UNITED STATES PATENT AND TRADEMARK OFFICE

_____

BEFORE THE PATENT TRIAL AND APPEAL BOARD

_____

SAMSUNG ELECTRONICS CO., LTD., and
SAMSUNG ELECTRONICS, AMERICA, INC.,
Petitioners,

v.

CERENCE OPERATING COMPANY,
Patent Owner.

_____

Case No. IPR2025-00458
U.S. Patent No. 8,682,671

_____

DECLARATION OF ERIC KOSKINEN IN SUPPORT OF PATENT OWNER'S PRELIMINARY RESPONSE TO THE PETITION

I, Eric Koskinen, hereby declare as follows:

I. Introduction

1. I am over the age of eighteen (18) years and otherwise competent to make this declaration.

2. I have been retained as an expert witness on behalf of Patent Owner for the above-captioned *inter partes* review ("IPR"). I understand that the petition for *inter partes* review involves U.S. Patent No. 8,682,671 ("the '671 Patent").

3. I make this declaration based on my personal knowledge, educational background and training, consideration of the materials I discuss herein, and my expert opinions.

4. My work in this case is being billed at my normal hourly consulting rate of $700, with reimbursement for actual expenses. My compensation is not related to the outcome of any proceeding involving these Asserted Patents. I have no personal interest in the outcome of the case.

5. In preparing this Declaration, I have reviewed and considered the '671 Patent, the Petition, the declaration submitted by Mr. Schmandt in this proceeding, and each document cited in my declaration.

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 2 of 53

II. Background and Qualifications

6. My qualifications for forming the opinions given in this declaration are summarized here and are addressed more fully in my curriculum vitae, which is submitted as Exhibit 2006.

7. I currently serve as a Tenured Associate Professor at Stevens Institute of Technology and, for several years, have served as a consultant, providing software-related services to law firms and corporations. Previously, I was a Lecturer/Researcher at Yale and a Professor at New York University. I completed an undergraduate degree (with Highest Honors) in Computer Science and Physics from The College of William and Mary, received a Master's degree in Computer Science from Brown University, and completed a Ph.D. in Computer Science from the University of Cambridge.

8. I have been working in the fields of computer science and software engineering since 2001, including the years 2005-2013 when I completed the Master's degree and the Doctorate mentioned previously.

9. I have work experience in all phases of commercial software design and operation, including requirements engineering, software architecture, software engineering, software deployment, and software project management. This work has included writing and developing software and browser-based web applications for Amazon.com (IMDb), Microsoft, and IBM. I have developed User Experiences

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 3 of 53

(UXs) and User Interfaces (UIs) for a variety of web applications including public- and internal- facing web applications. My experience also includes the development and deployment of countless multi-tier and distributed interacting systems.

10. I am a computer scientist, with expertise in the areas of programming languages, formal methods, concurrency, and systems. I have authored over 40 top-tier academic publications over the past 16 years and been cited over 2,000 times. I have served in international academic leadership positions including top-tier Steering Committees (e.g., ACM/IEEE Logic in Computer Science) and Program Committees (e.g., ACM Principles of Programming Languages (POPL), ACM/SIGPLAN Programming Language Design & Implementation (PLDI), and the International Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA)).

11. I have been awarded funding in excess of $6M from a variety of federal funding agencies, including three National Science Foundation (NSF) Awards, one DARPA Award, and two Awards from the Office of Naval Research (ONR).

12. I have authored top-tier conference publications in the field of programming languages, compilers, and automatic software verification, including articles in ACM Principles of Programming Languages (e.g., two papers in POPL 2023), ACM/SIGPLAN Programming Language Design & Implementation (PLDI 2017), and recent articles published in Object-Oriented Programming, Systems, Languages,

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 4 of 53

and Applications (OOPSLA 2019, 2020, 2022, and 2024) and Computer Aided Verification (CAV 2021). My publications have also received awards, e.g., a Distinguished Paper award at POPL 2023.

13. I have also authored top-tier conference publications in the field of systems, concurrency, and distributed systems (e.g., EuroSys 2008, PPoPP 2008, SPAA 2008a, SPAA 2008b, PODC 2017, and the Journal of Distributed Computing 2020), including recent works on multicore parallelization (OOPSLA 2022) and distributed ledgers, i.e., blockchains (PODC 2017 and Journal of Distributed Computing 2020). I have also served on program committees in these areas, including the International Conference On Principles Of Distributed Systems (OPODIS 2012), the ACM Symposium on Parallelism in Algorithms and Architectures (SPAA 2012), the ACM Symposium on Principles of Distributed Computing (PODC 2018), Principles and Practice of Parallel Programming (PPoPP 2018), and the International Conference on Networked Systems (NETYS 2023).

14. I have taught computer science courses on Data-Structures, Algorithms, Object-Oriented Programming, Compilers, and Societal Impacts of Information Technology. I have supervised PhD, Master's, and undergraduate students who have gone on to jobs in top companies or academic positions.

15. I also have an extensive background in software systems. I have developed a wide variety of software applications, including desktop applications, web-based

applications, network applications, and mobile applications. I have worked on a wide variety of computer infrastructure systems such as web servers, database servers, transactional systems, network systems, and APIs.

16. From 2003 to 2005, I was a Software Development Engineer at Amazon.com, where I developed industrial software applications including custom web servers, internal and external services, web-based applications, mobile applications and e-commerce applications, interfacing with internal payment systems.

17. As a Software Engineer at Amazon, I worked on the Internet Movie Database (IMDb). At IMDb I was involved in developing applications that constructed and queried high-performance databases, including Berkeley DBs (key/value Hash Tables), CDBs, MySQL database, PostgreSQL databases, high-performance Web Servers, and custom software abstractions for internal representation.

18. I have delivered dozens of invited lectures at universities and institutions around the world, including Oxford University, Massachusetts Institute of Technology, Cornell University, Princeton University, Yale University, Columbia University, Rice University, Tokyo University, Université Cité de Paris, ETH Zurich, Ecole Polytechnique Federale de Lausanne, University of Johannesburg South Africa, IBM, Microsoft, NEC SRI, and Oracle.

III. Legal Principles

19. I am not an attorney and do not offer opinions on the law in this declaration. For the purposes of this declaration, I have been informed about certain aspects of the law that are relevant to forming my opinions. My understanding of the law is as follows.

A. Person of Ordinary Skill

20. I have been informed that a person having ordinary skill in the art (i.e., a POSITA) is a hypothetical person who is presumed to have known the relevant art at the time the invention was made. I understand that the condition "at the time the invention was made" is imposed to rule out impermissible hindsight. I also understand that an expert is to analyze the prior art from the perspective of a person of ordinary skill in the art and not simply to provide his/her own personal conclusions. I further understand that the factors that may be considered in determining the level of ordinary skill include: (1) the problems encountered in the art; (2) the prior art solutions to the problems encountered in the art; (3) the rapidity of innovation; (4) the sophistication of the technology; and (5) the education level of active workers in the field. I understand that these factors need not all be considered for the analysis and that one or more of these factors may control.

21. I understand that Petitioners and their expert have proposed that a POSITA in the context of the '671 patent would have "possessed a Bachelor of Science (or

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 7 of 53

equivalent) in electrical engineering or computer science, and two years of work experience in the field of speech and natural language processing, including text-to-speech systems." Petition at 13-14. I further understand that Petitioners and their expert assert that "A greater amount of professional education could compensate for fewer years of work experience, and vice versa." *Id*. For the purposes of this Declaration I accept and agree with this POSITA definition.

22. As of February 12, 2010, I met, and in a number of aspects exceeded, the qualifications of a POSITA as defined herein. Additionally, all of my opinions in this declaration are from the perspective of a POSITA as defined here during the relevant timeframe.

B. Claim Construction

23. I understand that a claim construction inquiry begins and ends in all cases with the actual words of the claim. Apart from the written description and the prosecution history, the claims themselves provide substantial guidance as to the meaning of particular terms. I further understand that the context in which a term is used in the asserted claim can be highly instructive. The patent specification can also shed light on the meaning of claim terms.

24. I understand that, when conducting a claim construction inquiry, courts are not required to construe every limitation present in a patent's asserted claims. I

further understand that where a term is used in accordance with its plain meaning, the court should not recharacterize it using different language.

25. I understand that there is a "heavy presumption" that claim terms carry their full ordinary and customary meaning unless the accused infringer can show that the patentee expressly relinquished claim scope. The ordinary and customary meaning of a claim term is the meaning that the term would have to a POSITA at the time of the invention.

26. I understand that courts generally do not import limitations into claims from examples or embodiments appearing only in a patent's written description, even when a specification describes very specific embodiments of the invention or even only a single embodiment. Similarly, I understand that statements during patent prosecution do not limit the claims unless the statement is a clear and unambiguous disavowal of claim scope.

27. I further understand that a preamble of a patent claim is typically not limiting on the scope of the claim if the body of a claim defines a structurally complete invention. I understand that a preamble may be found to be limiting if it supplies the antecedent basis to a claim term in the claim's body, if it recites essential structure or steps, or is necessary to give life, meaning, or vitality to the claims. I understand that language in a claim preamble which merely states a purpose or intended use of the invention is not limiting.

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 9 of 53

C. Burden of Proof

28. I understand that Petitioners bear the burden of establishing that the challenged claims are invalid by clear and convincing evidence.

D. Anticipation

29. I am informed that for a claim to be anticipated, every limitation of the claimed invention must be found in a single prior art reference, either expressly or inherently, arranged as in the claim. I understand that the words in a reference are to be construed in light of the relevant surrounding circumstances in each case, and a reference in any event is good only for that which it discloses.

30. I further understand that anticipation cannot be established by multiple, distinct teachings that the artisan might combine to achieve the claimed invention. Additionally, I understand that when a reference does not disclose that the drawings are to scale and is silent as to dimensions, an accused infringer cannot infer dimensions to demonstrate invalidity. Finally, I understand that an element may be inherently disclosed only if it is necessarily present, not merely probably or possibly present, in the prior art.

E. Obviousness

31. I understand that the grounds for institution in this case only include allegations that challenged Claims 1-3, 7-9 and 12-13 are obvious. I also understand

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 10 of 53

that the scope of issues to be considered in an IPR are limited to those grounds set forth in the IPR Petition.

32. I have been informed that a patent claim is invalid as "obvious" under 35 U.S.C. § 103 in light of one or more prior art references if it would have been obvious to a person of ordinary skill in the art at the time of the invention, taking into account (1) the scope and content of the prior art, (2) the differences between the prior art and the claims, (3) the level of ordinary skill in the art, and (4) any so called "secondary considerations" of non-obviousness, which include "long felt need" for the claimed invention, commercial success attributable to the claimed invention, unexpected results of the claimed invention, skepticism of others, failure by others to achieve the claimed invention, and "copying" of the claimed invention by others.

33. I have been informed that a claim can be obvious in light of a single prior art reference or multiple prior art references. To be obvious in light of a single prior art reference or multiple prior art references, there must be some motivation for a POSITA to modify the single prior art reference, or combine two or more references, in a manner that provides the elements of the claimed invention. This reason may come from a teaching, suggestion, or motivation to combine, or may come from the reference(s) themselves, the knowledge or "common sense" of a POSITA, or from the nature of the problem to be solved, and this reason may be explicit or implicit from the prior art as a whole. I also understand that where there is a reason to modify

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 11 of 53

or combine the prior art to achieve the claimed invention, that it is also important that a POSITA would have a reasonable expectation of success. I have been informed that, under the law, the combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results.

34. It is my understanding that obviousness must be determined as of the date of the invention (*i.e.*, the effective filing date or priority date of the patent). Thus, in considering the reason or motivation to combine references, it is essential to avoid using hindsight. I also understand that the problem examined when considering obviousness may be a problem that would have motivated a POSITA at the time that the invention was made, not the specific problem solved by the invention. I understand that defining the problem in terms of its solution reveals improper hindsight in the selection of the prior art relevant to obviousness. Further, I understand that an overly narrow statement of the problem can represent a form of hindsight, because an inventive contribution may lie in defining the problem in a new, revelatory way. Similarly, I understand that an assertion that a person of ordinary skill could combine the references, rather than that they would have been motivated to do so, is an impermissible form of hindsight. Moreover, I understand that knowledge of a problem and motivation to solve it are entirely different from reason to combine particular references to reach the particular claimed method.

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 12 of 53

35. I am informed that the analysis of the reason to combine prior art should be explicit. Additionally, I understand that an obviousness analysis must include articulated reasoning with rational underpinnings to support the conclusion of obviousness. I understand that providing a full explanation of the reason to combine the references and the reasonable expectation of success is a necessary component of the obviousness inquiry, as inventions in most, if not all, instances rely upon building blocks long since uncovered, and claimed discoveries almost of necessity will be combinations of what, in some sense, is already known. However, I understand that conclusory statements fail to adequately explain why a person of ordinary skill would have a motivation to combine the potential prior art references.

36. I am informed that common sense, common wisdom, and common knowledge may be used to support a reason to combine, so long as the use of common sense in the analysis is explained with sufficient reasoning. But I understand that common sense can only be used to supply a limitation missing from the prior art where the technology is unusually simple and straightforward.

37. I am also informed that evidence suggesting reasons to combine cannot be viewed in a vacuum apart from evidence suggesting reasons not to combine. In analyzing motivation, both advantages and disadvantages must be considered. Teaching away, for example, is a statement in the prior art that either (i) discourages one from following the path of the claimed invention, or (ii) encourages one to

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 13 of 53

follow a path that diverges from the path of the claimed invention. The concept of teaching away bears directly on whether there was a reason to combine prior art. Known disadvantages in old devices which would naturally discourage a search for new inventions may also be taken into account in determining obviousness.

IV. Claim Construction

38. For the purposes of my analysis, I have given all the claim terms their plain and ordinary meaning, as would have been understood by a person having ordinary skill in the art at the time of the alleged invention, which I understand is February 12, 2010, having taken into consideration the language of the claims, the specification, and the prosecution history of record. I reserve the right to respond to any construction questions that may be raised in this matter.

V. Technology Background

A. The ’671 Patent

39. The ’671 patent claims cover methods and systems for use with speech-enabled applications, taking text strings and generating more natural speech during the text-to-speech synthesis process by adding contrastive stress to portions of the text strings to better emulate how a human would vocalize the corresponding phrase. According to the specification, prior art text-to-speech processes “produce[] speech output that is relatively monotone and flat, lacking the naturalness and emotional expressiveness of the naturally produced human speech….” ’671 patent at 8:11-15;

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 14 of 53

see also id., 8:34-36 ("conventional automatic speech synthesis technologies have not taken advantage of contrastive stress as an opportunity to improve intelligibility, naturalness and effectiveness of machine generated speech.").

40. Importantly, the '671 patent allows for identification of which words in a text string may need contrastive stress ***without*** prior tags, identification, or additional context provided by the user or application. *See id*., 9:14-15 ("In some embodiments, the synthesis system may automatically identify which tokens (e.g., words) should contrast with each other without any tags being included in the text input."). As the specification explains, "the synthesis system may further specifically identify which word(s) and/or syllable(s) should carry the contrastive stress." *Id*., 9:16-18. For example, "if a plurality of [words] in a text input contrast with each other, one, some or all of those [words] may be stressed when rendering the speech output." *Id*., 9:20-22. Then, "after identifying which word(s) and/or syllable(s) should carry contrastive stress, the synthesis system may apply the contrastive stress to the identified word(s) and/or syllable(s) through increased pitch, amplitude and/or duration, or in any other suitable manner." *Id*., 9:24-27.

41. The '671 patent's process is also dynamic, in that it can identify portions of a ***single*** sentence which may need contrastive stress (based on a comparison of other portions of that single sentence), or portions of ***multiple*** sentences which may need contrastive stress (based on a comparison of portions of the plurality of sentences).

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 15 of 53

'671 patent at 14:38-41 ("in accordance with some embodiments of the present disclosure, front-end 250 may be programmed to process text input 240 to identify one or more portions of text input 240 that should be rendered with contrastive stress to contrast with one or more other portions of text input 240."). For example, the '671 patent's process may "convert the plurality of text fields to word forms and identify portions that differ between them" and then "assign contrastive stress to be carried by the differing portions." *Id*., 35:62-64.

42. The '671 patent's process may also assign contrastive stress in multiple different ways, i.e., by applying contrastive stress to all differing portions of any two text strings, or by applying contrastive stress to just a subset, or just one of, the differing portions of the text string. '671 patent at 38:23-25 (explaining that "the contrastive stress pattern may involve applying stress to one, some or all of the plurality of text strings…."). For example, as shown in Figure 3A and the corresponding explanation in the specification, the text string "Fight number 1345 was originally scheduled to depart at 10:45 but is now scheduled to depart at 11:45" contains two times (i.e., "10:45" and "11:45") which may be contrasted with each other:

> In this example, "11:45 a.m.", the new time of the flight departure, contrasts with "10:45 a.m.", the original time of the flight departure. Specifically, "eleven" is the part of "11:45 a.m." that differs from and contrasts with the "ten" of "10:45 a.m." (i.e., the "forty_five" and the "a.m." do not differ or contrast). By carrying contrastive stress specifically on the word "eleven" (more particularly, on the syllable "-

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 16 of 53

> lev-", which is the syllable of main lexical stress in the word "eleven"), the resulting synthetic speech output may draw a listener's focus to the contrasting part of the sentence, and cause the listener to pay attention to the important difference between the "ten" of the original time and the "eleven" of the new time.

'671 patent at 22:67-23:12.

43. The '671 patent thus provided a novel way to use speech-enabled applications, resulting in more natural speech output via dynamic application of contrastive stress to select portions of analyzed text strings.

B. Walker[1]

44. Walker is a U.S. Patent Publication which published on December 6, 2001, but which was subsequently abandoned and never issued into a patent. Importantly, Walker is an example of the exact type of prior art system / method that the '671 patent was seeking to improve, i.e., a system / method that provides ***no*** contrastive stress during the text-to-speech process. Instead, Walker is directed to the concept of utilizing contextual analysis to confirm that words in a text string are being pronounced correctly (when said words could be read in multiple different ways depending on the context). *See, e.g.,* Walker, ¶[0002] ("The present invention is generally related to text-to-speech conversion methods and systems and, more particularly, to a text-to-speech method and system which convert written text into audible speech as a function of the context of the text."). For example, Walker

[1] US20010049602 ("Walker") (Ex-1005).

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 17 of 53

explains that the word "wind" could be pronounced phonetically as "wind" (i.e. the "wind is blowing") if the context of the phrase is weather, or alternatively pronounced as "wind" (i.e., "wind the clock") if the context of the phrase is time. *Id*., ¶[0005]. Thus, Walker is directed to an entirely different text-to-speech process than the '671 patent, and further lacks the core advantages the '671 patent provided over the prior art.

C. Matsumoto[2]

45. Matsumoto is a U.S. Patent Publication which later issued into a patent, but which subsequently expired due to the patent owner's failure to pay maintenance fees. Matsumoto describes a speech synthesizer which accepts input sentences, stores the input sentences into a matrix, and then collates new sentences with one or more of the stored sentences to calculate a variation degree between the newly entered sentence and one or more of the previously entered sentences. Matsumoto at Abstract. Matsumoto's speech synthesizer has a calculation unit that "calculates the variation coefficients (the speed coefficient, the pitch coefficient and the volume coefficient) by use of the variation degree obtained from the collation unit 14" and applies these variation degrees to the speed, volume, and pitch of the words in the subsequently analyzed sentence. *See, e.g*., Matsumoto ¶¶[0054]; *see also* ¶¶[0093]-[0103] (walking through operational example).

[2] US20070233492 ("Matsumoto") (Ex-1006).

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 18 of 53

46. While Matsumoto sought to fix issues with prior art speech synthesizers which resulted in purely monotone speech outputs (*see, e.g.* Matsumoto ¶[0005]), Matsumoto's speech synthesizer differs substantially from the '671 patent's process of analyzing portions of text strings and automatically adding contrastive stress as needed. For example, Matsumoto is designed to work only by comparing ***full*** sentences to each other, and only when the sentences have essentially the same structure (i.e., the same number of words and relating to the same general content, such as sentences discussing the weather in different locations). *Id*. This is illustrated below in Figure 8 of Matsumoto and the corresponding disclosure in the specification:

FIG. 8

	b=1	b=2	b=3	b=4	b=5	b=6
t (n=1) =1	KYOU NO	TOHOKU	CHIHOU NO	TENKI WA	HARE	DESU
t (n=2) =2	KYOU NO	KANTO	CHIHOU NO	TENKI WA	HARE	DESU
t (n=3) =3	KYOU NO	TOKAI	CHIHOU NO	TENKI WA	KUMORI TOKIDOKI AME	DESU
t (n=4) =4	ASU NO	KANSAI	CHIHOU NO	TENKI WA	HARE	DESU
t (n=5) =5	ASU NO	KANSAI	CHIHOU NO	SAITEIKION WA	JUDO	DESU
v (n, b)	0.5	1.08	0	1	1.83	0
C1 (n, b)	0.95	1.08	0.85	1.06	1.25	0.85

47. As shown in the above figure, the Matsumoto process involves collating each input sentence in an accumulation matrix. Matsumoto, FIG. 8; *see also id.,* ¶¶[0071]-[0072]. New sentences are then compared to prior sentences via a collation

process, which results in a variation degree "v(n, b)" being calculated for each word in the new sentence based on the degree of variation between that word and the words in the ***same*** numeric location as the prior sentences. *Id.,* ¶¶[0071]-[0082].

48. Additionally, because Matsumoto is designed to only work by comparing newly-input sentences to one or more previously accumulated sentences, Matsumoto's process only applies emphasis / deemphasis to ***subsequentl***y-entered sentences (i.e., emphasis / deemphasis is not applied to characters or words in the first sentence entered, only subsequent sentences). *Id.,* ¶¶[0071]-[0082], [0098]-[0100], [0105].

49. Matsumoto also requires ***significant*** user input and direction as part of the speech synthesis process, for instance, the user must select how many sentences should be analyzed when calculating the degree of variation (i.e., the "collation range", which includes setting the range of sentences, the range of time), setting minimum and maximum "variation coefficients," setting the "collation mode" (i.e., collating all accumulated sentences against the newly input sentence, or the "just-anterior sentence" against the newly-input sentence), and setting the interpolation interval. Matsumoto ¶¶[0056]-[0067].

50. Moreover, while the '671 patent allows for applying contrastive stress to ***portions*** of text strings, while not applying contrast to the remainder of said text strings, Matsumoto discloses examples where ***every*** word in a newly-collated

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 20 of 53

sentence has its speed, volume, and/or pitch adjusted based on the calculated variation coefficient, resulting in emphasis / de-emphasis of the ***entire*** analyzed text string. *Id.,* ¶¶[0087]-[0092] (explaining how the variation coefficient minimum and maximum values are used when determining the speed, pitch, and volume coefficients for each word).

51. In short, unlike the '671 patent, Matsumoto does not disclose any improved interoperability with a speech-enabled application, is limited to applications where a user is seeking to compare sentences with essentially the same sentence structure (i.e., number of words and general content), requires a significant amount of user input during the analysis process, and results in emphasis / de-emphasis of ***both*** contrasting ***and*** similar words (rather than adding contrast to only the words that differ).

D. Bellegarda[3]

52. Bellegarda is U.S. Patent directed to evaluating a newly-entered sentence, comparing it to previously-entered sentences, and assigning a "word prominence" value to words which appear more frequently in prior sentences. *See, e.g.*, Bellegarda at Abstract:

> A method for generating speech that sounds more natural may comprise generating synthesized speech having certain word prominence characteristics and applying a semantically-driven word prominence

[3] US7,313,523 ("Bellegarda") (Ex-1014).

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 21 of 53

> assignment model to specify word prominence consistent with the way humans assign word prominence. A speech representative of a current sentence is generated. The determination is made whether information in the current sentence is new or previously given in accordance with a semantic relationship between the current sentence and a number of preceding sentences. A word prominence is assigned to a word in the current sentence in accordance with the information determination.

53. Like Matsumoto, Bellegarda is about assessing ***subsequently*** entered sentences and assigning variation values to words in subsequent sentences based on the similarity of those words in prior sentences. Indeed, as Bellegarda explains, the focus of its process is to provide a "semantically-driven word prominence assignment model to assign word prominence characteristics consistent with the way humans assign work prominence." *Id*., 4:18-20. In short, Bellegarda is directed to identifying which words should be ***emphasized*** in ***subsequently***-entered sentences, by assessing the context of the previously-entered sentences and identifying which newly-entered words carry new information for the listener. *See id*. at Background.

E. Malsheen[4]

54. Malsheen is a U.S. Patent directed to expanding abbreviations of words. As the specification explains:

> An improved text-to-speech synthesizer that employs a text to speech converter, a text reader control procedure, a classifier procedure, an abbreviation expansion procedure, and an acronym/initialism expanding procedure is herein described. A classifier procedure is used to classify generate classification values for each word in the text message with regard to syntax, punctuation and membership in

[4] US5634084 ("Malsheen") (Ex-1007).

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 22 of 53

> predefined classes of words, the predefined classes of words including number, measurement units, geographic designations, and date/time values. An abbreviation expansion procedure evaluates, based on the classification values for words neighboring the identified words, which, if any, of the potential expansion values is applicable, and substitutes the potential expansion for the identified abbreviation word when evaluation yields a success value. An acronym/initialism expanding procedure identifies words in the text message that are acronyms and initialisms, parses pronounceable syllables within the identified words and generates a substitute string that can consist of any combination of letters, numbers, pronounceable syllables or multiple letter identifiers.

Malsheen at Abstract; *see also id.,* 1:5-10 ("The present invention relates generally to text to speech converters, and particularly to an extension of text to speech technology for improving the handling of text that includes abbreviations, acronyms, initialisms, non-standard compound words, numbers, dates, telephone numbers, fractions, times, and the like.").

F. Marple[5]

55. Marple is a U.S. Patent Publication directed to a speech synthesizer wherein a speech signal can be generated by concatenation from phonemes stored in a phoneme database. As the specification explains:

> Disclosed are novel embodiments of a speech synthesizer and speech synthesis method for generating human-like speech wherein a speech signal can be generated by concatenation from phonemes stored in a phoneme database. Wavelet transforms and interpolation between frames can be employed to effect smooth morphological fusion of adjacent phonemes in the output signal. The phonemes may have one prosody or set of prosody characteristics and one or more alternative prosodies may be created by applying prosody modification parameters

[5] US20080195391 ("Marple") (Ex-1008).

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 23 of 53

> to the phonemes from a differential prosody database. Preferred embodiments can provide fast, resource-efficient speech synthesis with an appealing musical or rhythmic output in a desired prosody style such as reportorial or human interest. The invention includes computer-determining a suitable prosody to apply to a portion of the text by reference to the determined semantic meaning of another portion of the text and applying the determined prosody to the text by modification of the digitized phonemes. In this manner, prosodization can effectively be automated.

Marple at Abstract; *see also id*., [0012] ("There is thus a need for a speech synthesizer and synthesizer method which is resource-efficient and can generate high quality speech from input text. There are further needs for a speech synthesizer and synthesizer method which can provide naturally rhythmic or musical speech and which can readily generate synthetic speech with one or more prosodies.").

VI. Petitioners Grounds for Requesting *Inter Partes* Review are Deficient

A. Ground 1A: Petitioners Fail To Show That Claims 1, 7-8, and 14-15 Would Have Been Obvious Based On Walker in view of Matsumoto

56. In my opinion, because Walker and Matsumoto are directed to entirely different text-to-speech processes than the '671 patent, they fail to disclose or render obvious multiple claim limitations.

1. Independent Claim 1

57. The steps of Claim 1 of the '671 patent are reproduced below:

> 1[pre]: A method for use with a speech-enabled application, the method comprising:
>
> 1[a]: receiving, from the speech-enabled application, input comprising a plurality of text strings;

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 24 of 53

> 1[b]: identifying a first portion of a first text string of the plurality of text strings as differing from a corresponding first portion of a second text string of the plurality of text strings, and a second portion of the first text string as not differing from a corresponding second portion of the second text string;
>
> 1[c]: assigning contrastive stress to the identified first portion of the first text string, but not to the identified second portion of the first text string;
>
> 1[d]: generating, using at least one computer system, speech synthesis output to render the plurality of text strings as speech having the assigned contrastive stress; and
>
> 1[e]: providing the speech synthesis output for the speech-enabled application.

58. As discussed further below, Walker in view of Matsumoto fails to disclose or suggest limitations 1[a], 1[b], 1[c], 1[d], or 1[e].

a. Limitation 1[a]: receiving, from the speech-enabled application, input comprising a plurality of text strings;

59. Walker in view of Matsumoto does not disclose or suggest this limitation.

60. Here, Petitioners argue that Walker discloses receiving "raw text" from "a speech-enabled application," but admit that Walker does ***not*** disclose a "plurality of text strings." Petition at 20-21. Petitioners nevertheless argue that this limitation would be obvious based on Walker in view of Matsumoto. *Id.* Specifically, Petitioners note that Walker discloses that one type of example text processed by its context detector is "a weather report for a particular city." *Id*. (citing Walker ¶[0024]). Petitioners then assert that it "would have been obvious to a POSITA that

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 25 of 53

Walker's weather report could pertain to more than one location/city or day," noting that "Matsumoto discloses a weather forecast to include the weather for different locations and days." *Id*. (citing Matsumoto ¶[0071]).

61. Petitioners also argue that the combination would be obvious to a POSITA because it would have amounted to combining prior art elements (i.e., Walker's context detector receiving a text weather report) according to a known method (Matsumoto teaching multiple weather reports for different regions) to yield predictable results (Walker's context detector receiving a text weather report for more than one location). Petition at 21.

62. As an initial matter, Petitioners support this argument with nothing substantive, other than a conclusory statement from their expert (*see* Petition at 20-21, citing Schmandt Decl., ¶107). Moreover, Petitioners merely state that a POSITA ***could*** modify Walker's contextual analysis engine with Matsumoto's text-to-speech engine, but without explaining ***why*** a POSITA would be motivated to combine Walker's processing of single text strings with Matsumoto's processing of multiple text strings. For example, Walker describes a process where a telephone user requests ***specific*** information from a voice application, and the application returns the ***specific*** information the user requested. As Walker explains, "[f]or example, the telephone user 12 may request a weather report for a ***particular*** city." Walker ¶[0024] (emphasis added). And then, "[i]n response to this request, voice

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 26 of 53

application 16 accesses text data sources 14 to find a text document having the weather report for the ***particular*** city." *Id*. (emphasis added). While Petitioners assert that it would be obvious to combine Walker and Matsumoto to provide weather reports for multiple cities, neither Petitioners, nor their expert, explain ***why*** a POSITA would be motivated to modify Walker to provide weather reports for multiple different cities ***that the user never asked about***.

63. As an illustration of this, using Walker's disclosed system, a user requesting a weather report for Brooklyn may have the below interaction:

- **User:** "What is the weather in Brooklyn today?"
- **Walker's Output:** "The weather in Brooklyn is windy today."

64. Conversely, under Petitioners' proposed combination of Walker and Matsumoto, the interaction would be:

- **User:** "What is the weather in Brooklyn today?"
- **Walker's + Matsumoto's Output:** "The weather in Brooklyn is windy today." "The weather in Tokyo is sunny today." "The weather in Los Angeles is cold today." "The weather in Seoul is cloudy today."

65. In short, while Petitioners argue that Walker's weather report "***could*** pertain to more than one location/city or day" (Petition at 20 (emphasis added)), Petitioners fail to explain ***why*** a POSITA would modify Walker to respond with weather reports for locations the user did not specifically request.

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 27 of 53

66. Moreover, even considering an example where a user requests information on multiple locations at once (an example which Walker does ***not*** disclose or suggest), there is no evidence that Walker's contextual analysis process could function with significantly expanded information requests. Specifically, Walker discloses a system where a user requests ***particular*** information, and the voice application accesses a specific "text data sources" to find a ***particular*** text document having the ***particular*** information the user requested. *See, e.g.,* Walker ¶¶[0015], [0024]. Nowhere does Walker disclose that its voice application could receive and process commands with ***multiple*** levels of instructions or nuances in each command, or that Walker's text data sources are configured to process data requests for multiple different issues simultaneously. Nor do Petitioners even argue that it would have been obvious, let alone possible, for Walker's voice application or text data sources to process multiple disparate commands simultaneously and retrieve data corresponding to each disparate information request. *Id.* Indeed, even today, speech applications have difficulty responding to commands outside of their pre-programmed response database. *See, e.g.,* Ex. 2007 (Adam Rogowski, *Scenario-Based Programming of Voice-Controlled Medical Robotic Systems*, Sensors 2022, 9520 (December 6, 2022)) ("we cannot always assume that the operator utters only the commands contained in the defined sublanguage."). In short,

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 28 of 53

Petitioners' attempt to contrive a hypothetical that meets the claims appears to be tainted with hindsight.

67. Petitioners also provide general arguments as to why a POSITA would be motivated to combine Walker and Matsumoto, none of which have merit. For example, Petitioners assert that "[a] POSITA would have appreciated that Matsumoto's speech synthesizer would have advanced Walker's objective of providing more natural synthesized speech for a weather report and provided additional details regarding how to implement such a TTS engine." Petition at 16 (citing Schmandt Decl., ¶¶88-89). Petitioners further assert that the combination would have been obvious to a POSITA because:

> [I]t would have amounted to using a known technique (Matsumoto's speech synthesizer for text weather reports) in a known device (Walker's text-to-speech system which converts text of weather reports into synthesized speech) ready for improvement (Walker discloses using available TTS engines) to yield predictable results (improving the naturalness and quality of synthesized speech of a weather report).

Petition at 17. But once again, Petitioners support these statements with nothing more than conclusory assertions from their expert. The issue of Matsumoto being directed towards Japanese and specifically, katakana characters, is not even addressed.

68. Additionally, as discussed above, Walker is directed to preventing incorrect pronunciations of words in a ***single*** word/phrase (based on the context of the analyzed sentence), while Matsumoto is directed to applying emphasis and de-

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 29 of 53

emphasis to ***subsequent*** sentences (based on a degree of variation between words or phrases in prior accumulated sentences). Neither Petitioners, nor their expert, explain how these two approaches would be compatible, or how combining these approaches would yield predictable results, especially given that one patent is directed towards English and the other is Japanese.

69. Moreover, neither Petitioners, nor their expert, address how Walker's adjustment to the pronunciation of any text in the analyzed text string may be affected by Matsumoto's ***subsequent*** adjustment of the tone, pitch, and speed of said text string. For example, raising the pitch of the word "really" may indicate surprise or disbelief, while lowering the pitch of the word "really" may indicate sarcasm or skepticism. Similarly, decreasing the speed of the word "really" (i.e., "reeeeally") can signal intrigue or suspicion, while increasing the speed (i.e., "*REALLY?!*") may imply urgency or excitement. Thus, it is possible that Matsumoto's adjustment of the tone, pitch, and speed of the processed text strings may actually impact how the words are perceived and conflict with Walker's core function of ensuring that words are correctly pronounced based on the context of the analyzed text string.

70. As another example, if the Walker process were to be applied to tonal languages, such as the Cantonese language, where adjusting the tone of a word may actually adjust the word's meaning, then combining Matsumoto's adjustment of the

tone, pitch, and speed of words may result in the incorrect pronunciation of the analyzed text string (i.e., the very result Walker is trying to prevent).

71. In short, Petitioners do not contend that either Walker or Matsumoto alone discloses or suggests the '671 patent's limitation of "receiving, from the speech-enabled application, input comprising a plurality of text strings," and Petitioners have not met their burden of establishing that a POSITA would have been motivated to modify or combine Walker and Matsumoto to achieve this limitation, or that a POSITA would have had reasonable success in combining or modifying these references.

b. Limitation 1[b]: identifying a first portion of a first text string of the plurality of text strings as differing from a corresponding first portion of a second text string of the plurality of text strings, and a second portion of the first text string as not differing from a corresponding second portion of the second text string;

72. Walker in view of Matsumoto does not disclose or suggest this limitation.

73. Once again, Petitioners do not assert that either Walker or Matsumoto alone disclose or suggest this limitation, instead asserting that Walker in view of Matsumoto discloses or suggests this limitation. Petition at 22-26. Once again, Petitioners are wrong. As discussed above, a POSITA would not have been motivated to combine or modify Walker and/or Matsumoto, and Petitioners offer no

additional reasoning to justify this combination for this limitation, beyond those already discussed above. *See id*.

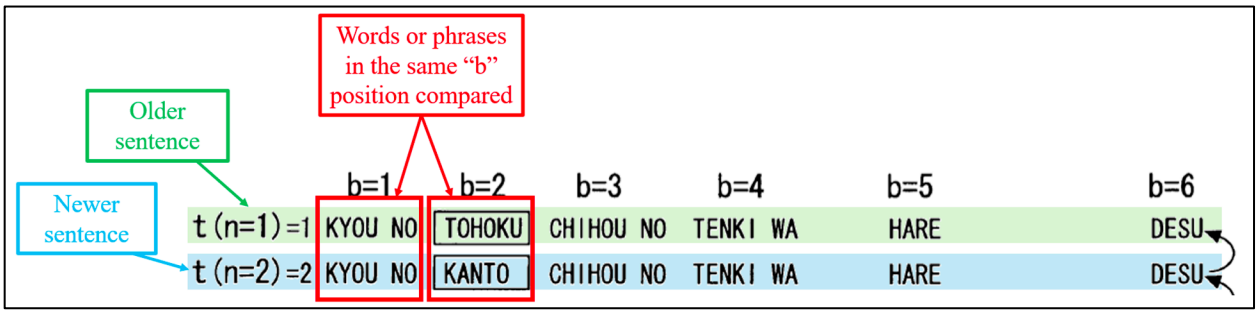
74. Moreover, while Petitioners rely on disclosure from Matsumoto regarding comparing newly-entered sentences with previously-entered sentences to identify which words may be similar or different (*see* Petition at 22-26, citing Matsumoto, ¶¶0058]-[0062], [0074]-[0079]), Petitioners mischaracterize the order of operations in Matsumoto. As discussed above, because Matsumoto is designed to only work by comparing newly-input sentences to one or more previously-accumulated sentences, Matsumoto's process only applies emphasis / deemphasis to ***subsequentl***y-entered sentences (i.e., emphasis / deemphasis is ***not*** applied to characters or words in the first sentence entered, only subsequent sentences). Matsumoto ¶¶[0071]-[0082], [0098]-[0100], [0105].

75. Thus, while independent claim 1 of the '671 patent requires "identifying a first portion of a ***first text string*** of the plurality of text strings as differing from a corresponding first portion of a ***second text string*** of the plurality of text strings, and a second portion of the ***first text string*** as not differing from a corresponding ***second portion*** of the second text string," Matsumoto actually does the opposite – i.e., Matsumoto identifies differing and similar portions of only the ***second***-entered (or subsequently entered) text strings. This is because, as discussed

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 32 of 53

above, Matsumoto's text-to-speech process is fundamentally different than the '671 patent.

76. Indeed, Petitioners appear to recognize this issue, as they inconsistently present the order of Matsumoto's process in their operative example. As shown in the below annotated example Petitioners present, Petitioners have taken Matsumoto's Figure 6 and correctly labeled the t(n=1) sentence as the "older" sentence (i.e., the first text string the engine received), and correctly labeled t(n=2) as the "newer" sentence (i.e., the second text string the engine received):



Petition at 25 (citing Matsumoto, FIG. 6 (annotated)).

77. Petitioners then ***flip*** the order of the entered sentences in an attempt to match the claim language, arguing that:

> A POSITA would have further understood that in the scenario discussed herein, "Kanto" appearing in location a(2, 2) corresponds to "a first portion of a first text string" and "kyou no" (today's) appearing in location a(2, 1) corresponds to "a second portion of the first text string." Schmandt ¶144.

Petition at 26.

78. As shown in the above argument from the Petition, Petitioners attempt to mischaracterize words from the ***second***-entered text string, i.e. "Kanto" appearing at location **a(2, 2)**, as being "***a first portion of a first text string***." *Id*. (emphasis added). But Kanto is ***not*** a portion of the ***first***-entered text string, it is a portion of the ***second***-entered text string, as evidenced by its matrix location **a(2, 2)**. Because, once again, unlike the '671 patent, Matsumoto is focused only on identifying differences in ***subsequently*** entered sentences, based on an analysis of the variation between the newly-entered sentence and subsequently-entered sentences.

79. Thus, *even if* Petitioners had met their burden of showing that a POSITA would be motivated to combine or modify Walker and/or Matsumoto–*and they have not*–Petitioners have still failed to meet their burden of showing that the combination of Walker and Matsumoto discloses or suggests the limitation "identifying a first portion of a first text string of the plurality of text strings as differing from a corresponding first portion of a second text string of the plurality of text strings, and a second portion of the first text string as not differing from a corresponding second portion of the second text string."

c. Limitation 1[c]: assigning contrastive stress to the identified first portion of the first text string, but not to the identified second portion of the first text string;

80. Walker in view of Matsumoto does not disclose or suggest this limitation.

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 34 of 53

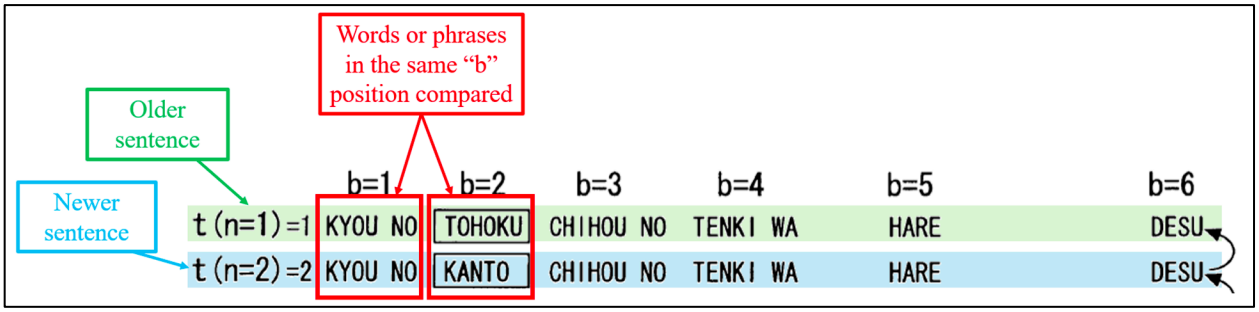
81. Once again, Petitioners do not assert that either Walker or Matsumoto alone disclose or suggest this limitation, instead asserting that Walker in view of Matsumoto discloses or suggests this limitation. Petition at 27-33. Once again, Petitioners are wrong. As discussed above, a POSITA would not have been motivated to combine or modify Walker and/or Matsumoto, and Petitioners offer no additional reasoning to justify this combination for this limitation, beyond those already discussed above. *See id*.

82. Moreover, once again, Petitioners mischaracterize the order of operations in Matsumoto to assert that this limitation is disclosed or suggested. Specifically, while the plain language of the claim requires "assigning contrastive stress to the identified ***first portion of the first text string***…," Petitioners once again identify portions of Matsumoto disclosing applying contrastive stress to a portion of the ***second*** text string entered. Petition at 29-30. Specifically, once again, Petitioners point to Matsumoto's emphasis of the word "Kanto" located in the ***second*** text string analyzed in the example in Figure 6, once again incorrectly labeling this as "the first portion of the first text string:"

> As discussed in §VIII.A.2.c, "Kanto" is identified as different from the corresponding word "Tohoku" and therefore, the calculated variation degree of "Kanto" is 1.0. Thus, a POSITA would have understood that Matsumoto's system calculates variation coefficients (C1, C2, C3) for "Kanto" above 1.0. Schmandt ¶¶166-173 (discussing the calculation of the volume variation coefficient for "Kanto"). Accordingly "Kanto" (i.e., the first portion of the first text string) is emphasized. Schmandt ¶178.

Petition at 29-30.

83. But once again, as shown even in Petitioners annotated example, Kanto is not a portion of the ***first*** entered text string, it is a portion of the ***second*** entered text string:



Petition at 25 (citing Matsumoto, FIG. 6 (annotated)).

84. This distinction is important because, as discussed above, one benefit the '671 patent invention has over Matsumoto is the ability to apply contrastive stress to portions of even a single text string, and even a single syllable, entered, while Matsumoto is limited to applying emphasis / de-emphasis only to ***subsequently*** analyzed sentences.

85. Additionally, Walker in view of Matsumoto does not disclose the second part of this claim limitation, i.e., applying contrastive stress to the first portion of the first text string, "***but not to the identified second portion of the first text string***." As discussed above, and as Petitioners admit, the example process shown in Matsumoto applies emphasis and de-emphasis to ***every*** word in the subsequently modified sentence. Matsumoto ¶¶[0087]-[0092] (explaining how the

variation coefficient minimum and maximum values are used when determining the speed, pitch, and volume coefficients for each word); Petition at 30-31 ("words that are the same are ***deemphasized*** because the variation coefficients for speed/duration (C1), pitch (C2), and volume (C3) are less than 1.0.") (emphasis added); *id.* at 72 ("***Matsumoto teaches deemphasizing corresponding words that are the same between two text strings***") (emphasis added).

86. Apparently recognizing this fatal issue to their invalidity arguments, Petitioners assert that:

> Nevertheless, a POSITA would have understood that deemphasizing words/phrases that are the same (e.g., by decreasing the pitch, amplitude and/or duration), as described in this specific example in Matsumoto, would still constitute "not [assigning contrastive stress] to the identified second portion of the first text string…."

Petition at 30-31.

87. Petitioners then engage in a form of claim construction analysis, essentially arguing that "contrastive stress" should be construed as ***only*** covering emphasizing words, while ***not*** covering de-emphasizing words. *Id*. at 30-32. But Petitioners are wrong. As an initial matter, Petitioners already declared elsewhere in their Petition that the claims should be given their "plain and ordinary meaning" (Petition at 14), and thus Petitioners' attempts to inject a favorable claim construction for this limitation to support their invalidity arguments should be rejected. Here, the "plain and ordinary meaning" of the term "contrastive stress," is

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 37 of 53

simply a long understood linguistic term, pertaining to stress which contrasts with other words in the text string, and neither Petitioners, nor their expert, have proffered any evidence suggesting a skilled artisan would understand the term differently.

88. Nor does Petitioners' selective quotation of the '671 patent specification support their claim construction argument. For example, while Petitioners cite to a portion of the '671 specification which states "synthesis system may apply the contrastive stress to the identified word(s) and/or syllable(s) ***through increased pitch, amplitude and/or duration***…" (Petition at 31 (emphasis in original)), Petitioners ignore the second part of that sentence, which states "***or in any other suitable manner.***" '671 patent at 9:22-27.

89. Indeed, the '671 specification explicitly discusses examples where ***different*** levels of contrastive stress are applied to different portions of modified text. *Id.*, 25:39-42 ("In some embodiments, synthesis system 200 may be programmed to render both fields with contrastive stress, ***but to apply a different level of stress to each field***….") (emphasis added); *see also id*., 25:42-54:

> For example, in some embodiments synthesis system 200 may be programmed to generate the output, "Flight number 1345 was originally scheduled to depart at 10:45 a.m., but is now scheduled to depart at 11:45 a.m. with the "10:45 a.m. rendered with ***anticipatory contrastive stress of the same or a different level*** as the stress applied to "11:45 a.m." In some embodiments, speech-enabled application 210 may be programmed to include one or more annotations in text input 300 to indicate which particular contrastive stress pattern is desired and/or what specific levels of stress are desired in association with individual contrasting fields.

90. The ’671 patent specification also explicitly discusses embodiments applying “lower” contrastive stress to some words compared to other words. *Id*., 28:12-16 (“In some embodiments, the level of stress assigned to a field or token that differs entirely from another field or token with which it contrasts ***may be lower*** than that assigned to fields or tokens for which only one or more portions differ.”) (emphasis added); *see also* MBO Labs., Inc. v. Becton, Dickinson & Co., 474 F.3d 1323, 1333 (Fed. Cir. 2007) (“A claim interpretation that excludes a preferred embodiment from the scope of the claim is rarely, if ever, correct.”) (internal citation omitted).

91. Moreover, while Petitioners attempt to support their claim construction argument by pointing to statements made during prosecution of the ’671 patent’s parent application (Petition at 31-32), Petitioners identify ***no*** portion of the parent application’s prosecution history which affirmatively states that contrastive stress is limited only to ***emphasis*** of words. Instead, the parent application’s prosecution history merely confirms that contrastive stress is a particular type of stress “applied in speech to words or syllables that are meant to contrast with each other” (Petition at 31) (citing ’610 FH, 147; ’671 patent at 8:17-20). And while Petitioners further argue that the ’610 patent applicant argued that “not every type of stress is contrastive stress,” Petitioners fail to identify ***any*** disclosure in the prosecution history where the applicant disavowed the de-emphasis of words.

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 39 of 53

92. As discussed above, contrary to Petitioners' arguments, the plain and ordinary meaning of "contrastive stress" is simply any stress which contrasts with other words in the text string. This plain and ordinary meaning is further supported by extrinsic evidence. For example, there are long held theories and debates around the concept of "contrastive stress" from Fred Dretske (*Contrastive Statements*, The Philosophical Review, Volume. 81, Number 4, Pages 411-437 (October 1979), attached hereto as Exhibit 2009), to his critics (Steven E Boer, *Meaning and Contrastive Stress*, The Philosophical Review, Volume. 88, Number 2, Pages 263-298 (April 1979), attached hereto as Exhibit 2009). More recently, a study published in the Journal of Speech, Language, and Hearing Research examined how speakers with dysarthria and healthy controls signal contrastive stress. Ex. 2008 (Patel and Campellone, *Acoustic and Perceptual Cues to Contrastive Stress in Dysarthria*, Journal of Speech, Language, and Hearing Research, Volume 52, Number 1, Pages 206-222 (February 2009), available at https://doi.org/10.1044/1092-4388(2008/07-0078)). There, the researchers found that while both groups used prosodic cues like pitch, intensity, and duration to emphasize stressed words, they also **markedly decreased these cues** for neighboring non-stressed words. *Id*. This relative reduction in prosodic features made the stressed word stand out more prominently to listeners. *Id*. This suggests that contrastive stress is a relative phenomenon: it's

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 40 of 53

not solely about increasing prominence on the target word but also about reducing prominence on surrounding words to create a clear contrast.

93. In short, Walker in view of Matsumoto does not disclose or suggest the limitation "assigning contrastive stress to the identified first portion of the first text string, but not to the identified second portion of the first text string;" because: 1) a POSITA would not be motivated to combine Walker and Matsumoto; 2) Walker and Matsumoto do not disclose or suggest "assigning contrastive stress to the identified first portion of the first text string;" and 3) Walker and Matsumoto do not disclose or suggest "but not [applying contrastive stress] to the identified second portion of the first text string."

d. Limitation 1[d]: generating, using at least one computer system, speech synthesis output to render the plurality of text strings as speech having the assigned contrastive stress; and Limitation 1[e]: providing the speech synthesis output for the speech-enabled application

94. Walker in view of Matsumoto does not disclose or suggest limitations 1[d] or 1[e].

95. Once again, Petitioners do not assert that either Walker or Matsumoto alone disclose or suggest limitations 1[d] or 1[e], instead asserting that Walker in view of Matsumoto discloses or suggests these limitations. Petition at 33-35. Once again, Petitioners are wrong. As discussed above, a POSITA would not have been motivated to combine or modify Walker and/or Matsumoto, and Petitioners offer no

additional reasoning to justify this combination for these limitations, beyond those already discussed above. *See id*. Thus, as with the above limitations, Walker in view of Matsumoto does not disclose or suggest limitations 1[d] or 1[e].

2. Independent Claim 8

96. The steps of Claim 8 of the ’671 patent are reproduced below:

> 8[pre]: At least one non-transitory computer-readable storage medium encoded with a plurality of computer-executable instructions that, when executed, perform a method for use with a speech-enabled application, the method comprising:
>
> 8[a]: receiving, from the speech-enabled application, input comprising a plurality of text strings;
>
> 8[b]: identifying a first portion of a first text string of the plurality of text strings as differing from a corresponding first portion of a second text string of the plurality of text strings, and a second portion of the first text string as not differing from a corresponding second portion of the second text string;
>
> 8[c]: assigning contrastive stress to the identified first portion of the first text string, but not to the identified second portion of the first text string;
>
> 8[d]: generating speech synthesis output to render the plurality of text strings as speech having the assigned contrastive stress; and
>
> 8[e]: providing the speech synthesis output for the speech-enabled application.

97. As shown above, claim 8 of the ’671 patent is a system version of the analogous method claim 1. Indeed, Petitioners cite to the same arguments for obviousness that they made for claim 1 of the ’671 patent for each respective limitation of claim 8. *See* Petition at 38 (“Walker in combination with Matsumoto

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 42 of 53

teaches Elements 8[a]-8[e] for the same reasons discussed in §§VIII.A.2.b-VIII.A.2.f"). Thus, Walker in view of Matsumoto does not disclose or suggest claim 8's limitations for the same reasons discussed above for the analogous limitations of claim 1 of the '671 patent.

3. Independent Claim 15

98. The steps of Claim 15 of the '671 patent are reproduced below:

> 15[pre]: A method for generating speech output via a speech-enabled application, the method comprising:
>
> 15[a]: generating, using at least one computer system executing the speech-enabled application, a plurality of text strings, each of the plurality of text strings corresponding to a portion of a desired speech output;
>
> 15[b]: inputting the plurality of text strings to at least one software module configured to identify a first portion of a first text string of the plurality of text strings as differing from a corresponding first portion of a second text string of the plurality of text strings, and a second portion of the first text string as not differing from a corresponding second portion of the second text string;
>
> 15[c]: receiving, from the at least one software module, speech synthesis output to render the plurality of text strings with contrastive stress assigned to the identified first portion of the first text string, but not to the identified second portion of the first text string; and
>
> 15[d]: generating, using the speech synthesis output, an audio speech output corresponding to the desired speech output.

99. As shown above, claim 15 of the '671 patent is similar to method claim 1, with the additional requirements that the method be performed on a computer system utilizing at least one software module. And as with claim 8 of the '671 patent, Petitioners cite to the same arguments for obviousness that they made for

claim 1 of the ’671 patent for each respective limitation of claim 15. *See* Petition at 39-46. Thus, Walker in view of Matsumoto does not disclose or suggest these limitations for the same reasons discussed above for the analogous limitations of claim 1 of the ’671 patent.

1. Dependent Claims 7 And 14

100. Claims 7 and 14 depend on independent claims 1 and 8, and thus Petitioners have failed to demonstrate that Walker in combination with Matsumoto discloses or renders obvious these claims for the same reasons discussed with respect to claims 1 and 8 of the ’671 Patent.

B. Ground 1B: Petitioners Fail To Show That Claims 1, 7-8, And 14-15 Would Have Been Obvious Based On Walker In View Of Matsumoto, And Bellegarda

101. Recognizing that Walker and Matsumoto fail to disclose or suggest limitations 1[c], 8[c], and 15[c] (regarding ***not*** applying contrastive stress to portions of the first text string which do not differ from portions of the second text string), Petitioners attempt to piece together a ***third*** reference, Bellegarda, to argue that these limitations would have been obvious in view of the combination of Walker, Matsumoto, and Bellegarda. Petition at 70-72. I disagree and as discussed below, a POSITA would not be motivated to combine or modify the references to achieve the invention of the ’671 patent. Moreover, even if a POSITA were to combine the

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 44 of 53

references, the combination ***still*** does not disclose multiple claim limitations, including limitations 1[c], 8[c], and 15[c].

102. First, a POSITA would not be motivated to combine or modify the references. As discussed above, a POSITA would not be motivated to combine Walker and Matsumoto to achieve the invention of the '671 patent. Walker and Matsumoto have different purposes, with Walker directed to correcting pronunciation errors in particularized text responses, while Matsumoto is directed to emphasizing and de-emphasizing entire text strings when processing a series of subsequent text strings with nearly identical sentence structures. *Id*.

103. Nor would a POSITA be motivated to combine Bellegarda with Walker and/or Matsumoto, let alone combine these references to achieve the invention of the '671 patent. Bellegarda is directed to evaluating a newly-entered sentence, comparing it to previously-entered sentences, and assigning a "word prominence" value to words which appear more frequently in prior sentences. *See, e.g*., Bellegarda at Abstract:

> A method for generating speech that sounds more natural may comprise generating synthesized speech having certain word prominence characteristics and applying a semantically-driven word prominence assignment model to specify word prominence consistent with the way humans assign word prominence. A speech representative of a current sentence is generated. The determination is made whether information in the current sentence is new or previously given in accordance with a semantic relationship between the current sentence and a number of preceding sentences. A word prominence is assigned to a word in the current sentence in accordance with the information determination.

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 45 of 53

104. According to Petitioners, "[a] POSITA would have found it obvious based on Bellegarda's teaching to set the minimum variation coefficient to 1.0 so that there would be no change to the variation coefficients for words/phrases determined to be the same." Petition at 72. As Petitioners argue:

> Only adjusting the prosody of words determined to be different between two text strings would have been obvious to a POSITA because it would have amounted to the simple substitution of one known element (Bellegarda teaches not adjusting the prosody of corresponding words that are the same between two text strings) for another (Matsumoto teaches deemphasizing corresponding words that are the same between two text strings) to obtain predictable results (in the Walker-Matsumoto-Bellegarda combination corresponding words that different between two text strings would continue to receive contrastive stress). Schmandt ¶349.

Id.

105. In short, Petitioners seek to add Bellegarda to Walker and Matsumoto to solve the problem of Matsumoto's process emphasizing / de-emphasizing the ***entirety*** of subsequent text strings (rather than only applying contrastive stress to just the differing portions). But Petitioners' arguments lack merit. As an initial matter, Petitioners mischaracterize Bellegarda's disclosure and purpose. Like Matsumoto, Bellegarda is about assessing ***subsequently*** entered sentences and assigning variation values to words in subsequent sentences based on the similarity of those words in prior sentences. Indeed, as Bellegarda explains, the focus of its process is to provide a "semantically-driven word prominence assignment model to assign word prominence characteristics consistent with the way humans assign work

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 46 of 53

prominence." Bellegarda at 4:18-20. In short, Bellegarda is directed to identifying which words should be ***emphasized*** in ***subsequently***-entered sentences, by assessing the context of the previously-entered sentences and identifying which newly-entered words carry new information for the listener. Whereas, the '671 has no such limitation as to subsequent ordering of sentences and rather, can be applied to arbitrary pairs of text strings. *See id*. at Background.

106. Put simply, like Matsumoto, Bellegarda is limited to adding contrastive stress to ***subsequently*** entered sentences, and thus does not add contrastive stress to the initially-entered sentence (i.e., the "first portion of the first text string" as required by the claims). Moreover, Petitioners only assert that a POSITA ***could*** combine Matsumoto's process of emphasizing / de-emphasizing entire sentences with Bellegarda's process of only adding prominence to differing words, but fail to offer any reasons why a POSITA ***would*** make such a modification. *See Belden Inc. v. Berk-Tek LLC*, 805 F.3d 1064, 1073 (Fed. Cir. 2015) (obviousness concerns "whether a skilled artisan not only *could* have made but *would* have been motivated to make the combinations or modifications of prior art to arrive at the claimed invention."). Petitioners offer no cogent reason why a POSITA would combine ***three*** different prior art references with entirely different purposes and processes, and Petitioners' attempt to cobble together a hypothetical that meets the claims is infected with hindsight.

107. Petitioners further argue that:

> A POSITA would have also recognized that not adjusting the prosody of corresponding words that are the same between two text strings would have reduced the computations required by the Walker-Matsumoto-Bellegarda combinations inasmuch as the TTS engine would not have needed to adjust the speed/duration, pitch and volume coefficients of corresponding words that are the same between two text strings.

Petition at 73. Petitioners attempt to combine the references does not make sense because it introduces a problem of processing power. But Petitioners do not explain how adding an ***additional*** text-to-speech processing engine would ***reduce*** the processing required by the total combined system. While the system would no longer need to adjust the speed/duration, pitch and volume coefficients of corresponding words that are the same between two text strings, the system would ***add*** the computational loads necessary to run Bellegarda's semantically-driven word prominence assignment model. Indeed, as Bellegarda explains, "[d]etermining which words are semantically related is, however, ***very complex*** due to the multi-faceted nature of semantic relationships." Bellegarda at 2:65-67 (emphasis added). Petitioners also fail to account for the significant differences in how Bellegarda and Matsumoto determine which words differ between compared text strings. For example, as discussed above, Matsumoto generates a matrix of sentences to compare words in the precise numerical order as words in the previously entered sentences. Conversely, Bellegarda recognizes that words may still be related to each other even

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 48 of 53

if they differ, and thus calls for a complex semantically-driven evaluation of each subsequent phrase. *See, e.g.,* Bellegarda at 2:54-58 ("Determining which words express the same concept involves not only words that are synonyms, but more generally, words that are semantically related to one another.").

108. Second, even if a POSITA would have been motivated to modify or combine the three references, the combination ***still*** does not disclose multiple claim limitations, including limitations 1[c], 8[c], and 15[c]. As an initial matter, Petitioners do not assert that Bellegarda discloses or suggests any other limitations beyond 1[c], 8[c], and 15[c], and thus the Walker-Matsumoto-Bellegarda combination fails to disclose the other limitations (i.e., limitations 1[a], 1[b], 1[d], and 1[e]) for the reasons discussed above for each respective limitation. Moreover, the addition of Bellegarda does not even solve the Walker-Matsumoto combination's deficiency with respect to limitations 1[c], 8[c], and 15[c]. As discussed above, Bellegarda is directed to emphasizing ***different*** words in ***subsequently-***entered sentences, while the second part of limitations 1[c], 8[c], and 15[c] relate to ***not*** adding contrastive stress to ***similar*** words in the ***first*** text string entered. Bellegarda says nothing about similar words and whether or not to stress them.

109. In short, a POSITA would not be motivated to combine or modify the references to achieve the invention of the '671 patent. Moreover, even if a POSITA

were to combine the references, the combination ***still*** does not disclose multiple claim limitations, including limitations 1[c], 8[c], and 15[c].

C. Grounds 2A through 3B: Petitioners Fail To Show That The Dependent Claims Would Have Been Obvious Based On Asserted Art

110. Petitioners' remaining grounds relate to arguments regarding the validity of certain dependent claims. Because these claims depend on independent claims 1, 8, and 15 discussed above, Petitioners have failed to demonstrate that the asserted prior art combinations disclose or renders obvious these claims for the same reasons discussed with respect to claims 1, 8, and 15 of the '671 Patent.

D. Objective Indicia of Non-Obviousness

111. As discussed further below, objective indicia of non-obviousness, including, for example, evidence of commercial success, long-felt but unsolved needs, and failure of others, further demonstrate the non-obviousness of the '671 patent.

112. Public information regarding Samsung's annual revenue shows that Samsung's annual revenue has been hundreds of billions of dollars per year for over a decade. *See, e.g.,* Demi Sher, *Samsung Statistics* (February 19, 2025) (available at https://www.investing.com/academy/statistics/samsung-facts/).

113. Additionally, while Samsung's Galaxy-devices have a number of features, there is a nexus between Samsung's significant sales and the improved text-to-speech functionality provided by the '671 patent. For example, Samsung

advertises the benefits of its text-to-speech functionality, including its "***revolutionary impact on communication***." *See, e.g.,* Text To Video AI (last accessed May 21, 2025) (available at https://texttovideoai.org/blog/samsung-text-to-speech); *see also* Galaxy Store (last accessed May 21, 2025) (available at https://galaxystore.samsung.com/detail/com.samsung.SMT) (advertising Samsung's text-to-speech application).

114. As another example of objective indicia of non-obviousness, the inventions of the '671 patent provide a solution to a long felt but unresolved need and demonstrate the failure of others to arrive at the solution prior to the invention of the '671 patent. For example, prior to the filing of the '671 patent and its related applications, the mobile phone and tablet industry suffered from text-to-speech applications which provided monotone and unnatural speech output. As discussed above, the '671 patent explains that prior art text-to-speech processes "produce[] speech output that is relatively monotone and flat, lacking the naturalness and emotional expressiveness of the naturally produced human speech…." '671 patent at 8:11-15; *see also id.,* 8:34-36 ("conventional automatic speech synthesis technologies have not taken advantage of contrastive stress as an opportunity to improve intelligibility, naturalness and effectiveness of machine generated speech."). Indeed, the very prior art references Petitioners rely upon in this Petition confirm this long-felt need. *See, e.g.*, Matsumoto ¶[0005] ("Further, in the case of

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 51 of 53

the speech in monotone, the hearer cannot concentrate on a want-to-hear point in the speech and might fail to hear the want-to-hear point."); Walker ¶[0003] ("Even small speech differences in what humans are accustomed to in normal conversation can cause large differences in how humans perceive the quality and naturalness of computer generated speech.").

115. Further, the '671 patent has been cited dozens of times by patent examiners and other applicants, as can be seen through any number of online and publicly available tools for determining such forward "citations." *See, e.g.*, '671 Patent's Google Patent Page (last accessed May 21, 2025) (available at https://patents.google.com/patent/US8682671B2/en?oq=8%2c682%2c671).

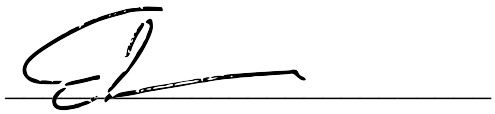
116. Moreover, the owners of these forward citing patents include Samsung's direct competitors. *See id*. (showing citations to patent publications assigned to competitors such as Motorola, Inc. and Apple Computer, Inc.).

117. In short, objective indicia of non-obviousness, including, for example, evidence of commercial success, long-felt but unsolved needs, failure of others, and unexpected results, demonstrate the non-obviousness of the '671 patent.

***

I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true.

May 27, 2025



Eric Koskinen

Cerence Operating Company, Ex. 2005
Samsung Electronics Co., Ltd. etc. v. Cerence Operating Company, IPR2025-00458
Page 53 of 53