail below, Conwell compares the hashed extracted features of the work to be identified to the features of the records in the database exclusively using an exact match comparison.

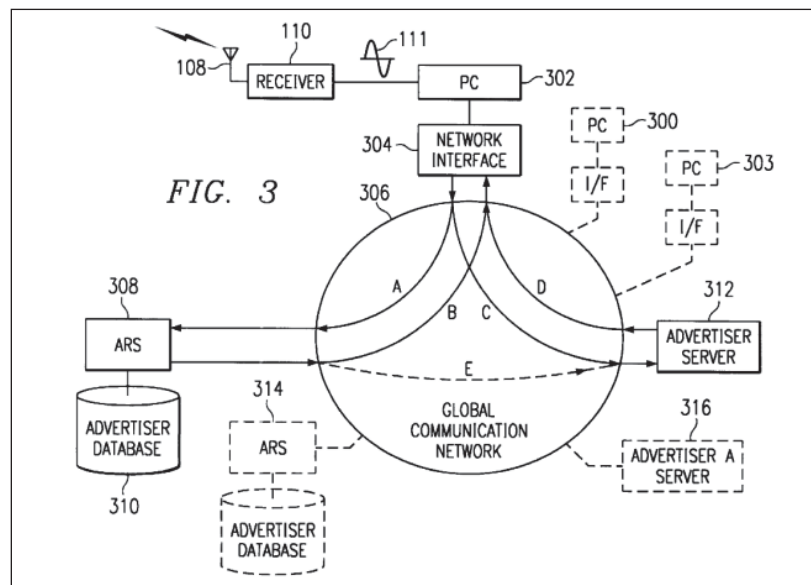
41. When implementing the search in Conwell to identify a match, comparing the hashed extracted features from a work with a record in the library using the disclosed lookup table (Conwell, 3:43-45) produces a binary result: either (1) there is an exact match; or (2) there is no exact match. *See* Conwell, Figure 3. Whether the hash of the extracted features of the work and the hashed extracted features of the record being compared are close (similar) or distant (dissimilar) is not considered, is not relevant, and cannot be determined using a nearness comparison of the hashes extracted in Conwell.

42. In addition, as I explain below in detail, while Conwell discloses using a sorted lookup table to store the hashed extracted features (*see* Conwell Figure 3, 3:43-45), Conwell does not identify any specific algorithm for performing the exact match comparison using the lookup-table and therefore does not identify either an exhaustive search or a non-exhaustive search. Either

approach could theoretically be used; however, neither approach is expressly or inherently disclosed.

4. Overview of Philyaw—Ex. 1014 (addressed in the ‘179 and ‘441 IPRs as a secondary reference).

43. Philyaw (Patent No. 6,098,106) discloses a system that uses identifying information embedded into either the sound or video portion of a broadcast signal to view corresponding information. Philyaw, Abstract; 1:66-2:8. “Figure 3 illustrates the system interactions over a global network” (Philyaw, 2:20-21):



44. Rather than using the searches claimed in the IPR patents to identify a work by comparing extracted features to a database of potential matches, the system in Philyaw embeds a “routing signal having routing information contained therein” into a broadcast program to identify what is being broadcast:

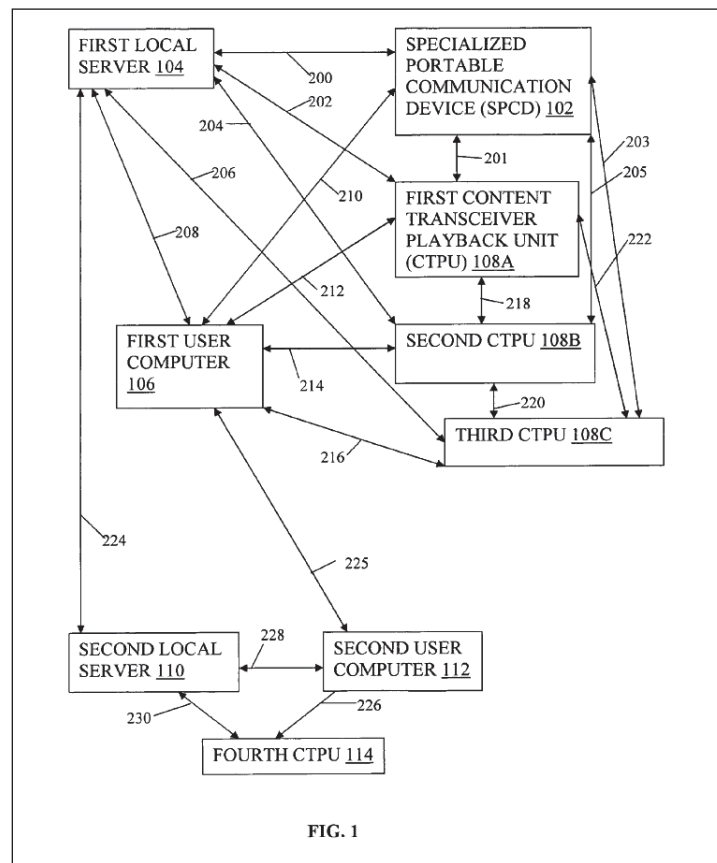
“A program is broadcasted having embedded therein a routing signal having routing information contained therein. The routing signal is then extracted from the broadcast. Thereafter, a personal computer is controlled to allow a user to retrieve the information from a storage region at the defined location, which defined location is located with the extracted routing information, providing it at the personal computer for use by the user.”

Philyaw, 2:1-9. Accordingly, Philyaw does not disclose any searching algorithm.⁵

5. Overview of Chen—Ex. 1008 (addressed in the ‘237 IPR as a secondary reference).

45. Chen (Patent No. 7,444,353) discloses a system for identifying music in a “music and information delivery” system. Chen, Abstract. Figure 1 “illustrates one embodiment of a music and information delivery system according to the teachings” of Chen. Chen, 4:7-9.

⁵ I note that the ‘179 and ‘441 IPR Petitions do not assert that Philyaw discloses the claimed “non-exhaustive neighbor search” (from the ‘179 and ‘441 claims) but instead exclusively relies on Ghias for this element in Ground 2 of the ‘179 and ‘441 IPRs (the only grounds in which Philyaw is asserted). *See* Pet. (‘179) 47-48, 51; Pet. (‘441) 47-48; 51.



A user can identify music that they can then access, for example, by downloading to a laptop or home computer. Chen, 1:58-66.

46. The system in Chen can “search a storage medium to identify and access the piece of music from the storage medium.” Chen, Abstract. But Chen does not provide any details as to how any search is performed. Accordingly, Chen does not disclose a(n):

- sub-linear time search to identify a neighbor (‘237, claims 1 and 5);
- approximate nearest neighbor search (‘237, claims 9 and 13);
- non-exhaustive search ... to identify a near neighbor (‘237, claim 25); or

- sublinear approximate nearest neighbor search (claim 33).

I note that neither the Petition nor the corresponding Declaration relies on Chen for these claimed elements but instead relies on Iwamura in Ground 3 of the ‘237 IPR, the only ground in which Chen is at issue. *See* Pet. (‘237) at 54-57.

III. General Findings.

47. Based on my analysis of:

- (a) the IPR Patents (the ‘237, ‘988, ‘179, and ‘441 patents);
- (b) the art asserted against the IPR Patents in the four IPRs;
- (c) the IPR Petitions;
- (d) Dr. Moulin’s Declarations (Exs. 1004 in each IPR) and deposition testimony (Ex. 2006);
- (e) the documents cited in the IPR Petitions;
- (f) the IPR Decisions; and
- (g) the other documents referenced in this Declaration,

I am of the opinion that the instituted claims of the IPR Patents⁶ are not unpatentable based on the grounds at issue in the IPRs.

⁶ I understand that the instituted claims are claims 1, 3–5, 7–9, 11–13, 15, 16, 21–27, 29, 30, 33, 34, 35, 37, and 38 of the ‘237 patent; claims 15–17, 21–28, 31–

48. Specifically, for the reasons and based on the analysis that I set forth below, I am of the opinion that:

‘237 patent:

- Ground 1: Claims 1, 3–5, 7–9, 11–13, 15, 16, 21–25, 29, 30, 33, 37, and 38 of the ‘237 patent are not anticipated by Iwamura under 35 U.S.C. § 102(e);
- Ground 2: Claims 1–3, 5–7, 9–11, 13–15, and 21–24 of the ‘237 patent are not anticipated by Ghias under 35 U.S.C. § 102(b);
- Ground 3: Claims 26, 27, 34, and 35 of the ‘237 patent are not obvious over Iwamura and Chen under 35 U.S.C. § 103;

‘998 patent:

- Ground 1: Claims 15–17, 21–23, 28, 31, and 51 of the ‘998 patent are not anticipated by Ghias under 35 U.S.C. § 102(b);
- Ground 2: Claims 22, 24–26, and 52 of the ‘998 patent are not obvious over Ghias under 35 U.S.C. § 103(a);

33, 51, and 52 of the ‘998 patent; claims 1–3, 6, 8–14, 18, 19, 21–27, 29–31, and 34–37 of the ‘179 patent; and claims 1–3, 6, 8–14, 18, 19, 21–27, 29, and 30 of the ‘441 patent.

- Ground 3: Claims 15–17, 21, 23, 27, 28, 31–33, 38, and 51 of the ‘998 patent are not anticipated by Iwamura under 35 U.S.C. § 102(e);

‘179 patent:

- Ground 1: Claims 1–3, 6, 8–14, 19, 21–26, 30, 31, and 34–37 of the ‘179 patent are not anticipated by Conwell under 35 U.S.C. § 102(e);
- Ground 2: Claims 1–3, 8, 10–14, 18, 19, 21–27, 29, 31, and 34–37 of the ‘179 patent are not obvious over Ghias and Philyaw under 35 U.S.C. § 103;

‘441 patent:

- Ground 1: Claims 1–3, 6, 8–14, 19, 21–26, and 30 of the ‘441 patent are not anticipated by Conwell under 35 U.S.C. § 102(e); and
- Ground 2: Claims 1–3, 8, 10–14, 18, 19, 21–27, 29, and 30 of the ‘441 patent are not obvious over Ghias and Philyaw under 35 U.S.C. § 103(a).

49. The information below presents the basis for my opinions that the challenged claims of the IPR Patents are not unpatentable based on the grounds at issue in the IPRs.

IV. General concerns with the IPR Petitions and Dr. Moulin’s Declarations (Exs. 1004 in each IPR).

50. It is my understanding that a Petition must set forth how a challenged claim is to be construed. In addition, it is my understanding that a Petition and corresponding declaration must then map the properly construed claim language to the teachings of the asserted art. Based on my review of the Petitions and corresponding Declarations, one skilled in the art would understand that the Petitions and corresponding Declarations fail to comply with these requirements.

51. First, the Petitions (and in particular, the ‘237 Petition) and corresponding Declarations fail to identify any construction of the phrase “approximate nearest neighbor search.” *See* Pet. (‘237) at 1-53. Claims 9 and 13 of the ‘237 patent include an “approximate nearest neighbor search.” *See* ‘237 patent claims 9 and 13. Because the Petition and Declarant do not identify any construction of “approximate nearest neighbor search,” they fail to map the properly construed language to the teaching of the prior art (*i.e.*, Iwamura, Ghias, and Chen).

52. Second, although Petitioner and Dr. Moulin identified constructions for certain terms,⁷ neither the Petition nor the corresponding Declaration maps the construed claim language to the teachings of the asserted art.

53. Moreover, one skilled in the art would understand that Petitioner's failure to map the properly construed claim language to the teachings of the prior art results in critical mistakes in the IPR Petitions and Declarations, as I illustrate using the following two examples:

54. Example 1: In his Declaration with respect to the '237 patent, Petitioner's Declarant, Dr. Moulin, confirmed (consistent with my understanding and the understanding of one of ordinary skill in the art) that a "sublinear search" is a search that has a sublinear relationship to the database size:

53. I understand and agree with Petitioner's position that the term "sublinear search" means "a search whose execution time **has a sublinear relationship to database size.**" For instance, a linear search of a 200-item database would take twice as long as a linear search of a 100-item database. By contrast, a sublinear search of a 200-item database would take less than twice as long as a sublinear search of a 100-item database, perhaps, for instance, 1.5 times as long.

⁷ See, e.g., Moulin Decl. ('237) ¶43 (addressing non-exhaustive search); Moulin Decl. ('237) ¶48 (addressing identify a neighbor / identify a near neighbor / nearest neighbor search); Moulin Decl. ('237) ¶53 (addressing sub-linear).

Moulin Decl. ('237) ¶53. In that same Declaration, Dr. Moulin also asserted that Ghias discloses a search that is sublinear based on the disclosure in Ghias of algorithms that have running times of “ $O(kn)$ or $O(n\log(m))$ ”:

faster than "brute force" searches. Ex. 1010 at 6:23-35. In particular, Ghias discloses searches whose execution times are proportional to the logarithm of the size of the data set (*id.* at 6:24-28 (" $O(kn)$ or $O(n\log(m))$ ")), which, as explained above in Section V(D), are sublinear (Ex. 1001 at 8:54-63).

Moulin Decl. ('237) ¶123.

55. In asserting that Ghias discloses a search that is sublinear, Dr. Moulin did not apply the construction of a sublinear search which requires that the search be sublinear with respect to the “size of the database” and not the number of pitch differences in the query which is the length of the query. Moulin Decl. ('237) ¶53. However, as stated in Ghias (and confirmed by Dr. Moulin at his deposition, *see* the deposition citations that I reference below), (1) “ m ” refers to the number of pitch differences in the query while “ n ” refers to the size of the string (song), and (2) each search algorithm identified in Ghias is linear with respect to the size of the data set. Accordingly, one skilled in the art would understand that Ghias does not disclose a sublinear search under Petitioner’s (and Dr. Moulin’s) own construction of sublinear.

56. Dr. Moulin's deposition transcript addressing the paragraph from his Declaration that I presented above (Moulin Decl. ('237) ¶123) demonstrates to one skilled in the art the problems that result from the Petitioner failing to apply Dr. Moulin's own construction to the art:

14	Q	And by "logarithm of the dataset," you
15	mean	logarithm of the size of the query, not the
16		dataset to be searched; is that true?
17	A	I believe we discussed that earlier.
18	Okay.	Should have -- here it is the log, indeed, of
19		the query dataset. So log(m).
20	Q	So it should -- this should be
21		interpreted -- when you write "dataset," you're --
22		what you mean is the query dataset, not the dataset
23		to be searched?
24	A	That's correct. So the logarithm of the
25		dataset is log(m), which is written below. Yes.
1	Q	Which is the query dataset?
2	A	That's right.

Moulin Depo. 154:14-155:2

IPR2015-00343, IPR2015-00345, IPR2015-00347, and IPR2015-00348
Declaration of George Karypis

12 Q Why is it -- why did you think it was
13 relevant to tell the Board that information? What
14 would they do with it?
15 MR. ELACQUA: Calls for speculation.
16 Objection.
17 THE WITNESS: I don't remember why I wrote it
18 that way. Again, I agree it could be written in a
19 more precise way. So it should have been the log of
20 the query dataset. The word "query" should have
21 appeared there.
22 BY MR. DOVEL:
23 Q Why don't you write that in so we don't
24 have that confusion.
25 MR. ELACQUA: Again, I'm going to object.
1 THE WITNESS: I've said it. So the word
2 "query" should be included before "dataset" in the
3 box on line 4.
4 BY MR. DOVEL:
5 Q Why do you object to writing it in there?
6 A Why should I write it? I just said it.

Moulin Depo. 155:12-156:6.

1 A Because it's very clear. I'm saying the
2 word "query" should have been incorporated before
3 "dataset."

Moulin Depo. 156:22-157:3.

6 Well, one possible reason you wanted to
7 give the Board this information is so that they
8 would misread it and be misled.
9 That's a possibility; right?
10 A It's not at all the -- the reason. So
11 it alone -- there are four documents. Some, they
12 are, like, 90 pages each. Some of the words could
13 have been better chosen. In particular, the word
14 "query" should have been there. I have acknowledged
15 that.
16 Q But assume --
17 A Again, I have acknowledged that this was
18 not written the best way.

Moulin Depo. 157:6-18.

57. Example 2: As I noted above, in his Declaration with respect to the '237 Patent, Dr. Moulin confirmed that a sublinear search is a search that has a sublinear relationship to the database size:

53. I understand and agree with Petitioner's position that the term "sublinear search" means "a search whose execution time has a sublinear relationship to database size." For instance, a linear search of a 200-item database

Moulin Decl. ('237) ¶53. Just a few pages later in that same Declaration, Dr. Moulin asserted:

72. It is my opinion that Iwamura further teaches how this search can be sublinear. For example, Iwamura discloses that different "search algorithms may be applied to perform melody searches," (*id.* at 10:2-3), such as the "Boyer-Moore algorithm." (*id.* at 9:63). "On the average the [Boyer-Moore] algorithm has a sublinear behaviour." Ex. 1017 at 1.

Moulin Decl. ('237) ¶72. In asserting that Iwamura discloses a search that is sublinear, Dr. Moulin did not apply his construction of a sublinear to the referenced Boyer-Moore algorithm.

58. As I explain in detail below and confirmed by Dr. Moulin, one skilled in the art would understand that the Boyer-Moore algorithm disclosed in Ghias does not disclose sublinear behaviors with respect to the size of the data set but only with respect to the query (or pattern)⁸ to be matched. Again, Dr. Moulin's deposition transcript addressing the paragraph from his Declaration that I presented above (Moulin Decl. ('237) ¶72) demonstrates to one skilled in the art the problems that result from Petitioner failing to apply Dr. Moulin's own construction to the art:

⁸ Dr. Moulin testified that "query," "pattern," and "probe" are "all synonymous in this context." Moulin Depo. 21:24-22:1. I agree with Dr. Moulin's testimony.

IPR2015-00343, IPR2015-00345, IPR2015-00347, and IPR2015-00348
Declaration of George Karypis

20 Q When you wrote this, were you trying to
21 convey that searching using the Boyer-Moore
22 algorithm would be sublinear with respect to the
23 size of the dataset being searched?
24 A No. I did not -- no.

Moulin Depo. 74:20-24.

9 Q When we read paragraph 72, are you
10 conveying to the reader that the Boyer-Moore
11 algorithm is sublinear with respect to the size of
12 the dataset being searched?
13 A No. All I do is quote a part of a paper
14 that shows why that algorithm is much faster than
15 brute force. That's all I'm doing. You're
16 inferring things I'm not saying or writing.

Moulin Depo. 69:9-16.

9 Q When we read paragraph 72, are you
10 conveying to the reader that the Boyer-Moore
11 algorithm is sublinear with respect to the size of
12 the dataset being searched?
13 A No. All I do is quote a part of a paper
14 that shows why that algorithm is much faster than
15 brute force. That's all I'm doing. You're
16 inferring things I'm not saying or writing.

Moulin Depo. 66:9-18.

23		When you wrote this sentence here on
24		page 29, did you think that someone at the Board
25		looking at this might think that you meant that a --
1		the Boyer-Moore algorithm was sublinear as used in
2		the patent claim?
3	A	I didn't think of it that way, no.

Moulin Depo. 75:23-76:3.

17	Q	And when you wrote this, were you trying
18		to convey to the Board that the Boyer-Moore
19		algorithm is one that, if you use it, it's sublinear
20		with respect to the size of the database?
21	A	No. No.

Moulin Depo. 67:17-21.

V. Claim Constructions.

59. In construing the claims of the IPR Patents, I understand that in this proceeding, the claims are interpreted using the broadest reasonable construction in light of the patent in which they appear. I also understand that there is a presumption that a claim term carries its ordinary and customary meaning to one of ordinary skill in the art at the time of the invention. In conducting my analysis of the claim elements below, I apply this understanding.

60. It is also my understanding that, for purposes of evaluating whether the IPR Petitions satisfied an initial threshold, the Board identified certain claim

constructions in its Decisions. It is my understanding that the preliminary constructions identified by the Board are:

“sub-linear”	“a search whose execution time scales with a less than linear relationship to the size of the data set to be searched” Decision (‘237) at 7.
“non-exhaustive search” / “non-exhaustive...search”	“a search that locates a match without a comparison of all possible matches” Decision (‘237) at 7; Decision (‘988) at 7; Decision (‘179) at 7; and Decision (‘441) at 7.
“neighbor search” / “identifying a neighbor”	“identifying a close, but not necessarily exact or closest, match” Decision (‘988) at 7; Decision (‘179) at 8; and Decision (‘441) at 7.
“neighbor” / “near neighbor”	“a close, but not necessarily exact or closest, match” Decision (‘237) at 8.
“approximate nearest neighbor search”	“identifying a close match that is not necessarily the closest match” Decision (‘237) at 9.

I address each construction in turn.

A. sub-linear (‘237 patent).

61. The Board’s preliminary construction of “sub-linear time search” is “a search whose execution time scales with a less than linear relationship to the size of the data set to be searched.” Decision (‘237) at 7. While the Board’s preliminary construction is a correct construction of “sub-linear time search,” there are apparently two possible interpretations of the Board’s construction:

Interpretation 1: consistent with the meaning of the phrase “size of the data set,” the “size of the data set” refers to the number of records in the data set being searched (such that the relevant linear relationship is with respect to the size of the data set, *i.e.*, the number of records in the data set), or

Interpretation 2: the “size of the data set” refers to the length of an individual record in the database (such that the linear relationship is with respect to the length of an individual record to be searched rather than the size of the data set).

62. As I explain below, the first interpretation is correct. Also, as I demonstrate below, the asserted art for this element (Iwamura and Ghias) does not disclose a sub-linear search under either interpretation. The (1) words actually used in the Board’s construction, the (2) specification of the ‘237 patent, and (3) the Petitioner’s Declarant, confirm that the proper interpretation of a sub-linear search, in the context of the ‘237 patent, is a search that is sublinear with the size

of the data set where the data set is the number of records in the database, not the length of an individual record in the data set.

1. the words used in the construction: “size of the dataset”

63. A data set, in the context of the ‘237, is a database (*i.e.*, set of records), not an individual record in a dataset or database. The Board incorporated the phrase “size of the dataset” in its preliminary construction based on the fact that “database” and “dataset” are “largely consistent” such that “database size” and “size of the data set” are “largely consistent.” Decision (‘237) at 7. As Petitioner’s Declarant confirmed, “dataset” and “database” are the same in the context of the IPR Patents:

14	Q	When you say "database," is that the same
15		thing as "dataset" in this context?
16	A	Yes.

Moulin Depo. 22:14-16.

11	Q	When you wrote "sublinear" here, you're
12		referring to sublinear with respect to the size of
13		the dataset being searched?
14	A	Yes. Okay. Just to be clear, the dataset
15		is the whole database, okay, the database of songs.

Moulin Depo. 110:11-15. Consistent with the understanding of one skilled in the art, I agree with Petitioner’s Declarant—that “database” and “dataset” are the same in the context of the IPR Patents.

64. A database comprises all records in the database; a single record in a database is not a database. My understanding is consistent with the testimony of Petitioner's Declarant:

11	So just to make it clear, N is not the
12	size of the database; N is the size of a single
13	song.

Moulin Depo. 89:4-13.

65. Moreover, I observed that the origin of the phrase “data set” in the Board’s construction of sub-liner is the Patent Owner’s Preliminary Response for the ‘237 patent in which Patent Owner specifically clarified that the “‘size of the data set’ is the number of potential matches in the data set (*i.e.*, the ‘number of entries in the search database.’” Preliminary Response (‘237) at 9-10 (*quoting* Moulin Decl. (‘237) ¶54). I note that neither the Patent Owner (in its Preliminary Response) nor the Petitioner (in its Petition or Declaration) stated or suggested that sublinear should be based on the length of an individual query in the dataset.⁹

⁹ This is the natural growth mechanism for such a problem. For example, if there are X songs in a database and X additional songs are added, in one dimension the database has doubled (2X). In general, there is no reason to think that the distribution of record sizes changes as the size of the database (number of record in the database) increases. If each record is a fixed length vector, the database size

2. ‘237 specification.

66. As I noted above, it is my understanding that a claim in an unexpired patent is given its broadest reasonable construction in light of the specification of the patent in which it appears. It is also my understanding that the best source for discerning the proper context of claim terms is the patent specification.

67. “Sub-linear” indicates a relationship between two quantities which is less than linear. Linearity describes “[t]he relationship existing between two quantities when a change in a second quantity is directly proportionate to a change in the first quantity.” Ex. 2007 (Modern Dictionary of Electronics) at 425 (1999). “Sub-” is a prefix indicating “under” or “below.”

68. The claim language identifies “time” as one of the quantities being related. In the expression “a sub-linear time search,” “sub-linear time” is an adjective phrase modifying “search.” The ‘237 specification identifies the number of records in the data set (“N”) as the variable that is sub-linear with respect to time.

will double when the number of records doubles. Even if each record is a digital representation of an entire song, there is no reason to think that the new songs have a time / length statistical distribution that is significantly different from the songs that already exist in the database.

69. First, the specification identifies a problem with prior art searches—that the searches are “linear” with respect to the number of records in the data set (“N”)¹⁰—not with respect to the length of an individual record in the database being searched:

“If binary search was possible, then a database containing N vectors would require at most $\log(N)$ comparisons. In previous work, it was not uncommon to perform a linear search of all N entries, perhaps halting the search when the first match is found. On average, this will require N/2 comparisons.¹¹ If N is large, this search can be computationally very expensive.”

‘237, 8:54–63.¹²

¹⁰ In the ‘237 specification, the variable “N” indicates the number of entries in the database being searched—“a database containing N vectors.” ‘237, 8:54–55.

¹¹ N/2 is the expected average result for an exhaustive or linear search of a dataset with one match.

¹² A search algorithm that requires N/2 comparisons has a running time that scales linearly with respect to N. As N increases by one, the search’s running time increases by, on average, one half of the running time of a single comparison. Thus, as N increases, the running time of a search requiring N/2 comparisons increases by a directly proportionate amount, *i.e.*, linearly.

“Consider a situation in which one out of 100,000 possible commercials is to be identified. Each 30-second commercial consists of 900 video frames. If all 900 frames are stored in the database, then $N=90,000,000$. Even if only every 10th video frame is stored in the database, its size is still nine million. While databases of this size are now common, they rely of [sic] efficient search to access entries, i.e., they do not perform a linear search. A binary search of a 90,000,000-item database requires less than 20 comparisons. In contrast, a linear search will require an average of 45,000,000!”

‘237, 21:14–23.

70. In both of the instances from the ‘237 specification that I presented above, the ‘237 specification describes prior art search techniques as “linear” with respect to “N”—the number of records in the database being searched—not with respect to the length of an individual record in the database.

71. Second, the ‘237 specification identifies search techniques that achieve a sub-linear search time with respect to the number of records in the database being searched (not with respect to the length of an individual record being searched):

“Other forms of matching include those based on clustering, kd-trees, vantage point trees and excluded middle vantage point forests are possible and will be discussed in more detail later. . . . Thus, for example, a sub-linear search time can be achieved.”

‘237, 8:64–9:7.

“A number of possible data structures are applicable including kd-trees and vantage point trees. These data structures and associated search algorithms organize a N-point dataset (N=90,000,000 in out previous example) so that sub-linear time searches can be performed on average.”

‘237, 21:56–60. Clustering, kd-trees, vantage point trees and excluded middle vantage point all achieve sub-linear behavior by reducing the number of records being search, *e.g.*, by discarding clusters (buckets) of potential matches, not by reducing the length of an individual record being searched. These methods prune parts of the search space (*i.e.*, data records to be searched) and this is why they are efficient.¹³

¹³ The Yianilos paper incorporated by reference into the ‘237 patent (‘237, 8:65-9:6) explains: “We introduce the idea of aggressive pruning and give a family of practical algorithms, an idealized analysis, and describe experiments. Our main result is that search complexity measured in terms of *d*-dimensional inner product operations, is i) strongly sublinear with respect to the data set size *n* for moderate *R*, ii) asymptotically, and as a practical matter, independent of dimension.” *See* Ex. 2010 (P. N. Yianilos) at 1.

72. Again, in all instances, the ‘237 specification describes techniques as “linear” with respect to “N”—the number of records in the database being searched—not with respect to the length of an individual record in the database.

3. Petitioner’s Declarant.

73. According to Petitioner (consistent with the understanding of one skilled in the art), a “sublinear” search is “a search whose execution time has a sublinear relationship to database size”—where the database size is the number of records in the database, not the length of an individual record in the data set.

74. A “database” consists of all records in the data set; a “database” is not one individual record in the database—an individual record is not a “database.” Dr. Moulin confirmed that sublinear in the context of the ‘237 patent is based on the size of database (a “concept that’s common in [his] field” Moulin 8:10-14), not the length of an individual record in the database:

53. I understand and agree with Petitioner's position that the term "sublinear search" means "a search whose execution time has a sublinear relationship to database size." For instance, a linear search of a 200-item database would take twice as long as a linear search of a 100-item database. By contrast, a sublinear search of a 200-item database would take less than twice as long as a sublinear search of a 100-item database, perhaps, for instance, 1.5 times as long.

Moulin Decl. (‘237) ¶53. As Petitioner’s Declarant—Dr. Moulin—confirmed:

24 Q And what it refers to, to be more precise,
25 is that if we have a sublinear search, that means
1 that the search time is going to grow at less than a
2 linear relationship compared to the size of the
3 database as we increase the size of the database?
4 A That is correct.

Moulin Depo. 13:24-14:4.

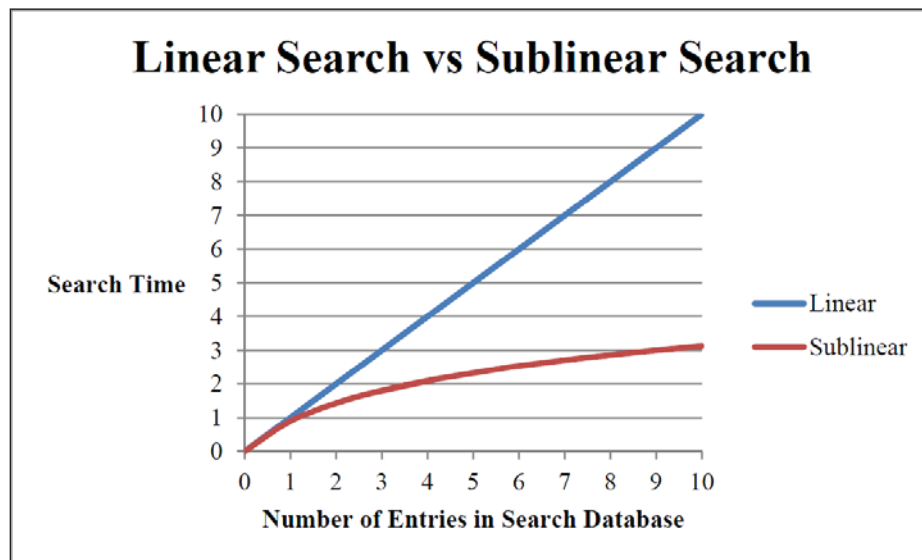
16 Q Well, as you understood the term
17 "sublinear" as it's used in the patent, it's about
18 the execution time as we increase the size of the
19 database; is that right?
20 A In the patent, yes.
21 Q Yes.
22 A Yes.

Moulin Depo. 103:16-22; 16:4-12; 24:1-12.

75. The Petition and corresponding Declaration interpreted “database size” as the number of records in the database (not the length of an individual record to be searched):

“For instance, a linear search of a 200-item database would take twice as long as a linear search of a 100-item database. By contrast, a sublinear search of a 200-item database would take less than twice as long as a sublinear search of a 100-item database, perhaps, for instance, 1.5 times as long.”

Pet. (‘237) at 6; Moulin Decl. (‘237) ¶ 53.



Moulin Decl. ('237) ¶54 (showing that the execution time of a sub-linear time search increases with a less than linear relationship to the “number of entries in the search database”).

“[I]t is my opinion that a search whose execution time is proportional to the logarithm of the size of the data set (e.g., a search with execution time proportional to $A \log(BN)$, where A and B are constants, and N is the number of entries in the database) is sublinear.”

Moulin Decl. ('237) ¶56.

76. One skilled in the art would understand that each explanation of “sub-linear” in the Petition and Declaration demonstrates that the sub-linearity of a search depends on the number of entries in a database, not on other factors, such as the length of an individual record in the database. Nowhere does Petitioner or the

Declarant suggest that the relevant sublinear search is with respect to the length of an individual record to be searched.¹⁴

77. I note that in its Decision, the Board did not present any analysis or reasoning for interpreting sub-linear relative to the length of an individual record being searched rather than the number of records in the dataset. *See* Pet. ('237) at 7. And the example presented by the Board in its Decision equates “data set” with the number of records in the database (“N”), not the length of an individual record:

linear relationship to the size of the data set to be searched.” One example of such a sub-linear search would be a search with an execution time proportional to the logarithm of the size of the data set (“N”), where a doubling of N would lead to an execution time proportional to $\log(2N)$.

¹⁴ In Petitioner’s analysis, the length of individual records is not a factor in evaluating whether an algorithm is a sub-linear time search or a linear time search. If the length of individual records were a factor in evaluating the sub-linearity of a search, then such lengths would have to be addressed in determining whether a given algorithm is sub-linear and Petitioner’s examples would have to account for those lengths. The Petition does not rely upon any variable other than the number of items in the database to be searched to determine whether a search is sub-linear. The Petition determines whether a search has a sub-linear running time exclusively with respect to the number of entries in the database to be searched.

Decision ('237) at 7.

78. Importantly, one skilled in the art would understand that under any possible interpretation of “sublinear” in the context of the ‘237 patent (as well as the general context of search algorithms), a search algorithm is sublinear only with respect to the size of the dataset (the size of the reference database), not the size of the query (pattern) of the work to be identified using the search. As Petitioner’s expert confirmed, consistent with my understanding, whether a prior art search “scales based upon the size of the query or pattern” would not “be accurately assessing the ‘237 patent claims.” Moulin Depo. 25:22-26:8.

11	Is it the case that for a search to be
12	sublinear as it's used in the '237 patent, it's not
13	enough for it to have execution time that is
14	sublinear in relationship to the size of the
15	pattern; it must also be sublinear in relationship
16	to the size of the database?
17	A When I read "sublinear" in, say, Claim 5
18	of the patent, as we just did, I understand
19	sublinear to mean in relation with the size of the
20	database. It does not say anything about in
21	relation with the size of the query.

Moulin Depo. 26:11-21.

IPR2015-00343, IPR2015-00345, IPR2015-00347, and IPR2015-00348
Declaration of George Karypis

4 Q Would it be wrong, then, to assess whether
5 a search is sublinear by applying a definition that
6 said, "It's sublinear if this execution time has a
7 sublinear relationship to the size of the query or
8 the pattern"?

9 A It would not be very relevant. Again,
10 mathematically, it can be done. Everything can be
11 done. But it would not be relevant, from an
12 engineering viewpoint, for an application like this.

Moulin Depo. 25:4-12.

25 Let's assume that the Patent Trial and
1 Appeal Board was presented with prior art that
2 presented a search that was linear with respect to
3 the size of the database, but it was sublinear with
4 respect to the size of the pattern.

5 Would that prior art demonstrate a
6 sublinear search as it's used in Claim 25?

7 A People say it's a linear search, again,
8 because it's in relation with the size of the
9 database. And as you just said, that complexity is
10 still linear; so people would say it is a linear
11 search.

12 Q It is not a sublinear search?

13 A It's only sublinear in terms of the size
14 of the query, which is generally not the
15 parameter of -- the relevant parameter.

Moulin Depo. 26:25-27:15.

IPR2015-00343, IPR2015-00345, IPR2015-00347, and IPR2015-00348
Declaration of George Karypis

16 Q Is it the case that if we had a piece of
17 prior art that was linear with respect to the size
18 of the database but sublinear with respect to the
19 size of the query or the pattern, that that prior
20 art would not teach a sublinear search as it's used
21 in Claim 25?
22 A Again, if one understands sublinear to be
23 in terms in relation to the size of the database,
24 that would be a -- a linear search.

Moulin Depo. 27:16-24.

4 Let's assume we've got a piece of prior
5 art that scales at a sublinear relationship with the
6 size of the pattern or query but it scales at a
7 linear relationship with the size of the database
8 that's being searched.
9 Would that prior art demonstrate or
10 disclose a sublinear search as it's used in the
11 claims of the '237 patent?
12 A No. Again, because my understanding is
13 the claims of the '237 patent, whenever there's
14 mention of "sublinear," it means in terms of the
15 database size. It does not say it explicitly; it's
16 my inference based on my knowledge and my expertise.

Moulin Depo. 28:4-16. I agree with the Petitioner's Declarant—that “sublinear” in the context of the IPR Patents is with respect to the size of the database, not the size of the query or pattern.

B. non-exhaustive search (‘237, ‘988, ‘179, and ‘441 patents).

- 1. The Board’s preliminary construction of “non-exhaustive search” is consistent with the understanding of one of ordinary skill in the art: “a search that locates a match without a comparison of all possible matches.”**

79. A “non-exhaustive” search is a search that uses an algorithm designed to locate a match without comparing the work to all records in the database. A “non-exhaustive” search uses an intelligent algorithm to narrow the database to only a subset of potential matches. *See e.g.*, Moulin Decl. (‘988) ¶12 (“algorithms that increased efficiency by intelligently searching only a subset of potential matches (*i.e.*, ‘non-exhaustive’ algorithms)”); Pet. (‘237) at 3 (“search algorithms that increased efficiency by intelligently searching only a subset of potential matches (*i.e.*, ‘non-exhaustive’ algorithms)”).

80. For example, if there are 100 records in a database, a non-exhaustive search could use an intelligent algorithm to exclude 75 records from the search such that only 25 would be searched during the comparison process. As the specifications of the IPR Patents observe, these non-exhaustive “forms of matching include[ing] those based on clustering, kd-trees, vantage point trees and excluded middle vantage point forests” do not systematically compare the work to be identified to each record. *See e.g.*, ‘179, 9:14-17. Each of these examples uses an intelligent algorithm to narrow the database to only a subset of potential matches.

81. A “non-exhaustive” search can be contrasted with an “exhaustive” search. An exhaustive search systematically checks whether each potential match matches the work to be identified until a match is found, “perhaps halting the search when the first match is found.” ‘237, 8:59-61. An exhaustive search is “a very general problem-solving technique that consists of systematically enumerating all possible candidates for the solution and checking whether each candidate satisfies the problem’s statement.” Ex. 2001 (the “solution” here refers to a record and not a section within that record).

82. If there are 100 records in the database, an “exhaustive” search does not narrow the potential matches but instead systematically compares the work with each record to determine a match (if there is one). Systematically comparing the work to be identified with each potential match until a match is identified rather than using intelligence to narrow the search candidates is also referred to as using “brute force.” Moulin Decl. (‘988) ¶44 (“a brute force search conducts a comparison of every item in a search database”); Ex. 2001.