

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

PATENT OWNER'S PRELIMINARY RESPONSE

TABLE OF CONTENTS

I. INTRODUCTION	1
II. PROCEDURAL HISTORY	4
III. LEVEL OF ORDINARY SKILL IN THE ART	6
IV. CLAIM CONSTRUCTION	6
V. LEGAL STANDARDS.....	7
VI. PETITIONERS’ GROUNDS FOR REQUESTING <i>INTER PARTES</i> REVIEW ARE DEFICIENT	9
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	9
1. Walker and Matsumoto Are Directed To Fundamentally Different Approaches To Text-to-Speech Than The ’671 Patent.....	9
a. The ’671 Patent.....	9
b. Walker.....	12
c. Matsumoto	13
2. Because Walker and Matsumoto Are Directed To Fundamentally Different Approaches To Text-To-Speech Than The ’671 Patent, They Fail To Disclose Or Render Obvious Multiple Claim Limitations	16
a. Independent Claim 1	17
(1) Limitation 1[a]: receiving, from the speech-enabled application, input comprising a plurality of text strings;	17
(2) 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;	24

(3) 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; . 27	
(4) 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	34
b. Independent Claim 8.....	35
c. Independent Claim 15.....	36
d. Dependent Claims 7 And 14.....	37
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.....	37
C. Ground 2A: Petitioners Fail To Show That Claims 2-5, 9-12, And 16-19 Would Have Been Obvious Based On Walker In View Of Matsumoto, And Malsheen	43
D. Ground 2B: Petitioners Fail To Show That Claims 2-5, 9-12, And 16-19 Would Have Been Obvious Based On Walker In View Of Matsumoto, Bellegarda, And Malsheen	43
E. Ground 3A: Petitioners Fail To Show That Claims 6, 13, And 20 Would Have Been Obvious Based On Walker In View Of Matsumoto, And Marple	44
F. Ground 3B: Petitioners Fail To Show That Claims 6, 13, And 20 Would Have Been Obvious Based On Walker In View Of Matsumoto, Bellegarda, and Marple	44
G. Objective Indicia Support Non-Obviousness	44
VII. CONCLUSION	48

TABLE OF AUTHORITIES

Cases

<i>Abbott Laboratories v. Sandoz, Inc.</i> , 544 F.3d 1341 (Fed. Cir. 2008)	21, 41
<i>Advanced Cardiovascular Systems, Inc. v. Medtronic, Inc.</i> , 265 F.3d 1294 (Fed. Cir. 2001)	33
<i>Arctic Cat Inc. v. Bombardier Recreational Prods. Inc.</i> , 876 F.3d 1350 (Fed. Cir. 2017)	7
<i>Belden Inc. v. Berk-Tek LLC</i> , 805 F.3d 1064 (Fed. Cir. 2015)	8, 19, 40
<i>Dennison Mfg. Co. v. Panduit Corp.</i> , 475 U.S. 809 (1986)	22, 41
<i>Graham v. John Deere Co. of Kansas City</i> , 383 U.S. 1 (1966)	7, 45
<i>Granite Constr. Co. v. United States</i> , 962 F.2d 998 (Fed. Cir. 1992)	8
<i>In re Cyclobenzaprine Hydrochloride Extended-Release Capsule Patent Litig.</i> , 676 F.3d 1063 (Fed. Cir. 2012)	7
<i>In re Kahn</i> , 441 F.3d 977 (Fed. Cir. 2006)	8
<i>In re Magnum Oil Tools Int’l, Ltd.</i> , 829 F.3d 1364 (Fed. Cir. 2016)	8, 18, 22
<i>In re Stepan Co.</i> , 868 F.3d 1342 (Fed. Cir. 2017)	7
<i>Intelligent Bio-Systems, Inc. v. Illumina Cambridge Ltd.</i> , 821 F.3d 1359 (Fed. Cir. 2016)	8
<i>MBO Labs., Inc. v. Becton, Dickinson & Co.</i> , 474 F.3d 1323 (Fed. Cir. 2007)	32

Polaris Indus., Inc. v. Arctic Cat, Inc.,
882 F.3d 1056 (Fed. Cir. 2018)..... 45

Starhome GmbH v. AT & T Mobility LLC,
743 F.3d 849 (Fed. Cir. 2014)..... 30

Virtek Vision International ULC v. Assembly Guidance Systems, Inc.,
97 F.4th 882 (Fed. Cir. 2024)..... 8

WBIP, LLC v. Kohler Co.,
829 F.3d 1317 (Fed. Cir. 2016)..... 7, 45

Statutes

35 U.S.C. § 103..... 7

PATENT OWNER’S EXHIBIT LIST

No.	Description
2001	Docket Navigator Stay Statistics
2002	March 28, 2025, 2 nd Amended DCO
2003	Docket Navigator Median Time-to-Trial Statistics
2004	Samsung’s September 9, 2024, Invalidity Contentions
2005	Expert declaration of Eric Koskinen (“Koskinen Decl.”)
2006	Koskinen CV
2007	Adam Rogowski, <i>Scenario-Based Programming of Voice-Controlled Medical Robotic Systems</i> , Sensors 2022, 9520 (December 6, 2022)) (“we cannot always assume that the operator utters only the commands contained in the defined sublanguage.”).
2008	Patel and Campellone, <i>Acoustic and Perceptual Cues to Contrastive Stress in Dysarthria</i> , 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)).
2009	Fred Dretske (<i>Contrastive Statements</i> , The Philosophical Review, Volume. 81, Number 4, Pages 411-437 (October 1979)
2010	Steven E Boer, <i>Meaning and Contrastive Stress</i> , The Philosophical Review, Volume. 88, Number 2, Pages 263-298 (April 1979)).
2011	District Court Infringement Contentions

I. Introduction

Petitioners Samsung Electronics Co., Ltd. and Samsung Electronics, America, Inc., (collectively “Samsung” or “Petitioners”) petition for *inter partes* review (“IPR”) of U.S. Patent No. 8,682,671 (the “’671 Patent”) should be denied for four reasons.

First, Petitioners’ primary references are directed to entirely different text-to-speech processes than the ’671 patent. Indeed, Petitioners’ primary reference, Walker, is directed to the type of prior art systems and methods that the ’671 patent explicitly sought to improve. For example, the ’671 patent claims cover methods and systems designed to take text strings and generate 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. As the ’671 patent explains, 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.”).

Conversely, 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 a text string is being pronounced correctly (when said text string could be read in multiple different ways depending on the context). *See, e.g.*, Walker, ¶[0001] (“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 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. Notably, while Walker is a U.S. Patent Publication which published on December 6, 2001, it was subsequently abandoned and never issued into a patent.

Petitioners’ second core reference, Matsumoto, is also directed to a fundamentally different (and largely impractical) text-to-speech process than the ’671 patent. Indeed, while Matsumoto is a U.S. Patent Publication which later issued into a patent, it subsequently expired due to the patent owner’s failure to pay maintenance fees. Specifically, Matsumoto describes a speech synthesizer which

accepts input sentences, stores the input sentences into a matrix, and then compares words in the new sentences with words in the *precise* numerical order as words in the prior sentences. Matsumoto's speech synthesizer then utilizes the calculated variation degree to apply variations to the speed, volume, and pitch of each of the words in the subsequently-analyzed sentence. *See, e.g.*, Matsumoto ¶¶[0093]-[0103] (walking through operational example).

Because Matsumoto's process involves a matrix-style analysis of previously-entered sentences, its text-to-speech modification can *only* functionally be applied to subsequently-entered sentences, and *only* when the subsequently-entered sentences are nearly identical in structure to the previously-entered sentences. Moreover, while the '671 patent allows for automatic application of contrastive stress in text strings, Matsumoto contemplates *significant* user input and direction as part of the speech synthesis process, including having the user select how many sentences should be analyzed when calculating the degree of variation, setting minimum and maximum "variation coefficients," setting the "collation mode" (i.e., collating all accumulated sentences against the newly input sentence, or just the last-entered sentence against the newly-input sentence), and setting the interpolation interval. Matsumoto ¶¶[0056]-[0067].

Additionally, 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 sentence has its speed, volume, and/or pitch adjusted. Matsumoto ¶¶[0087]-[0092].

Second, because Petitioners’ primary references are directed to entirely different text-to-speech processes than the ’671 patent, the combination of these references fail to disclose or suggest multiple claim limitations.

Third, a POSITA would not be motivated to combine or modify Petitioners’ references, and would not have had a reasonable expectation of success in combining or modifying these references to achieve the ’671 patent’s invention.

Fourth, 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.¹

II. Procedural History

On March 15, 2024, Patent Owner sued Petitioners in district court for infringing Patent Owner’s patents, including the ’671 patent. *Cerence Operating Company v. Samsung Electronics Co. Ltd. et al.*, 2:24-cv-00181, ECF No. 1 (E.D. Tex. March, 15 2024).

¹ Pursuant to the Director’s March 26, 2025, memorandum regarding Interim Processes for PTAB Workload Management (“March 26, 2025, Memo”), on April 28, 2025, Patent Owner Cerence Operating Company submitted a briefing requesting that the Director exercise discretion under 35 U.S.C. § 314(a) and issue a decision denying institution of the Petition. As discussed in that briefing, the *Fintiv* factors all weigh in favor of discretionary denial, including the weak nature of the Petition’s merits, which are addressed herein.

On January 21, 2025, Petitioners filed their petition for *Inter Partes* Review, challenging claims 1-20 of the '671 patent. Petitioners structure their asserted grounds for unpatentability as follows:

- **Ground 1A:** Obviousness of claims 1, 7-8, and 14-15 in view of Walker² and Matsumoto³;
- **Ground 1B:** Obviousness of claims 1, 7-8, and 14-15 in view of Walker, Matsumoto, and Bellegarda⁴;
- **Ground 2A:** Obviousness of claims 2-5, 9-12, and 16-19 in view of Walker, Matsumoto, and Malsheen⁵;
- **Ground 2B:** Obviousness of claims 2-5, 9-12, and 16-19 in view of Walker, Matsumoto, Bellegarda, and Malsheen;
- **Ground 3A:** Obviousness of claims 6, 13, and 20 in view of Walker, Matsumoto, and Marple⁶;
- **Ground 3B:** Obviousness of claims 6, 13, and 20 in view of Walker, Matsumoto, Bellegarda, and Marple.

² US20010049602 (“Walker”) (Ex-1005).

³ US20070233492 (“Matsumoto”) (Ex-1006).

⁴ US7,313,523 (“Bellegarda”) (Ex-1014).

⁵ US5634084 (“Malsheen”) (Ex-1007).

⁶ US20080195391 (“Marple”) (Ex-1008).

III. Level of Ordinary Skill in the Art

Petitioners propose that a person of ordinary skill in the art at the time of the priority date would have “possessed a Bachelor of Science (or 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. “A greater amount of education could compensate for less work experience, and vice versa.” *Id.*

Patent Owner reserves the right to offer a different level of ordinary skill in this or in other proceedings concerning the '671 patent. However, Patent Owner does not believe that the issues presented in this Preliminary Response turn on the precise level of ordinary skill in the art and believes that the Board could assume Petitioners' proposed level of ordinary skill for the purposes of deciding whether to institute.

IV. Claim Construction

Petitioners state that “no express constructions of the claims are necessary” and that “the claims should be given their plain and ordinary meaning.” Petition at 14. Patent Owner agrees that no construction of the terms is required to decide the patentability of the challenged claims. As Patent Owner explains further below, Petitioners' invalidity arguments should be rejected when applying the plain and ordinary meaning to the relevant claim terms.

V. Legal Standards

The U.S. Supreme Court set forth the framework for applying the statutory language of 35 U.S.C. § 103 in *Graham v. John Deere Co. of Kansas City*:

Under § 103, the scope and content of the prior art are to be determined; differences between the prior art and the claims at issue are to be ascertained; and the level of ordinary skill in the pertinent art resolved. Against this background, the obviousness or nonobviousness of the subject matter is determined. Such secondary considerations as commercial success, long felt but unsolved needs, failure of others, etc., might be utilized to give light to the circumstances surrounding the origin of the subject matter sought to be patented.

Graham v. John Deere Co. of Kansas City, 383 U.S. 1, 17–18 (1966).

An obviousness determination requires finding that a POSITA would have been motivated to modify or combine teachings in the prior art. *In re Stepan Co.*, 868 F.3d 1342, 1345–46 (Fed. Cir. 2017). “Whether an ordinarily skilled artisan would have been motivated to modify the teachings of a reference is a question of fact.” *WBIP, LLC v. Kohler Co.*, 829 F.3d 1317, 1327 (Fed. Cir. 2016) (citations omitted). “[W]here a party argues a skilled artisan would have been motivated to combine references, it must show the artisan ‘would have had a reasonable expectation of success from doing so.’” *Arctic Cat Inc. v. Bombardier Recreational Prods. Inc.*, 876 F.3d 1350, 1360–61 (Fed. Cir. 2017) (quoting *In re Cyclobenzaprine Hydrochloride Extended-Release Capsule Patent Litig.*, 676 F.3d 1063, 1068–69 (Fed. Cir. 2012)).

Moreover, 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.” *Belden Inc. v. Berk-Tek LLC*, 805 F.3d 1064, 1073 (Fed. Cir. 2015). The “mere fact that these possible arrangements existed in the prior art” does not provide a motivation to substitute the arrangement in one prior art embodiment with the arrangement in another.” *Virtek Vision International ULC v. Assembly Guidance Systems, Inc.*, 97 F.4th 882, 887 (Fed. Cir. 2024). Because a motivation to modify or combine is part of the *prima facie* case for obviousness that must be set forth in the petition, a petitioner should not be permitted to present “an entirely new rationale to explain why one of skill in the art would have been motivated to combine” in its reply. *Intelligent Bio-Systems, Inc. v. Illumina Cambridge Ltd.*, 821 F.3d 1359, 1370 (Fed. Cir. 2016).

“[R]ejections on obviousness grounds cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.” *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006). The Board “may reject even uncontroverted expert testimony when it is intrinsically unpersuasive.” *Granite Constr. Co. v. United States*, 962 F.2d 998, 1006 (Fed. Cir. 1992). In short, the petitioner cannot satisfy its burden of proving obviousness by employing “mere conclusory statements.” *In re Magnum Oil Tools Int’l, Ltd.*, 829 F.3d 1364, 1380 (Fed. Cir. 2016).

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

Petitioners contend that Walker in view of Matsumoto renders obvious claims 1, 7-8, and 14-15 of the '671 patent. Petitioners are wrong. Walker and Matsumoto are directed to entirely different text-to-speech processes than the '671 patent and thus fail to disclose or suggest multiple claim limitations. Moreover, a POSITA would not have been motivated to combine Walker, which describes a process for identifying the correct pronunciation of a word based on the context of the dialogue, with Matsumoto, which describes a process for emphasizing and de-emphasizing *all* of the words in a sentence (without consideration of how any word may be pronounced differently based on the context). Additionally, because Matsumoto's adjustments to the tone, pitch, and speed of any given word may conflict with Walker's adjustment to the pronunciation of said word, a POSITA would not have a reasonable expectation of success in combining these references.

1. Walker and Matsumoto Are Directed To Fundamentally Different Approaches To Text-to-Speech Than The '671 Patent

a. The '671 Patent

As discussed above, 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. Ex. 2005 (Koskinen Decl.), ¶37. As the '671 patent explains, 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.”).

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.

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). Ex. 2005 (Koskinen Decl.), ¶41; *see also* '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.

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. Ex. 2005 (Koskinen Decl.), ¶42; '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 “-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.

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

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. Ex. 2005 (Koskinen Decl.), ¶44. 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, ¶[0001] (“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 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

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. Ex. 2005 (Koskinen Decl.), ¶45; Matsumoto at Abstract. Matsumoto’s speech synthesizer

then utilizes the calculated variation degree to apply variations to the speed, volume, and pitch of each of the words in the subsequently analyzed sentence. *See, e.g.*, Matsumoto ¶¶[0093]-[0103] (walking through operational example).

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. Ex. 2005 (Koskinen Decl.), ¶46. 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

As shown in the above figure, the Matsumoto process involves accumulating 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].

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 *subsequently*-entered sentences (i.e., emphasis / deemphasis is not applied to characters or words in the first sentence entered, only subsequent sentences). Ex. 2005 (Koskinen Decl.), ¶48; *see also* Matsumoto ¶¶[0071]-[0082], [0098]-[0100], [0105].

Matsumoto also contemplates *significant* user input and direction as part of the speech synthesis process, including having the user select how many sentences should be analyzed when calculating the degree of variation (i.e., the “collation range”), setting minimum and maximum “variation coefficients,” setting the “collation mode” (i.e., collating all accumulated sentences against the newly input sentence, or just the last-entered sentence against the newly-input sentence), and setting the interpolation interval. Matsumoto ¶¶[0056]-[0067].

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 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. Ex. 2005 (Koskinen Decl.), ¶50; *see also* 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).

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). Ex. 2005 (Koskinen Decl.), ¶51.

2. Because Walker and Matsumoto Are Directed To Fundamentally Different Approaches To Text-To-Speech Than The '671 Patent, They Fail To Disclose Or Render Obvious Multiple Claim Limitations

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. Ex. 2005 (Koskinen Decl.), ¶56.

a. Independent Claim 1

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;

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.

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].

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

Walker in view of Matsumoto does not disclose or suggest this limitation. Ex. 2005 (Koskinen Decl.), ¶59.

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 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]).

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.

As an initial matter, Petitioners support this argument with nothing more than a conclusory statement from their expert (*see* Petition at 20-21, citing Schmandt Decl., ¶107), but this conclusory, unsupported assertion is entitled to no weight. *See In re Magnum Oil Tools Int’l, Ltd.*, 829 F.3d 1364, 1380 (Fed. Cir. 2016) (“To satisfy

its burden of proving obviousness, a petitioner cannot employ mere conclusory statements.”).

Moreover, Petitioners merely assert that a POSITA *could* modify Walker’s contextual analysis engine with Matsumoto’s text-to-speech engine, but wholly fail to explain *why* a POSITA would be motivated to combine Walker’s processing of single text strings with Matsumoto’s processing of multiple text strings. *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.”). For example, Walker describes a process where a user requests *specific* information over the telephone 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 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*. Ex. 2005 (Koskinen Decl.), ¶62.

As an illustration of this, using Walker’s disclosed system, a user requesting a weather report for Los Angeles may have the below interaction:

- **User:** “What is the weather in Los Angeles today?”
- **Walker’s Output:** “The weather in Los Angeles is windy today.”

Conversely, under Petitioners’ proposed combination of Walker and Matsumoto, the interaction would be:

- **User:** “What is the weather in Los Angeles today?”
- **Walker’s + Matsumoto’s Output:** “The weather in Los Angeles is windy today.” “The weather in Tokyo is sunny today.” “The weather in New York City is cold today.” “The weather in Seoul is cloudy today.”

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.

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. Ex. 2005 (Koskinen Decl.), ¶66. Specifically, Walker discloses a system where a user requests *particular*

information, and the voice application accesses “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. Ex. 2005 (Koskinen Decl.), ¶66. Nor do Petitioners even argue that it would have been obvious 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, Petitioners’ attempt to cobble together a hypothetical that meets the claims is infected with hindsight. *Abbott Laboratories v. Sandoz, Inc.*, 544 F.3d 1341, 1348 (Fed. Cir. 2008) (“in addressing the question of obviousness a judge must not pick and choose isolated elements from the prior art and combine them so as to yield the invention in question if such a

combination would not have been obvious at the time of the invention”) (quoting *Dennison Mfg. Co. v. Panduit Corp.*, 475 U.S. 809, 810 (1986) (per curiam)).

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, which are entitled to no weight. *See In re Magnum Oil Tools Int’l, Ltd.*, 829 F.3d at 1380.

Additionally, as discussed above, Walker is actually directed to preventing incorrect pronunciations of words in a *single* text string (based on the context of the analyzed text string), while Matsumoto is directed to applying emphasis and de-emphasis to *subsequent* text strings (based on a degree of variation between words

or phrases in prior accumulated text strings). Neither Petitioners, nor their expert, explain how these two approaches would be compatible, or how combining these approaches would yield predictable results.

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. Ex. 2005 (Koskinen Decl.), ¶69. 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. In these types of situations, Matsumoto's adjustment of the tone, pitch, and speed of the processed text strings may actually conflict with Walker's core function of ensuring that words are correctly pronounced based on the context of the analyzed text string.

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).

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. Ex. 2005 (Koskinen Decl.), ¶71.

- (2) 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;

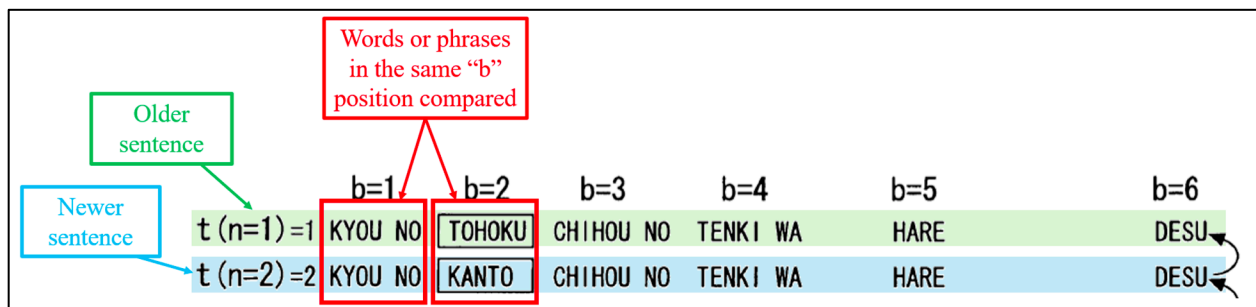
Walker in view of Matsumoto does not disclose or suggest this limitation. Ex. 2005 (Koskinen Decl.), ¶72.

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.*

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 *subsequently*-entered sentences (i.e., emphasis / deemphasis is *not* applied to characters or words in the first sentence entered, only subsequent sentences). Ex. 2005 (Koskinen Decl.), ¶74; *see also* Matsumoto ¶¶[0071]-[0082], [0098]-[0100], [0105].

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

Indeed, Petitioners appear to recognize this issue, as they flip 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)).

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.

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 text strings, based on an analysis of the variation between the newly-entered sentence and subsequently-entered sentences. Ex. 2005 (Koskinen Decl.), ¶78.

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.” Ex. 2005 (Koskinen Decl.), ¶79.

- (3) 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;

Walker in view of Matsumoto does not disclose or suggest this limitation. Ex. 2005 (Koskinen Decl.), ¶80.

Once again, Petitioners do not assert that either Walker or Matsumoto alone disclose or suggest this limitation, instead asserting that Walker in view of

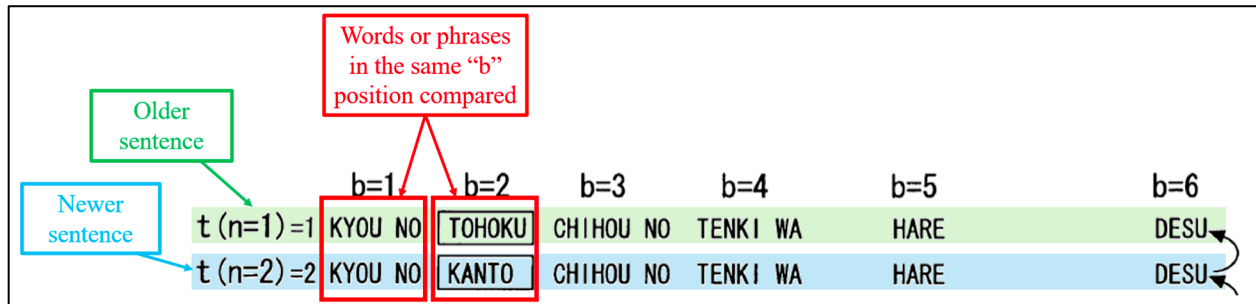
Matsumoto discloses or suggests this limitation. Petition at 27-33. Once again, Petitioners are wrong. As discussed above, a POSITA would not have been motivated to combine or modify Walker and/or Matsumoto, and Petitioners offer no additional reasoning to justify this combination for this limitation, beyond those already discussed above. *See id.*

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.

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)).

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 entered, while Matsumoto is limited to applying emphasis / de-emphasis only to *subsequently* analyzed sentences.

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.*" Ex. 2005 (Koskinen Decl.), ¶85. 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. *Id.*; see also 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).

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.

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. *Starhome GmbH v. AT & T Mobility LLC*, 743 F.3d 849, 857 (Fed. Cir. 2014) (“In the absence of an express intent to impart a novel meaning to claim terms, an inventor's claim terms take on their ordinary meaning.”). Here, the “plain and ordinary meaning” of the

term “contrastive stress,” is simply any stress which contrasts with other words in the text string (Ex. 2005 (Koskinen Decl.), ¶87), and neither Petitioners, nor their expert, have proffered any evidence suggesting a skilled artisan would understand the term differently.

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.

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.

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).

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. Petition at 31-32; *see also Advanced Cardiovascular Systems, Inc. v. Medtronic, Inc.*, 265 F.3d 1294 (Fed. Cir. 2001) (“We note, however, that statements made during prosecution of ancestor patent applications do not automatically bind related patents.”).

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. Ex. 2005 (Koskinen Decl.), ¶92. This plain and ordinary meaning is further supported by extrinsic evidence. For example, 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\(2008/07-0078\)](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 not solely about increasing prominence on the target word but also about reducing prominence on surrounding words to create a clear contrast.

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.” Ex. 2005 (Koskinen Decl.), ¶93.

- (4) 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

Walker in view of Matsumoto does not disclose or suggest limitations 1[d] or 1[e]. Ex. 2005 (Koskinen Decl.), ¶94.

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].

b. Independent Claim 8

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.

As shown above, claim 8 of the '671 patent is a system version of the analogous method claim 1. Ex. 2005 (Koskinen Decl.), ¶97. 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 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. Ex. 2005 (Koskinen Decl.), ¶97.

c. Independent Claim 15

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.

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. Ex. 2005 (Koskinen Decl.), ¶99. 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. Ex. 2005 (Koskinen Decl.), ¶99.

d. Dependent Claims 7 And 14

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. Ex. 2005 (Koskinen Decl.), ¶100.

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

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

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.*

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. Ex. 2005 (Koskinen Decl.), ¶103. 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.

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.

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. Ex. 2005 (Koskinen Decl.), ¶105. 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 word 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. *See id.* at Background; Ex. 2005 (Koskinen Decl.), ¶105.

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). Ex. 2005 (Koskinen Decl.), ¶106. 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. *Abbott Laboratories v. Sandoz, Inc.*, 544 F.3d 1341, 1348 (Fed. Cir. 2008) (“in addressing the question of obviousness a judge must not pick and choose isolated elements from the prior art and combine them so as to yield the invention in question if such a combination would not have been obvious at the time of the invention”) (quoting *Dennison Mfg. Co. v. Panduit Corp.*, 475 U.S. 809, 810 (1986) (per curiam)).

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. But Petitioners do not explain how adding an *additional* text-to-speech processing engine would *reduce* the processing required by the total combined system. Ex. 2005 (Koskinen Decl.), ¶107. 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. *Id.* Indeed, as Bellegarda explains, “[d]etermining which words are semantically related is, however, *very complex* due to the multifaceted 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. Ex. 2005 (Koskinen Decl.), ¶108. 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 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.”).

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]. Ex. 2005 (Koskinen Decl.), ¶108. 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. *Id.* 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]. *Id.* 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. *Id.*

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. Ground 2A: Petitioners Fail To Show That Claims 2-5, 9-12, And 16-19 Would Have Been Obvious Based On Walker In View Of Matsumoto, And Malsheen

Claims 2-5, 9-12, and 16-19 depend on independent claims 1, 8, and 15, and thus Petitioners have failed to demonstrate that Walker in combination with Matsumoto and/or Malsheen discloses or renders obvious these claims for the same reasons discussed with respect to claims 1, 8, and 15 of the '671 Patent. Ex. 2005 (Koskinen Decl.), ¶110.

D. Ground 2B: Petitioners Fail To Show That Claims 2-5, 9-12, And 16-19 Would Have Been Obvious Based On Walker In View Of Matsumoto, Bellegarda, And Malsheen

Claims 2-5, 9-12, and 16-19 depend on independent claims 1, 8, and 15, and thus Petitioners have failed to demonstrate that Walker in combination with Matsumoto, Bellegarda, and/or Malsheen discloses or renders obvious these

claims for the same reasons discussed with respect to claims 1, 8, and 15 of the '671 Patent. Ex. 2005 (Koskinen Decl.), ¶110.

E. Ground 3A: Petitioners Fail To Show That Claims 6, 13, And 20 Would Have Been Obvious Based On Walker In View Of Matsumoto, And Marple

Claims 6, 13, and 20 depend on independent claims 1, 8, and 15, and thus Petitioners have failed to demonstrate that Walker in combination with Matsumoto, and/or Marple discloses or renders obvious these claims for the same reasons discussed with respect to claims 1, 8, and 15 of the '671 Patent. Ex. 2005 (Koskinen Decl.), ¶110.

F. Ground 3B: Petitioners Fail To Show That Claims 6, 13, And 20 Would Have Been Obvious Based On Walker In View Of Matsumoto, Bellegarda, and Marple

Claims 6, 13, and 20 depend on independent claims 1, 8, and 15, and thus Petitioners have failed to demonstrate that Walker in combination with Matsumoto, Bellegarda, and/or Marple discloses or renders obvious these claims for the same reasons discussed with respect to claims 1, 8, and 15 of the '671 Patent. Ex. 2005 (Koskinen Decl.), ¶110.

G. Objective Indicia Support Non-Obviousness

Objective evidence of a patented invention's non-obviousness—such as commercial success, industry praise, long felt need, and failure of others—“may often be the most probative and cogent evidence in the record[]” and “must be

considered in *every* case.” *WBIP, LLC v. Kohler Co.*, 829 F.3d 1317, 1328-37 (Fed. Cir. 2016) (original emphasis) (citing cases); *see also, e.g., Polaris Indus.*, 882 F.3d at 1071-72. Such evidence serves “to ‘guard against slipping into use of hindsight’ ... and to resist the temptation to read into the prior art the teachings of the invention.” *Graham v. John Deere Co. of Kan. City*, 383 U.S. 1, 36 (1966) (citation omitted).

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, demonstrate the non-obviousness of the ’671 patent. For example, as shown in Patent Owner’s infringement contentions served in the parallel District Court case, the Samsung devices accused of infringing the ’671 patent include Samsung’s Galaxy devices, which make up a large share of Samsung’s total profits. *See, e.g., Ex. 2011 (Infringement Contentions)* at 3-4. Moreover, 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/>). 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).

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 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.”). And while Walker and Matsumoto sought to solve some issues related to poor text-to-speech prior art systems and methods, neither of them were successful in solving this problem. Indeed, Walker’s patent application was abandoned, and Matsumoto’s patent expired due to failure of the assignee to even pay the nominal maintenance fees (indicating the assignee’s awareness of the impracticality of the Matsumoto design).

Further, the industry has recognized the value of the claimed solutions in the ’671 patent. This recognition has come in various forms, including: (a) recognition of the claimed solutions in technical and marketing materials; (b) awards; (c) praise; (d) citations and disclosures in later-filed patent applications. Indeed, 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?q=8%2c682%2c671>).

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.).

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.

VII. Conclusion

For the reasons set forth above, the Petition fails to establish a reasonable likelihood that Petitioners would succeed in showing that the prior art discloses or renders obvious any of the challenged claims, and the Board should deny institution on that basis.

Date: May 27, 2025

Respectfully submitted,

/Qi (Peter) Tong/
Qi (Peter) Tong (Reg. No. 74,292)
RUSS AUGUST & KABAT
8080 N. Central Expy, Suite 1503
Dallas, TX 75206
(310) 826-7474

Attorney for Patent Owner

CERTIFICATE OF WORD COUNT

I certify that there are 10,786 words in this paper, excluding the portions exempted under 37 C.F.R. § 42.24(a)(1), according to the word count tool in Microsoft Word.

/Qi (Peter) Tong/
Qi (Peter) Tong (Reg. No. 74,292)

CERTIFICATE OF SERVICE (37 C.F.R. § 42.6(e)(1))

The undersigned hereby certifies that the above document was served on May 27, 2025, by filing this document through the P-TACTS system as well as delivering a copy via electronic mail upon the following attorneys of record for the Petitioner:

Ali R. Sharifahmadian
ali.sharifahmadian@arnoldporter.com
Patrick Reidy
patrick.reidy@arnoldporter.com
Douglas L. Clark
douglas.clark@arnoldporter.com
Olivia Matthews
olivia.mathews@arnoldporter.com
xSamsungCerenceAP@arnoldporter.com

Date: May 27, 2025

Respectfully submitted,

/Qi (Peter) Tong/
Qi (Peter) Tong (Reg. No. 74,292)
RUSS AUGUST & KABAT
8080 N. Central Expy, Suite 1503
Dallas, TX 75206
(310) 826-7474

Attorney for Patent Owner