

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 IPR2025-00460

U.S. Patent No. 8,914,291

Petition for *Inter Partes* Review of

U.S. Patent No. 8,914,291

TABLE OF CONTENTS

PETITIONERS' EXHIBIT LIST	v
LISTING OF CHALLENGED CLAIMS.....	vii
I. INTRODUCTION	1
II. GROUNDS FOR STANDING (37 C.F.R. §42.104(A))	1
III. STATEMENT OF PRECISE RELIEF REQUESTED FOR EACH CLAIM CHALLENGED	1
IV. OVERVIEW OF THE '291 PATENT	3
V. OVERVIEW OF THE PRIMARY PRIOR ART REFERENCES.....	6
A. <i>Walker</i> (Ex-1005).....	6
B. <i>Matsumoto</i> (Ex-1006)	9
VI. LEVEL OF ORDINARY SKILL IN THE ART	13
VII. CLAIM CONSTRUCTION	13
VIII. DETAILED EXPLANATION OF GROUNDS.....	14
A. Ground 1A: <i>Walker</i> in Combination With <i>Matsumoto</i> And <i>Malsheen</i> Renders Obvious Claims 1-4, 6-11, 13-17, and 19-20.....	14
1. Independent Claim 1	14
a. 1[pre].....	14
b. 1[a].....	15
c. 1[b].....	18
(i) "Token" as used in the '291 claims and specification.....	19
(ii) Motivation to Combine <i>Walker</i> and <i>Matsumoto</i> ...	21

(iii)	Example text weather report that a POSITA would have understood to be taught by the <i>Walker-Matsumoto</i> combination	25
(iv)	Motivation to combine <i>Walker-Matsumoto</i> with <i>Malsheen</i>	27
(v)	Application of the <i>Walker-Matsumoto-Malsheen</i> combination to the example text weather report	30
d.	1[c]	34
e.	1[d]	41
f.	1[e]	43
2.	Independent Claim 8	44
a.	8[pre]	44
b.	8[a]-8[b]	44
c.	8[c]-[g]	46
3.	Independent Claim 14	47
a.	14[Pre]	47
b.	14[a]-[e]	49
4.	Dependent Claims 2-4, 6-7, 9-11, 13, 15-17, and 19-20	50
a.	Claim 2	50
(i)	2[a]	50
(ii)	2[b]	53
b.	Claim 3	55
c.	Claim 4	56

d.	Claim 6.....	57
(i)	6[a].....	57
(ii)	6[b]-6[c].....	59
e.	Claim 7.....	60
f.	Claims 9-11, 13.....	62
g.	Claims 15-17 and 19-20	62
B.	Ground 1B: <i>Walker</i> in Combination With <i>Matsumoto</i> , <i>Malsheen</i> , and <i>Bellegarda</i> Renders Obvious Claims 1-4, 6-11, 13-17, and 19-20.....	63
1.	1[c], 8[e], 14[c]	63
2.	All Other Features of Claims 1-4, 6-11, 13-17, and 19-20.....	66
IX.	SECONDARY CONSIDERATIONS	68
X.	THE DISCRETIONARY FACTORS FAVOR INSTITUTING TRIAL	69
A.	35 U.S.C. §314(a).....	69
1.	Stay.....	69
2.	Trial Date	70
3.	Diligence/Investment	71
4.	Overlap.....	72
5.	Parties.....	72
6.	Other considerations.	73
B.	35 U.S.C. §325(d).....	73
XI.	MANDATORY NOTICES UNDER 37 C.F.R. § 42.8.....	73
A.	Real Parties-in-Interest.....	73

1.	Related Matters	73
B.	Lead and Backup Counsel.....	74
C.	Service Information.....	75
D.	Power of Attorney	75
XII.	FEES	75

PETITIONERS' EXHIBIT LIST

Exhibit No.	DESCRIPTION
1001	U.S. Patent 8,914,291 (“’291Pat”)
1002	Declaration of Christopher M. Schmandt (“Schmandt”)
1003	Curriculum Vitae of Christopher M. Schmandt
1004	File History of U.S. Patent No. 8,914,291 (“’291FH”)
1005	U.S. Patent App. No. 2001/0049602 (“Walker”)
1006	U.S. Patent App. No. 2007/0233492 (“Matsumoto”)
1007	U.S. Patent 5,634,084 (“Malsheen”)
1008	W3C Speech Synthesis Markup Language (SSML) Ver. 1.1 WD4 , September 4, 2007 (“W3C”)
1009	U.S. Patent 7,010,489 (“Lewis”)
1010	U.S. Patent 7,177,807 (“Schmid”)
1011	U.S. Patent 6,865,533 (“Addison”)
1012	U.S. Patent 7,725,321 (“Bennett”)
1013	File History of U.S. Patent No. 8,447,610 (“’610FH”)
1014	U.S. Patent 7,313,523 (“Bellegarda”)
1015	U.S. Provisional Application No. 60/205,000
1016	<i>Speech Synthesis</i> , in <i>Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition</i> (Daniel Jurafsky & James H. Martin Oct. 13, 2007) (“Jurafsky”)

Exhibit No.	DESCRIPTION
1017	U.S. Patent 8,682,671 (“’671Pat”)
1018	U.S. Patent 8,447,610 (“’610Pat”)
1019	Redline comparison of the specification of the ’291 patent (as available on www.patentcenter.uspto.gov) to the specification of the ’610 patent (as available on www.patentcenter.uspto.gov)
1020	Docket Control Order from <i>Cerence Operating Co. v. Samsung Electronics Co., Ltd. et al.</i> , No. 2:24-cv-00181-JRG-RSP (E.D. Tex.) (Dkt. 52)
1021	Cover Pleading of Plaintiff Cerence Operating Company’s Disclosure of Asserted Claims and Infringement Contentions (July 18, 2024)
1022	Summary of all patents/applications in the ’291 patent’s family

LISTING OF CHALLENGED CLAIMS

Claim 1	
1[pre]	A method for providing speech output for a speech-enabled application, the method comprising:
1[a]	receiving from the speech-enabled application a text input comprising a text transcription of a desired speech output;
1[b]	identifying at least one first portion of at least one token of the text input that differs from at least one corresponding first portion of at least one other token of the text input, and at least one second portion of the at least one token that does not differ from at least one corresponding second portion of the at least one other token;
1[c]	assigning contrastive stress to the identified at least one first portion of the at least one token, but not to the identified at least one second portion of the at least one token;
1[d]	generating, using at least one computer system, an audio speech output corresponding to at least a portion of the text input, the audio speech output comprising at least one portion carrying contrastive stress corresponding to the at least one first portion of the at least one token of the text input, to contrast with at least one other portion of the audio speech output corresponding to the at least one corresponding first portion of the at least one other token of the text input; and
1[e]	providing the audio speech output for the speech-enabled application.
Claim 2	
2[pre]	The method of claim 1, wherein the identifying comprises:

2[a]	identifying a plurality of tokens of the text input of a same text normalization type for which a contrastive stress pattern is to be applied; and
2[b]	identifying the at least one token to be rendered with contrastive stress from among the plurality of tokens of the same text normalization type.
Claim 3	
3	The method of claim 2, wherein the same text normalization type is a text normalization type selected from the group consisting of: an alphanumeric sequence type, an address type, a Boolean value type, a currency type, a date type, a digit sequence type, a fractional number type, a proper name type, a number type, an ordinal number type, a telephone number type, a flight number type, a state name type, a street name type, a street number type, a time type and a zipcode type.
Claim 4	
4	The method of claim 2, wherein the plurality of tokens are identified based at least in part on at least one indication in the text input that the contrastive stress pattern is desired in association with the plurality of tokens.
Claim 6	
6[pre]	The method of claim 2, wherein identifying the plurality of tokens comprises:
6[a]	tokenizing the text input;
6[b]	automatically identifying the text normalization type of the plurality of tokens; and
6[c]	automatically determining that the contrastive stress pattern is to be applied for the plurality of tokens.

Claim 7	
7	The method of claim 2, wherein the at least one token to be rendered with contrastive stress is identified based at least in part on an order of the plurality of tokens in the text input.
Claim 8	
8[pre]	Apparatus for providing speech output for a speech-enabled application, the apparatus comprising:
8[a]	a memory storing a plurality of processor-executable instructions; and
8[b]	at least one processor, operatively coupled to the memory, that executes the instructions to:
8[c]	receive from the speech-enabled application a text input comprising a text transcription of a desired speech output;
8[d]	identify at least one first portion of at least one token of the text input that differs from at least one corresponding first portion of at least one other token of the text input, and at least one second portion of the at least one token that does not differ from at least one corresponding second portion of the at least one other token;
8[e]	assign contrastive stress to the identified at least one first portion of the at least one token, but not to the identified at least one second portion of the at least one token;
8[f]	generate an audio speech output corresponding to at least a portion of the text input, the audio speech output comprising at least one portion carrying contrastive stress corresponding to the at least one first portion of the at least one token of the text input, to contrast with at least one other portion of the audio speech output corresponding to the at least one corresponding first portion of the at least one other token of the text input; and

8[g]	provide the audio speech output for the speech-enabled application.
Claim 9	
9[pre]	The apparatus of claim 8, wherein the at least one processor executes the instructions to identify the at least one token at least in part by:
9[a]	identifying a plurality of tokens of the text input of a same text normalization type for which a contrastive stress pattern is to be applied; and
9[b]	identifying the at least one token to be rendered with contrastive stress from among the plurality of tokens of the same text normalization type.
Claim 10	
10	The apparatus of claim 9, wherein the same text normalization type is a text normalization type selected from the group consisting of: an alphanumeric sequence type, an address type, a Boolean value type, a currency type, a date type, a digit sequence type, a fractional number type, a proper name type, a number type, an ordinal number type, a telephone number type, a flight number type, a state name type, a street name type, a street number type, a time type and a zipcode type.
Claim 11	
11	The apparatus of claim 9, wherein the at least one processor executes the instructions to identify the plurality of tokens based at least in part on at least one indication in the text input that the contrastive stress pattern is desired in association with the plurality of tokens.
Claim 13	

13[pre]	The apparatus of claim 9, wherein the at least one processor executes the instructions to identify the plurality of tokens at least in part by:
13[a]	tokenizing the text input;
13[b]	automatically identifying the text normalization type of the plurality of tokens; and
13[c]	automatically determining that the contrastive stress pattern is to be applied for the plurality of tokens.
Claim 14	
14[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 providing speech output for a speech-enabled application, the method comprising:
14[a]	receiving from the speech-enabled application a text input comprising a text transcription of a desired speech output;
14[b]	identifying at least one first portion of at least one token of the text input that differs from at least one corresponding first portion of at least one other token of the text input, and at least one second portion of the at least one token that does not differ from at least one corresponding second portion of the at least one other token;
14[c]	assigning contrastive stress to the identified at least one first portion of the at least one token, but not to the identified at least one second portion of the at least one token;
14[d]	generating an audio speech output corresponding to at least a portion of the text input, the audio speech output comprising at least one portion carrying contrastive stress corresponding to the at least one first portion of the at least one token of the text input, to contrast with at least one other portion of the audio speech

	output corresponding to the at least one corresponding first portion of the at least one other token of the text input; and
14[e]	providing the audio speech output for the speech-enabled application.
Claim 15	
15[pre]	The at least one non-transitory computer-readable storage medium of claim 14, wherein the identifying comprises:
15[a]	identifying a plurality of tokens of the text input of a same text normalization type for which a contrastive stress pattern is to be applied; and
15[b]	identifying the at least one token to be rendered with contrastive stress from among the plurality of tokens of the same text normalization type.
Claim 16	
16	The at least one non-transitory computer-readable storage medium of claim 15, wherein the same text normalization type is a text normalization type selected from the group consisting of: an alphanumeric sequence type, an address type, a Boolean value type, a currency type, a date type, a digit sequence type, a fractional number type, a proper name type, a number type, an ordinal number type, a telephone number type, a flight number type, a state name type, a street name type, a street number type, a time type and a zipcode type.
Claim 17	
17	The at least one non-transitory computer-readable storage medium of claim 15, wherein the plurality of tokens are identified based at least in part on at least one indication in the text input that the contrastive stress pattern is desired in association with the plurality of tokens.

Claim 19	
19[pre]	The at least one non-transitory computer-readable storage medium of claim 15, wherein identifying the plurality of tokens comprises:
19[a]	tokenizing the text input;
19[b]	automatically identifying the text normalization type of the plurality of tokens; and
19[c]	automatically determining that the contrastive stress pattern is to be applied for the plurality of tokens.
Claim 20	
20	The at least one non-transitory computer-readable storage medium of claim 15, wherein the at least one token to be rendered with contrastive stress is identified based at least in part on an order of the plurality of tokens in the text input.

I. INTRODUCTION

Petitioners Samsung Electronics Co., Ltd. and Samsung Electronics America, Inc. (collectively, “Petitioners” or “Samsung”) requests *inter partes* review of claims 1-4, 6-11, 13-17, and 19-20 of U.S. Patent No. 8,914,291 (“the ’291 patent”).

II. GROUNDS FOR STANDING (37 C.F.R. §42.104(A))

Petitioners certify that the ’291 patent is available for *inter partes* review and that Petitioners are not barred or estopped from requesting *inter partes* review to challenge the claims on the grounds herein.

III. STATEMENT OF PRECISE RELIEF REQUESTED FOR EACH CLAIM CHALLENGED

For purposes of this proceeding only, Petitioners apply the earliest alleged priority date of the ’291 patent: February 12, 2010.¹ Pre-AIA 35 U.S.C. §102 applies.

¹ The claims of the ’291 patent are not entitled to claim priority to the February 12, 2010 filing date of U.S. Patent 8,949,128 (*see* Ex. 1022) because the specification of the ’128 patent lacks, *inter alia*, written description support for at least limitations 1[b], 1[c], 8[d], 8[e], 14[b] and 14[c]. At best, the claims of the ’291 patent are entitled to a priority date no earlier than August 9, 2010 (i.e., the filing date of U.S. Patent 8,571,870). All of the prior art relied on in this Petition predate February 12,

Petitioners respectfully request review and cancellation under 35 U.S.C. §311 of claims 1-4, 6-11, 13-17, and 19-20 (“Challenged Claims”) in combination with:²

Reference	Date(s)	Prior Art At Least Under
US20010049602 (“ <i>Walker</i> ”) (Ex-1005)	12/06/2001 (published)	§102(b)
US20070233492 (“ <i>Matsumoto</i> ”) (Ex-1006)	10/04/2007 (published)	§102(b)
US5634084 (“ <i>Malsheen</i> ”) (Ex-1007)	05/27/1997 (issued)	§102(b)
US7,313,523 (“ <i>Bellegarda</i> ”) (Ex-1014)	12/25/2007 (issued)	§102(b)

Ground	Claims	Basis
1A	1-4, 6-11, 13-17, and 19-20	§103: <i>Walker, Matsumoto, Malsheen</i>
1B	1-4, 6-11, 13-17, and 19-20	§103: <i>Walker, Matsumoto, Malsheen, Bellegarda</i>

2010 by more than one year. Accordingly, the Board need not resolve the priority date issue when examining the grounds in this Petition.

² Petitioners do not concede that any Challenged Claims satisfy other requirements for patentability that cannot be raised in IPR, including 35 U.S.C. §§101 and 112.

IV. OVERVIEW OF THE '291 PATENT

The '291 patent relates to “[t]echniques for generating synthetic speech with contrastive stress.” *'291Pat*, Abstract. The '291 patent aims to improve text-to-speech technology by adding stress where a human speaker would naturally apply stress in order to help the listener understand the speaker. *See id.*, 8:20-23, 8:32-48. For example, in an utterance such as “not nineteen, but twenty” the speaker may “emphasize[] the number ‘twenty’ to distinguish and contrast it from the other number ‘nineteen.’” *Id.*, 12:38-44.

The '291 patent achieves this effect by first receiving “a text transcription of a desired speech output.” *Id.*, 4:45-46. A synthesis system “analyze[s] the text input to determine which portion(s) of the speech output should carry contrastive stress.” *Id.*, 9:5-8. Then, the synthesis system “may automatically identify which tokens (e.g., words) should contrast with each other.” *Id.*, 9:13-15. “[A]fter 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[.]” *Id.*, 9:23-27.

Next, the system generates “an audio speech output corresponding to at least a portion of the text input” that contains “at least one portion carrying contrastive

stress to contrast with at least one other portion of the audio speech output.” *Id.*, 4:46-50.

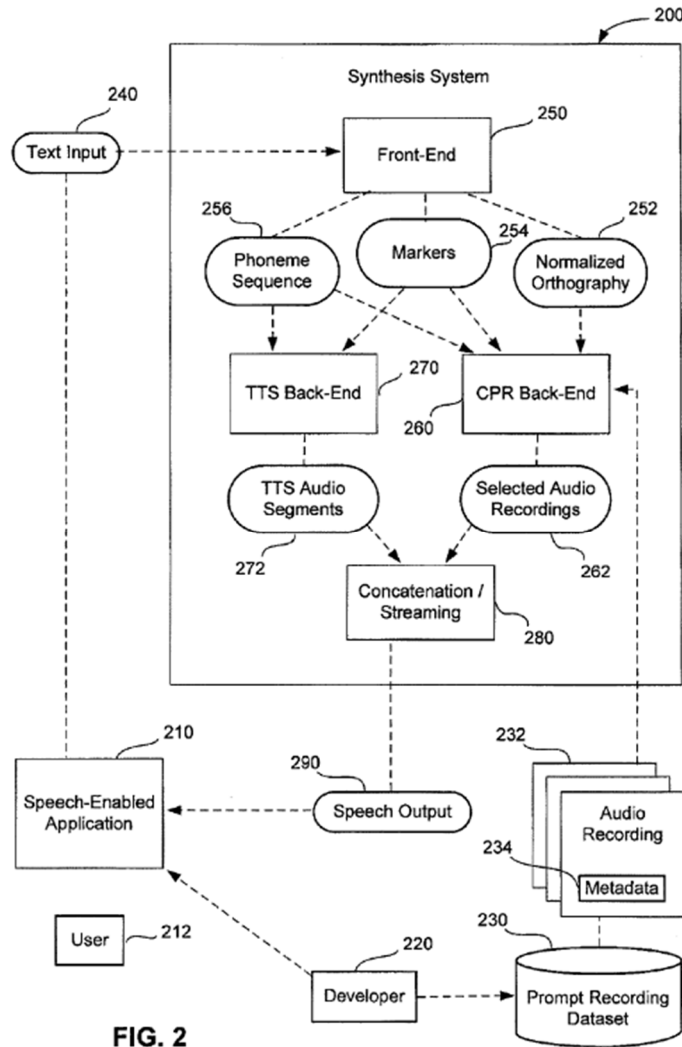
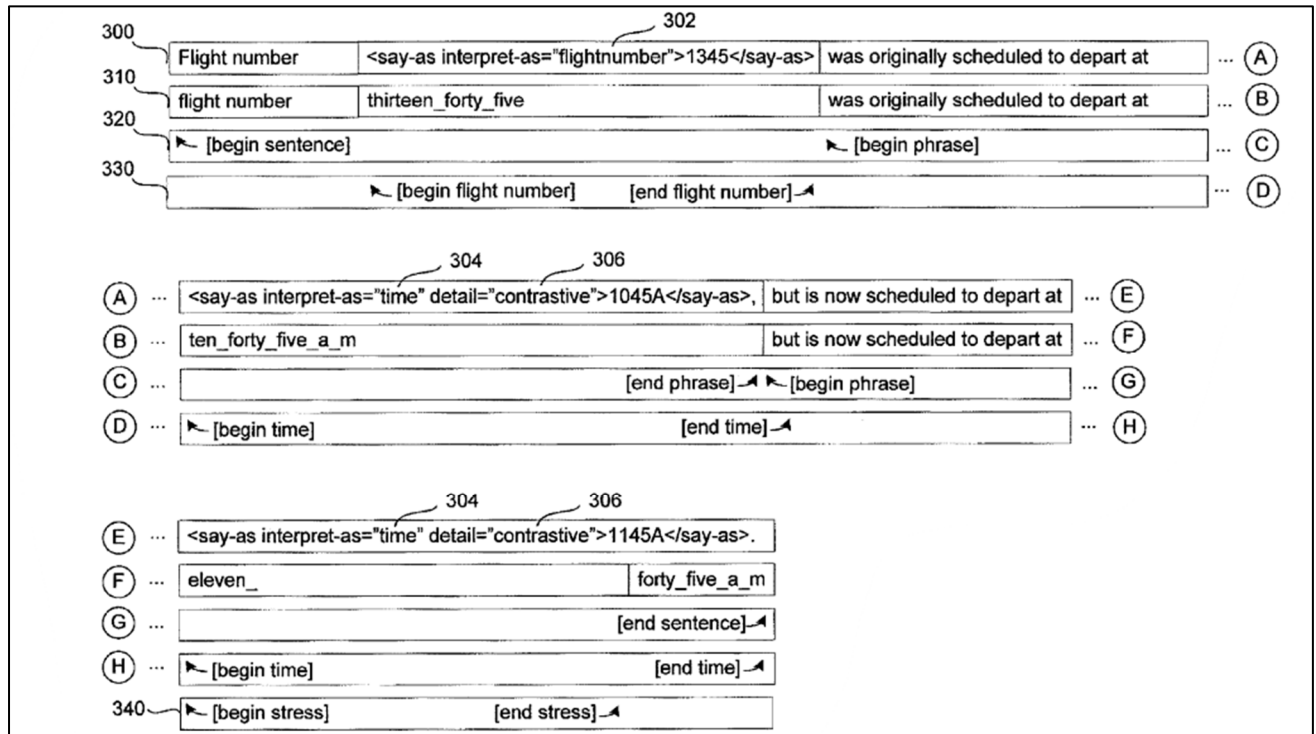


FIG. 2

'291Pat, FIG. 2.

Using the example of figure 3A below, “the [begin stress] and [end stress] markers 340 delineate the word ‘eleven’ within the time ‘11:45 a.m.’ as the specific portion of the speech output that should carry contrastive stress. 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).” *’291Pat*, 22:64-23:5.



’291Pat, FIG. 3A.

The *’291* patent further discloses that the identification of portions of text that should contrast may be identified through “normalized orthography” which encompasses converting text into full word forms based on the context of the text. *See ’291Pat*, 15:3-14. Using the example of figure 3A, “10:45 a.m.” and “11:45 a.m.” are both flight times, and therefore, “are of the same text normalization type

(e.g., ‘time’).” *Id.*, 24:2-15. Identification of the text normalization type aids synthesis system 200 in “identifying portions of a text input that are meant to contrast with each other[.]” *Id.*, 24:16-20; *see also id.*, 16:20-42.

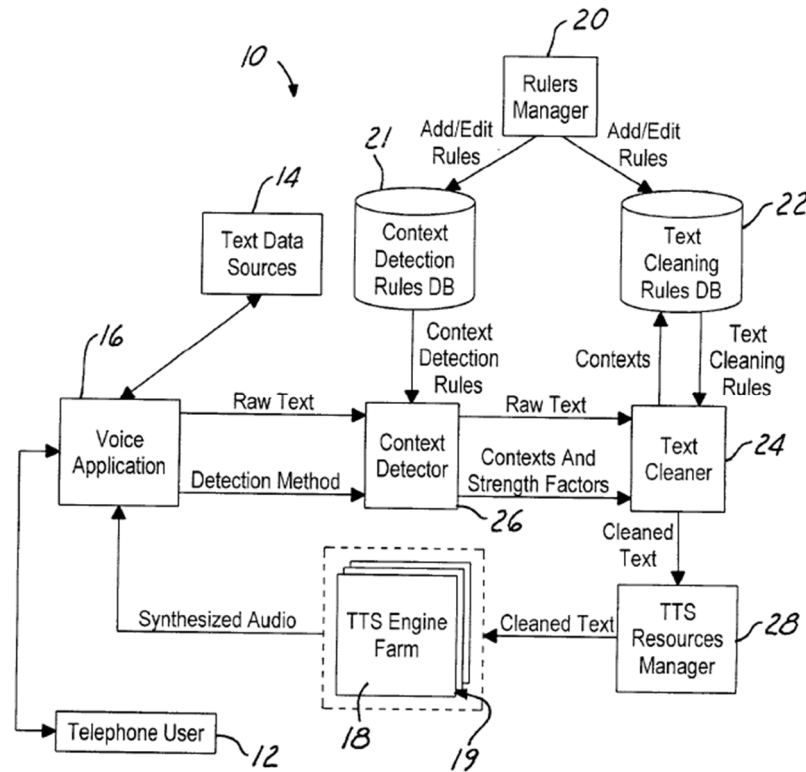
V. OVERVIEW OF THE PRIMARY PRIOR ART REFERENCES

The prior art references analyzed in the grounds below are analogous to the ’291 patent because they are within the same field of endeavor as the ’291 patent and are reasonably pertinent to one or more problems addressed by the ’291 patent. *Schmandt* ¶¶43-47.

A. *Walker* (Ex-1005)

Walker relates to “a text-to-speech method and system which convert[s] written text into audible speech as a function of the context of the text.” *Walker* ¶[0002]. *Walker* explains that synthesized speech from TTS engines often results in unnatural speech delivery due to the different contexts of words. *See id.*, ¶¶[0003]-[0005]. Thus, “without knowing the context of the document” the TTS engine may incorrectly convert text into speech. *Id.* In combination with these issues, *Walker* discloses a communication system that “convert[s] written text into audible speech as a function of the context of the text” (*id.*, ¶[0006]) to “enhance the quality of a TTS conversion of the text into speech” (*id.*, ¶[0007]).

As shown in figure 1, *Walker* discloses “communication system 10” which is “a voice portal platform for enabling a telephone user 12 to access written text ... from text data sources 14.” *Id.*, ¶[0023], FIG. 1. “Text data sources 14 may include ... *weather condition providers.*”³ *Id.*, ¶[0023].



Walker, FIG. 1.

Walker discloses that “[i]n response to a request for text or information from telephone user 12, communication system 10 locates and converts the requested text

³ All **bold/italics/color** emphases and annotations added unless noted otherwise.

into speech and then provides the speech to the telephone user via a voice application 16.” *Id.*, ¶[0023]. For example, a “telephone user 12 may request a weather report for a particular city” and, in response, “voice application 16 accesses text data sources 14 to find a text document having the weather report for the particular city” and then “receives an electronic copy of the weather report text from a text data source 14.” *Id.*, ¶[0024].

Walker further discloses preprocessing components “to determine the context of the text, i.e., context detection, and then modify[ing] the text based on the context of the text, i.e., text cleaning.” *Id.*, ¶[0025]. Specifically, *Walker*’s preprocessing includes a context detector 26 that uses “context detection rules to search the text for key words and phrases associated with each context detection rule set in order to determine the context of the text” (*id.*, ¶[0027]), and a text cleaner (or text normalizer) 24 that modifies “specific words, phrases, abbreviations, acronyms, and pronunciation in the text in accordance with a text cleaning rule set to modify the text so that the modified text sounds natural when converted into speech” (*id.*, ¶[0028]).

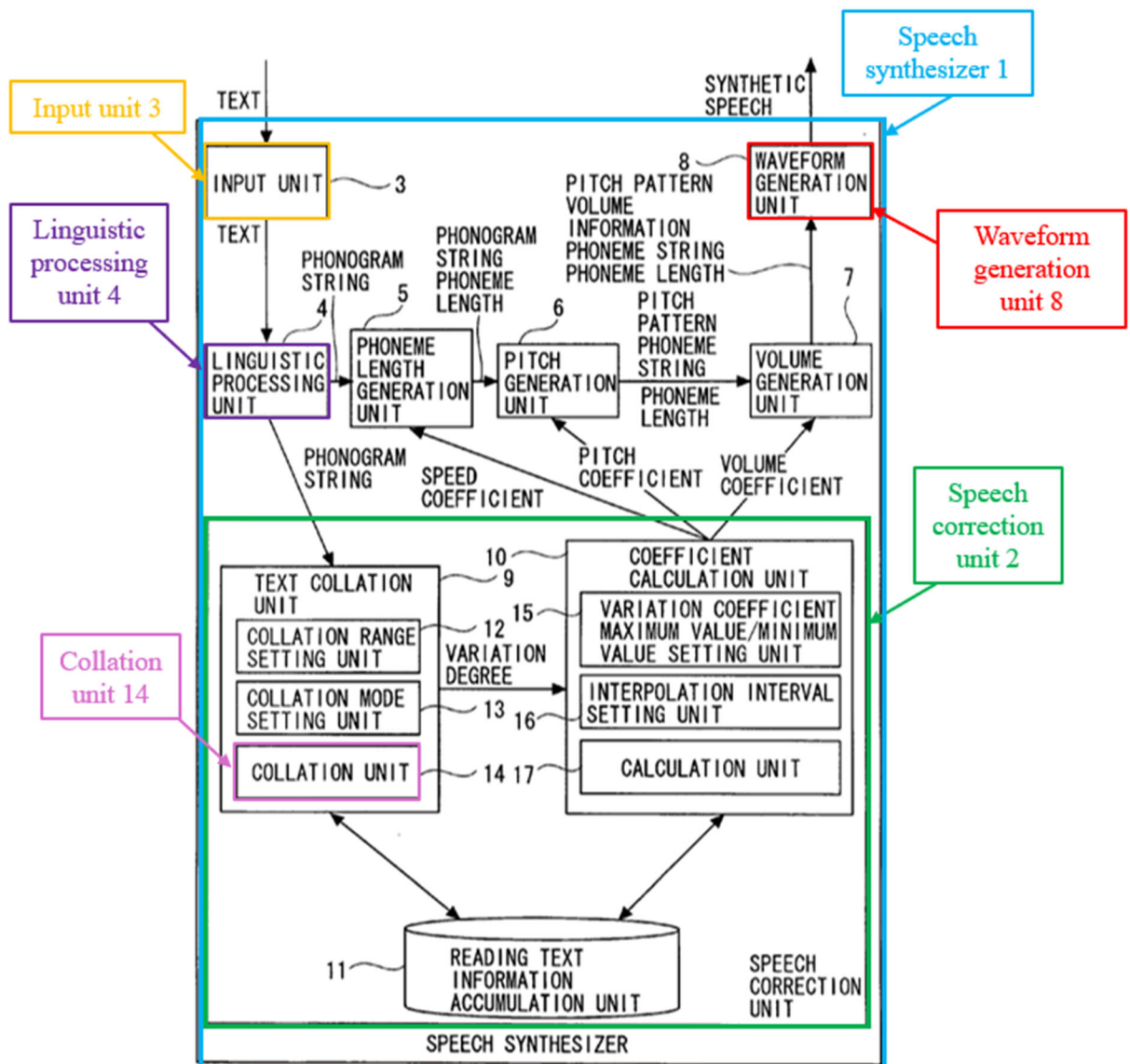
After this preprocessing, cleaned text is provided to a “TTS resource manager 28” that “distributes the cleaned text to the appropriate TTS engine 18” in TTS engine farm 19 to synthesize text into speech. *Id.*, ¶[0034]. For example, *Walker*

discloses that “a TTS engine 18 of a TTS engine farm 19 in communication system 10 converts or synthesizes the weather report text from voice application 16 into computer generated audio speech” and then “provides the audio speech of the weather report text to telephone user 12 via voice application 16.” *Id.*, ¶[0024].

B. *Matsumoto* (Ex-1006)

Matsumoto relates to “providing a hearer with [] easy-to-hear synthetic speech.” *Matsumoto*, Abstract. *Matsumoto* explains that in certain instances such as a “weather forecast,” speech synthesizers receive “almost the same sentences[.]” *Id.*, ¶[0005] (“[Today’s weather in the Hokkaido region is fair.] (kyou no Hokkaido chihou no tenki wa hare desu.), [Today’s weather in the Tohoku region is fair.] ([kyou no Tohoku chihou no tenki wa hare desu.])”). However, if “the speech synthesizer vocalizes such sentences in a monotone” the hearer may fail to hear the “want-to-hear point.” *Id.* Accordingly, *Matsumoto*’s objective is to provide “the hearer with the easy-to-hear synthetic speech” (*id.*, ¶[0015]), that exhibits high naturalness by controlling prosodic changes in the sentence (*see id.*, ¶[0122]). Prosodic changes include variations in speed, pitch, and volume when words/phrases are synthesized into speech. *See id.*, ¶¶[0004], [0040]-[0042], [0103]-[0105]; *Schmandt* ¶60.

As shown in figure 1, *Matsumoto* discloses a “speech synthesizer 1” that “includes a speech correction unit 2, an input unit 3, a linguistic processing unit 4 ... and a waveform generation unit 8.” *Id.*, ¶[0036].



Matsumoto, FIG. 1 (annotated).

“[I]nput unit 3 accepts text data of a sentence” (*id.*, ¶[0037]) and “linguistic processing unit 4 executes a morphological analysis about the text (sentence) and segments this text (sentence) into morphemes (the minimum unit having a meaning in language)” (*id.*, ¶[0038]). After the text is segmented, “speech correction unit 2 accumulates the phonogram strings (sentences) acquired from the text data inputted by the input unit 3, then obtains a variation degree” between the accumulated sentences. *Id.*, ¶[0044]. The variation degree is calculated by collation unit 14 (*see id.*, ¶¶[0048],[0077]) and describes how each word or phrase in a sentence differs from a corresponding word or phrase in a previous sentence. *See id.*, ¶¶[0044],[0074].

In particular, *Matsumoto* discloses that each sentence is assigned a sequential value “n” (with n=1 being the oldest sentence) and each word or phrase is assigned a sequential value “b.” *See id.*, ¶¶[0072]-[0075]. Accordingly, each word or phrase is described by the coordinates “a(n, b).” *See id.*, ¶¶[0074]-[0075], FIG. 6.

	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

Matsumoto, FIG. 6.

The collation unit 14 then “compares ... two sets of a(n, b) having the same value of the variable b and different values of the variable n.” *Id.*, ¶[0074]. For example, as shown above, “in the process of the collation between a(1, 1) ([today's]) ([kyou no]) and a(2, 1) ([today's]) ([kyou no]), [the collation unit 14] judges that contents of the phrases are the same [and] in the collation between a(1, 2) ([Tohoku]) and a(2, 2) ([Kanto]), judges that the contents of the phrases are different.” *Id.*

The calculated variation degree obtained by collation unit 14 is used to “calculate[] the variation coefficients (the speed coefficient, the pitch coefficient and the volume coefficient)” between each corresponding word or phrase. *Id.*, ¶[0054]. The variation coefficients are then used to adjust the speed, pitch, and volume for each word or phrase. *See id.*, ¶¶[0103]-[0105].

Matsumoto further discloses that “waveform generation unit 8” generates synthetic speech “from the phoneme string, the phoneme length, the pitch pattern

and the volume information” based on the variation coefficients and outputs synthetic speech to “the voice output device (not shown) such as the speaker connected to the speech synthesizer 1.” *Id.*, ¶[0107].

VI. LEVEL OF ORDINARY SKILL IN THE ART

A person of ordinary skill in the art (“POSITA”) at the time of the alleged invention of the ’291 patent,⁴ 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. A greater amount of education could compensate for less work experience, and vice versa. *Schmandt* ¶23.

VII. CLAIM CONSTRUCTION

Petitioners are unaware of any “prior claim construction determination” related to the ’291 patent. *See* 37 C.F.R. §42.100(b). Given the close correlation and substantial identity between the prior art references and the Challenged Claims, that no express constructions of the claims are necessary to assess whether the prior art

⁴ The POSITA analysis would not change regardless of which of the two priority dates discussed in note 1 applies. *Schmandt* ¶23.

reads on the Challenged Claims. Thus, the claims should be given their plain and ordinary meaning.⁵ *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc).

VIII. DETAILED EXPLANATION OF GROUNDS

A. Ground 1A: *Walker* in Combination With *Matsumoto* And *Malsheen* Renders Obvious Claims 1-4, 6-11, 13-17, and 19-20

1. Independent Claim 1

a. 1[pre]

To the extent the preamble is limiting, *Walker* teaches this feature.⁶ *Schmandt* ¶¶73-79.

⁵ Petitioners reserve all rights to raise claim construction arguments and other arguments in any parallel or future litigation concerning the '291 patent. For example, comparing the claims to the accused products in the litigation may raise controversies that require construction of certain claim terms.

⁶ Petitioners use the term “teaches” as including both express teachings or those fairly suggested to a POSITA. *In re Baird*, 16 F.3d 380, 383 (Fed. Cir. 1994); *In re Keller*, 642 F.2d 413, 425 (CCPA, 1981) (“The test for obviousness is ... what the combined teachings of the references would have suggested to those of ordinary skill in the art.” (citations omitted)).

Walker discloses a method in which a telephone user 12 requests information and communication system 10 locates and converts the requested text into speech which is provided to the user via voice application 16. *See Walker* ¶¶[0018], [0023], FIG. 1. An example of a text to be converted to speech provided in *Walker* is a weather report. *See id.*, ¶[0023]. A TTS engine 18 within *Walker*'s system converts the text into synthesized speech and provides the audio speech of the text weather report to telephone user 12 via voice application 16. *See id.*, ¶¶[0023]-[0024], [0034]. *Walker*'s voice application 16 is consistent with how the '291 patent describes a "speech enabled application." *See '291Pat*, 1:26-28 (describing a "speech-enabled application" as "capable of providing output to a human user in the form of speech."); *see also id.*, 1:31-53, 9:35-60; *Schmandt* ¶¶77-78.

Accordingly, *Walker* teaches a system having a voice application for performing a method of converting text into speech ("a method for providing speech output for a speech-enabled application"). *Schmandt* ¶79.

b. 1[a]

Walker in combination with *Matsumoto* teaches this feature. *Schmandt* ¶¶80-90.

Walker discloses that context detector 26 receives raw text (i.e., unprocessed text) from voice application 16. *Walker* ¶[0030] (describing the "electronic copy of

the text document provided by voice application 16 to context detector 26 is labeled ‘Raw Text’ in FIG. 1”), FIG. 1. As discussed in §VIII.A.1.a, a POSITA would have understood that voice application 16 is a speech-enabled application. Thus, a POSITA would have understood context detector 26 receives raw text from a speech-enabled application 16. *Schmandt* ¶81.

Walker further discloses that one type of text received by the context detector 26 from voice application 16 is “a weather report for a particular city.” *Id.*, ¶[0024]. While *Walker* discloses that the weather report is for a particular city, it would have been obvious to a POSITA that *Walker*’s weather report could pertain to more than one location/city or day. For example, *Matsumoto* discloses a weather forecast to include the weather for different locations and days.⁷ See *Matsumoto* ¶[0071] (describing “[t]oday’s weather” in the “Tohoku region” and “[t]omorrow’s weather” in the “Kansai region”); see also *id.*, ¶¶[0005], [0073]; *Schmandt* ¶¶82-83.

It would have been obvious to modify *Walker* based on *Matsumoto*’s disclosure such that the context detector 26 receives from the voice application 16 raw text that pertains to the weather report for more than one location/city because,

⁷ A POSITA would have understood weather reports and weather forecasts to include similar information. *Schmandt* ¶84.

as *Matsumoto* teaches, a user (e.g., a telephone user 12 interacting with voice application 16 in *Walker*'s system) may be interested in the weather from more than one location/city. *Schmandt* ¶85.

The combination would also have been obvious to a POSITA because it would have amounted to combining prior art elements (*Walker* teaches context detector 26 receiving a text weather report) according to a known method (*Matsumoto* teaches that a text weather report can pertain to more than one location/city or day) to yield predictable results (the context detector receiving a text weather report that pertains to more than one location/city or day). *Schmandt* ¶86. Further, the combination would have been obvious to a POSITA because it amounted to a simple substitution of a known element (context detector receiving a text weather report for a particular city) for another (context detector receiving a text weather report for more than one location/city or day) to obtain predictable results (to determine the context of text as pertaining to a weather report). *Schmandt* ¶87.

Thus, a POSITA would have found it obvious for *Walker*'s weather report to include two or more sentences (or text input), each describing the weather for a different location/city or day, as taught by *Matsumoto*. *Schmandt* ¶88.

As discussed in §V.A, *Walker* converts written text into synthesized speech output. A POSITA would have understood the synthesized speech output of the

Walker-Matsumoto combination to be a “desired speech output” because *Walker* discloses a user using *Walker*’s system to request that a text weather report be outputted as synthesized speech. *Walker*, ¶¶[0023]-[0024]. *Schmandt* ¶89.

Accordingly, the *Walker-Matsumoto* combination teaches a context detector receiving from a voice application (“receiving from the speech-enabled application”) the raw text of a weather report that a user wanted to hear as speech output (“a text input comprising a text transcription of a desired speech output”). *Schmandt* ¶90.

c. 1[b]

Walker in combination with *Matsumoto* and *Malsheen* teaches this feature. *Schmandt* ¶¶91-160.

In the discussion that follows, Petitioners first discuss a POSITA’s understanding of “token” as used in the claims and specification of the ’291 patent. Next Petitioners discuss why a POSITA would have found it obvious to combine the teachings of *Walker* and *Matsumoto*. Then, Petitioners discuss an example text weather report that a POSITA would have understood to be taught by the *Walker-Matsumoto* combination. Thereafter, Petitioners discuss why a POSITA would have found it obvious to further modify the *Walker-Matsumoto* combination based on *Malsheen* to further process the example text weather report prior to converting it

into synthesized audio output. Finally, Petitioners explain how the *Walker-Matsumoto-Malsheen* combination would operate on the example text weather report.

(i) “Token” as used in the ’291 claims and specification

A POSITA would have understood “token” as used in the claims and specification of the ’291 patent to encompass a word or lexical unit of text separated by white spaces. *See*, ’291*Pat*, 9:13-15 (“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.”), 31:56-61 (referring to a “tokenizer” as “pars[ing] text input 350 into a sequence of *individual tokens on the order of single words ... the individual parsed tokens are represented as separated by white space* in the normalized orthography 360.”); *see also id.*, 25:62-67 (“words/word sequences (or equivalently, tokens/token sequences)”), 26:2-5 (“one or more linking tokens such as ‘originally’, ‘but’ and ‘is now’”), 26:19-20 (“linking token ‘originally’”), 26:21 (“linking tokens ‘but is now’”), 36:23-25 (“At act 410, the text input may be tokenized, i.e., parsed into individual tokens on the order of single words.”); *Schmandt* ¶93. Indeed, this is consistent with how a POSITA would

have understood the term “token” to be used in the art. *Schmandt* ¶¶94-95 (discussing *W3C*).

A POSITA would have also understood the '291 patent to use the terms “token” and “field” interchangeably. *See, e.g., '291Pat*, 9:8-10 (“the application may include tags in the text input to identify ***tokens or fields*** that should contract with each other.”), 28:62-63 (“Identifying specific portions of one or more ***fields or tokens*** to be rendered to carry contrastive stress ...”), 27:56-60 (“In some embodiments, the entire ***token ‘1145A’*** may be identified to carry contrastive stress, while in other embodiments synthesis system 200 may next proceed to identify only one or more specific portions of the ***field ‘1145A’*** to be rendered to carry contrastive stress”); *see also id.*, 32:63-66 (referring to “10:30 a.m.,” “11:30 a.m.,” and “12:30 p.m.” as tokens), 33:16-21 (referring to “10:30 a.m.,” “11:30 a.m.,” and “12:30 p.m.” as fields); *Schmandt* ¶96.

The '291 patent explains that tokens may have a text normalization type, and that a normalized orthography corresponding to the inputted text token may be generated. *Schmandt* ¶97; *see '291Pat.*, 36:29-36 (“As discussed above, text normalization types of various tokens may be identified ... [and] a normalized orthography corresponding to the text input may be generated.”). Examples of text normalization types mentioned in the '291 patent include state names. *'291Pat*,

24:45-51. The '291 patent describes “normalized orthography” as encompassing the conversion of abbreviations into full words or word sequences. *See* '291 *Pat.*, 15:3-14 (“Front-end 250 may also be programmed to expand **abbreviations** and acronyms **into full words and/or word sequences**, and to convert numerals, symbols and other meaningful characters to word forms, using appropriate language-specific rules based on the context in which these items occur in text input 24.”), 36:37-40 (“[T]he normalized orthography may represent a standardized spelling out of the words included in the text input, which for some tokens may depend on their text normalization type.”); *Schmandt* ¶98.

(ii) Motivation to Combine *Walker* and *Matsumoto*

A POSITA would have been motivated to combine the teachings of *Walker* and *Matsumoto*, and would have had a reasonable expectation of success in doing so. *Schmandt* ¶¶99-115.

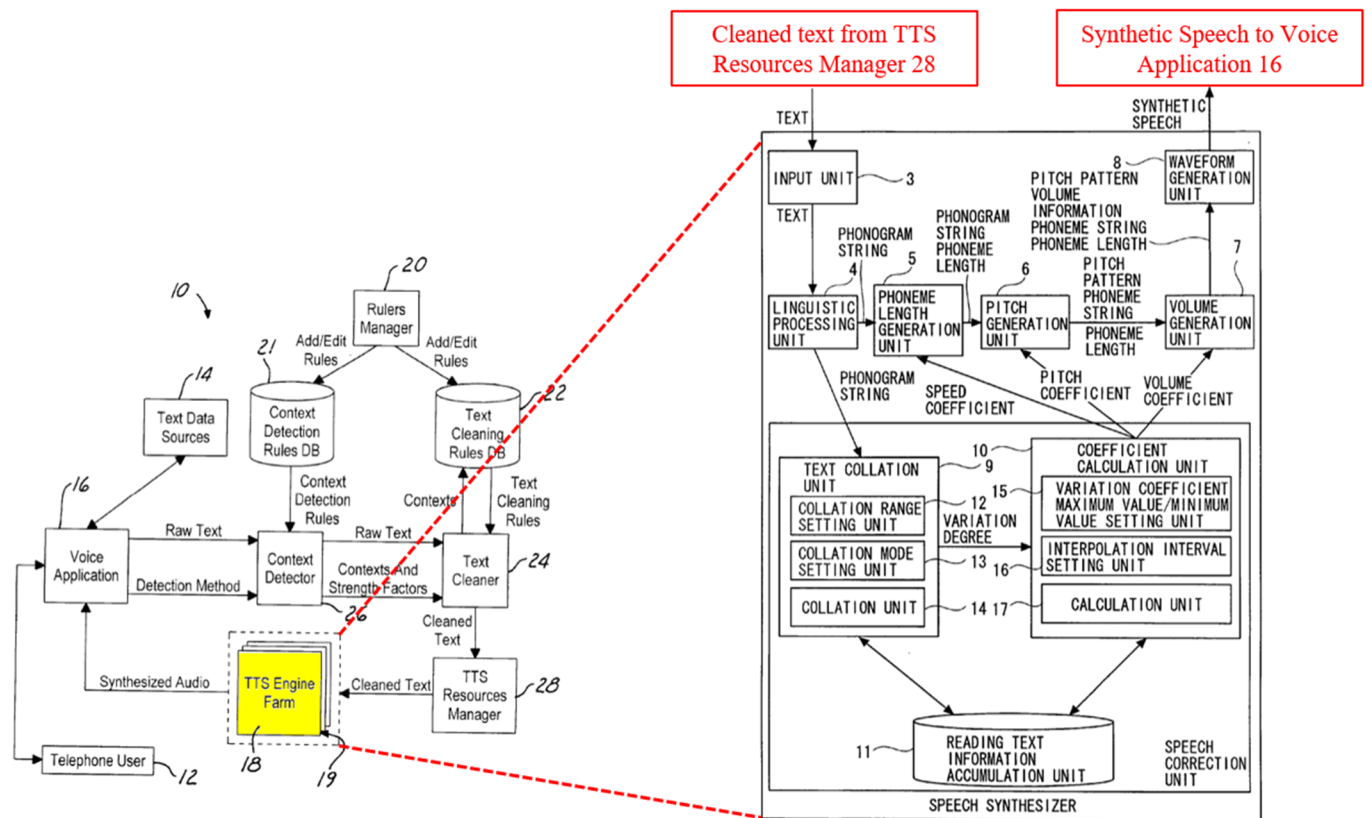
Walker teaches a system and method for improving the naturalness of synthesized speech using the text’s context. *See* §V.A. *Walker* teaches obtaining more natural and higher quality speech synthesis by using a context detector and text cleaner to preprocess raw text received by a voice application. *See id.*, ¶¶[0030]-[0033]; *see also id.*, ¶¶[0011]-[0015], [0034]-[0038]. *Walker* discloses that one type of text that is pre-processed is “a weather report for a particular city” (*id.*, ¶[0024])

which is then provided to a TTS resource manager 28 that “distributes the cleaned text to the appropriate TTS engine 18” in TTS engine farm 19 to synthesize text into speech (*id.*, ¶[0034]). *Walker*, however, does not provide extensive implementation details regarding the TTS engines 18 within the TTS engine farm 19, including as they would relate to TTS engines for synthesizing weather reports (or similar text documents) or how such a TTS engine would improve the naturalness of synthesized speech. *Schmandt* ¶102; *See also Walker* ¶[0034]. Thus, a POSITA would have been motivated to find implementation details for a TTS engine capable of synthesizing text weather reports into speech with high naturalness to implement *Walker*’s teachings. *Schmandt* ¶103. In so looking, a POSITA would have readily found *Matsumoto*. *Schmandt* ¶103.

Matsumoto discloses a speech synthesizer that aims to provide easy-to-hear synthesized speech that exhibits high naturalness, for example, in a weather forecast. *See Matsumoto* ¶¶[0121]-[0122]; *see also id.*, ¶¶[0005], [0014], [0015], [0121]. It was well-known that speech synthesizers like *Matsumoto*’s were deployed in connection with speech applications, such as *Walker*’s voice application. *Schmandt* ¶105. Indeed, *Walker* itself discloses that its system uses a TTS engine (i.e., a speech synthesizer) working in conjunction with a voice application to convert text to

synthesized speech. See *Walker* ¶[0024]; *Schmandt* ¶¶105-108 (discussing *Lewis* and *Schmid*).

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. *Schmandt* ¶¶109-111. Thus, a POSITA would have found it obvious to incorporate *Matsumoto*'s teachings regarding its speech synthesizer into *Walker* as one of *Walker*'s TTS engines 18 in TTS farm 19 for converting text weather reports into more naturally flowing, easy-to-hear synthesized speech, as conceptually illustrated in the figure below. *Schmandt* ¶112.



The combination would also have been obvious to a POSITA because it 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). *Schmandt* ¶113.

A POSITA would have had a reasonable expectation of success in making the modification because *Matsumoto* discloses additional details for a TTS engine that improves the naturalness of synthesized speech that would have been within a

POSITA's ability to implement in *Walker*. *Schmandt* ¶114. For example, both *Walker* and *Matsumoto* synthesize speech from documents. See *Walker*, ¶¶[0004]-[0005], [0024], [0030]; *Matsumoto*, ¶¶[0005], [0037]-[0038]. Additionally, *Walker* already discloses using a speech synthesizer that receives text input and outputs synthetic speech, and it would have been within a POSITA's ability to program *Matsumoto*'s speech synthesizer as one of *Walker*'s TTS engines, because *Walker* produces text which is sent to the TTS engines, and *Matsumoto* takes text as its input. *Schmandt* ¶115.

(iii) Example text weather report that a POSITA would have understood to be taught by the *Walker-Matsumoto* combination

For the reasons discussed in §VIII.A.1.b, the *Walker-Matsumoto* combination teaches a text weather report comprising two or more sentences describing the weather for two or more locations/cities. *Matsumoto*'s figure 6 provides an example set of such sentences.

	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

Matsumoto, FIG. 6.

While *Matsumoto*'s examples refer to a location by only a single name (e.g., Tohoku or Kanto), a POSITA would have understood that a weather report can refer to a location by city and state names, and that state names are often written using the abbreviation for that state's name, for example: The weather in Greenville, NC is fair. *Schmandt* ¶117.

While *Matsumoto*'s figure 6 shows five sentences (i.e., t(n=1) through t(n=5)), a POSITA would have understood *Matsumoto*'s system to be capable of operating on two or more sentences. *Schmandt* ¶118. For example, the calculation of a variation degree of a given word is based on a comparison to the word in the same "b" position in the prior sentence. See *Matsumoto* ¶¶[0074], [0077]-[0079], Expression 1; *Schmandt* ¶¶119-120. *Matsumoto* further discloses that the collation can be based on the number of sentences (e.g., two sentences), including comparing

the latest sentence with the just-anterior sentence. *See Matsumoto* ¶¶[0058]-[0059], [0061]-[0062], FIGs. 2-3; *Schmandt* ¶¶121-125.

Accordingly, the *Walker-Matsumoto* combination teaches producing a synthetic speech output for a text weather report comprised of two sentences in which the location information includes an abbreviation representing a state name, such as the example weather report illustrated below:

The weather in Greenville, NC is fair.

The weather in Greenville, SC is fair.

Schmandt ¶126.

(iv) Motivation to combine *Walker-Matsumoto* with *Malsheen*

A POSITA would have been motivated to further modify the *Walker-Matsumoto* combination in view of *Malsheen*'s disclosure and would have had a reasonable expectation of success in doing so. *Schmandt* ¶¶127-139.

As detailed in the preceding section, a POSITA would have understood the *Walker-Matsumoto* combination to teach producing a synthetic speech output for a text weather report comprised of the following two example sentences:

The weather in Greenville, NC is fair.

The weather in Greenville, SC is fair.

For the reasons discussed in §VIII.A.1.c(i), (“token” as used in the ’291 claims and specification) a person of ordinary skill in the art would have understood each word of the example weather report above—including the abbreviations “NC” and “SC”—to comprise a “token” as that term is used in the ’291 patent claims and specification.

While *Walker* provides examples of expanding abbreviations into full words (see, *Walker* ¶[0033]), it does not discuss in great detail on how the context cleaner should modify or change abbreviations representing a location (e.g., representing a state name) before the text is sent to the appropriate TTS engine. *Schmandt* ¶130.

Malsheen discloses a text-to-speech converter for improving the handling of text that includes abbreviations. *Malsheen*, 1:5-10. For example, *Malsheen* discloses a text classifier procedure that “determines and classifies the syntax aspects of each word” including making determinations as to whether a word is a member of a particular class (e.g., geographic names). *Id.*, 5:41-50; see also 6:20-39. The text classifier procedure “calls the Text Expander procedure 140” (*id.*, 7:20-22), which “determine[es] what expansion of the word is required” (*id.*, 7:10-12). Additional text processing may be performed by the “Text Expander procedure 140” to expand state abbreviations into a full words. See *id.*, 8:31-32 (“the ‘state’ rule indicates that ‘mo’ should be expanded to ‘Missouri’”); see also *id.*, 9:19-10:62 (explaining how

the text “S F, CA” would be expanded to “San Francisco California.”). A POSITA would have understood *Malsheen*’s reference to “expand[ing]” an abbreviation, such as a state name, to “a full word string” to mean that the abbreviation of a state is written in a full word form. *Schmandt* ¶134.

A POSITA would have been motivated to incorporate *Malsheen*’s teaching that abbreviations representing a location should be converted to written words into the *Walker-Matsumoto* combination such that abbreviations appearing as a combination of letters in the text weather report are converted to full text by the text cleaner 24 to ensure proper pronunciation by the speech synthesizer (i.e., “South Carolina” instead of “s c”). *Schmandt* ¶135.

A POSITA would have found the combination obvious because it amounted to the use of a known technique (*Malsheen*’s teaching to expand abbreviations into full words before sending text to a TTS engine) to improve a similar system (*Walker-Matsumoto* teaches converting text weather reports incorporating abbreviations representing locations to synthesized speech) in the same way (abbreviations representing locations in the *Walker-Matsumoto* text weather reports are converted to full words as taught by *Malsheen* before the weather report is sent to a TTS engine). *Schmandt* ¶136. Additionally, a POSITA would have had a reasonable expectation of success in making the modification because *Malsheen* discloses

additional details for text normalization implemented by *Walker*'s context cleaner that would have been within a POSITA's ability to implement. *Schmandt* ¶137.

Thus, a POSITA would have understood that in the *Walker-Matsumoto-Malsheen* combination, the text cleaner would convert the abbreviation "NC" to "North Carolina" and the abbreviation "SC" to "South Carolina," such that the example text weather report sent to the TTS engine would comprise:

The weather in Greenville, North Carolina is fair.

The weather in Greenville, South Carolina is fair.

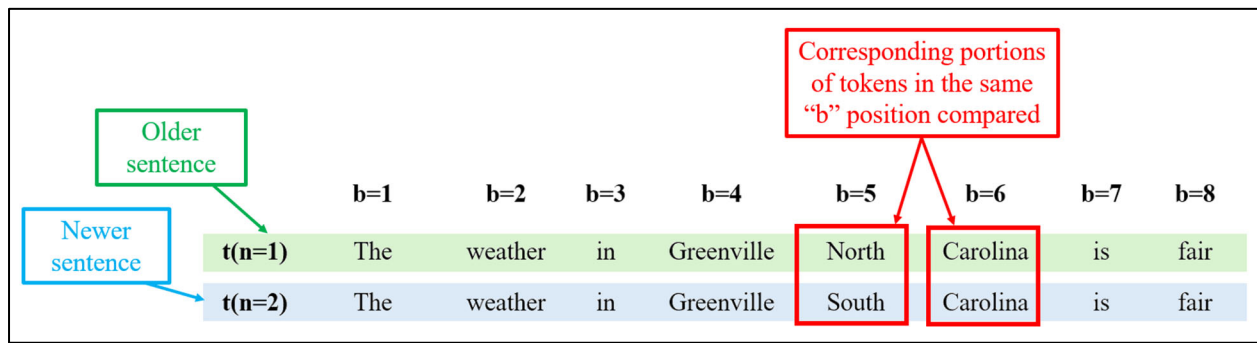
Schmandt ¶138. A POSITA would have further understood that this process of expanding abbreviations of state names to full words is analogous to what the '291 patent refers to as "normalized orthography" as discussed in §VIII.A.1.c(i). *Schmandt* ¶139.

(v) Application of the *Walker-Matsumoto-Malsheen* combination to the example text weather report

In the discussion that follows, it is assumed that the weather report received by *Walker*'s voice application 16 from text data source 14 is the example weather report provided in §VIII.A.1.c(iii) and that *Matsumoto*'s TTS system is configured to perform collation based on the number of sentences (i.e., two sentences) and to collate the latest sentence with only the anterior sentence.

As discussed in §VIII.A.1.c(ii), a POSITA would have found it obvious to incorporate *Matsumoto*'s TTS system into *Walker*'s TTS engine farm as the TTS engine for synthesizing text weather reports into speech. Further, a POSITA would have understood that *Walker-Matsumoto-Malsheen*'s speech synthesizer would operate on the “cleaned” sentences of the example weather report, as discussed in §VIII.A.1.c(iv).

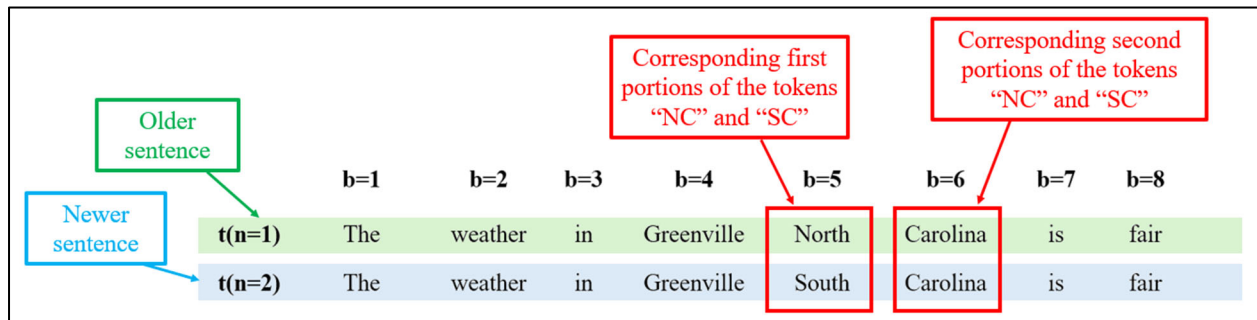
Matsumoto discloses its speech synthesizer includes collation unit 14 that calculates the variation degree of the new sentence from the previous sentence which describes how each word or phrase in a new sentence differs from a corresponding word or phrase in a previous sentence. *See Matsumoto* ¶¶[0044], [0074]; *Schmandt* ¶143. Specifically, *Matsumoto* discloses that collation unit 14 compares two sets of a(n, b). *See id.*, ¶[0074] (disclosing that a(1, 1) “kyou no” and a(2, 1) “kyou no” are compared while a(1, 2) “Tohoku” and a(2, 2) “Kanto” are compared); *see also id.*, ¶[0075]. Thus, using the example weather report discussed herein, a(1, 5) “North” and a(2, 5) (“South”) are compared while a(1, 6) “Carolina” and a(2, 6) “Carolina” are compared. *Schmandt* ¶144.



As shown, *Matsumoto*'s collation unit would identify that "North" is different from "South" and that "Carolina" is the same as "Carolina," each of which constitute a "portion" of different tokens, "NC" and "SC," respectively. *See Matsumoto*, ¶[0074] (discussing with reference to figure 6 that "in the process of the collation between a(1, 1) ([today's]) ([kyou no]) and a(2, 1) ([today's]) ([kyou no])," collation unit 14 "judges that contents of the phrases are the same" and "in the collation between a(1, 2) ([Tohoku]) and a(2, 2) ([Kanto]), judges that the contents of the phrases are different."). *Schmandt* ¶145. *Matsumoto* further discloses that when corresponding words or phrases are determined to be the same (e.g., "Carolina" and "Carolina") the variation degree is "0," but if determined to be different (e.g., "North" and "South"), the variation degree is "1." *See id.*, ¶[0080]; Expression 1; *Schmandt* ¶¶146-156 (discussing the calculation of the variation degree).

A POSITA would have understood that *Matsumoto*'s collation unit 14 compares *corresponding portions* of tokens because "South" and "North" both

appear in the b=5 position and “Carolina” and “Carolina” both appear in the b=6 position. *Schmandt* ¶157.



A POSITA would have understood that in the scenario discussed herein, “South” appearing in location a(2, 5) corresponds to “one first portion of at least one token” and “North” appearing in location a(1, 5) corresponds to “one corresponding first portion of at least one other token.” *Schmandt* ¶158. Additionally, a POSITA would have understood that “Carolina” appearing in location a(2, 6) corresponds to “one second portion of the at least one token” and “Carolina” appearing in location a(1, 6) corresponds to “one corresponding second portion of the at least one other token.” *Schmandt* ¶158. Indeed, the ’291 patent describes comparing portions (i.e., “eleven” and “ten”) of tokens (i.e., “11:45 a.m.” and “10:45 a.m.”). *See* ’291*Pat*, 28:27-31, 22:67-23:5 (“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).), 28:62-67.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches identifying a first portion (i.e., “North” and “South”) of two tokens (i.e., “NC” and “SC”) that are different and have a variation degree of “1” (“identifying at least one first portion of at least one token of the text input that differs from at least one corresponding first portion of at least one other token of the text input”) and a second portion (i.e., “Carolina” and “Carolina”) of two tokens (i.e., “NC” and “SC”) that are the same and have a variation degree of “0” (“and at least one second portion of the at least one token that does not differ from at least one corresponding second portion of the at least one other token”). *Schmandt* ¶160.

d. 1[c]

Walker in combination with *Matsumoto and Malsheen* teaches this feature. *Schmandt* ¶¶161-202.

As discussed in §VIII.A.1.c(v), in the example weather report discussed in this ground, “South” in location a(2, 5) corresponds to “one first portion of at least one token,” and “North” in location a(1, 5) corresponds to “one corresponding first portion of at least one other token.”

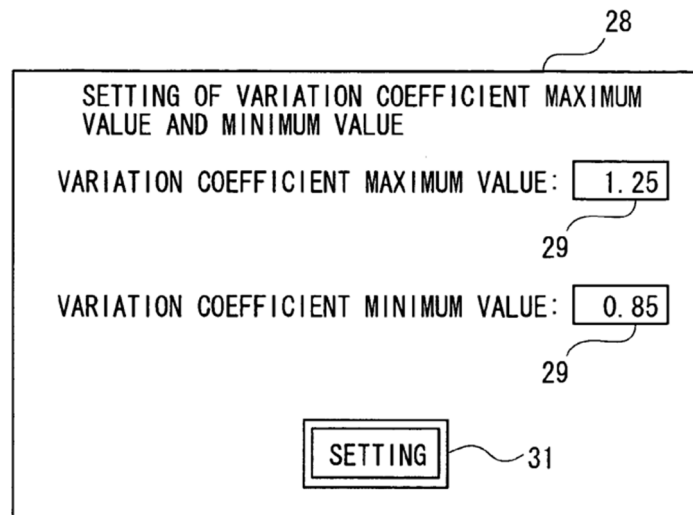
Matsumoto discloses “calculation unit 17” uses the “variation degree obtained from the collation unit 14” to “calculate[] the variation coefficients (the speed coefficient, the pitch coefficient and the volume coefficient)” between corresponding words or phrases of two or more sentences. *Matsumoto* ¶[0054]. The variation coefficients C1 (speed), C2 (pitch), and C3 (volume), are used to adjust the prosody of words/phrases in text when synthesized into speech. *See id.*, ¶¶[0103]-[0105].

A POSITA would have understood that the prosodic changes to “speed” recited in *Matsumoto* refers to changes in “duration” because it describes the speed of synthesized speech (i.e., how long it takes to speak a given word). *See Matsumoto* ¶¶[0010], [0011]-[0014], [0103]; *Schmandt* ¶¶165-170 (discussing *Bellegarda* and *Lewis*).

Calculation unit 17 assigns the variation coefficients for speed/duration (C1), pitch (C2), and volume (C3) to the phoneme length generation unit 5, the pitch generation unit 6, and the volume generation unit 7, respectively. *See id.*, ¶[0054]. The variation coefficients are then used to adjust the speed/duration, pitch, and volume for each word or phrase of the new sentence. *See id.*, ¶¶[0103]-[0105]; *see also id.*, ¶¶[0042] (“The phoneme length generation unit 5, the pitch generation unit

6 and the volume generation unit 7 make, only when assigned a variation coefficient, the correction by use of the assigned variation coefficient.”), [0088]-[0091].

Matsumoto further discloses a user can set a maximum and minimum value for the variation coefficients, which are upper and lower limits of the variation coefficients C1 (speed/duration), C2 (pitch), and C3 (volume). *See id.*, ¶¶[0064]-[0065]; *Schmandt* ¶173. *Matsumoto*’s figure 4 discloses an example in which the variation coefficients (C1, C2, C3) have a user-set maximum value of 1.25 and minimum value of 0.85. *See Matsumoto* ¶[0091].



Matsumoto, FIG. 4.

When calculation unit 17 calculates a variation coefficient that is more than the user-set maximum value, the maximum variation coefficient value is used. *See id.*, ¶[0091]. Similarly, when calculation unit 17 calculates a variation coefficient

that is less than the user-set minimum value, the minimum variation coefficient value is used. *See id.* A POSITA would have understood that in *Matsumoto*'s system, a variation coefficient above 1.0 would emphasize a word/phrase, a variation coefficient below 1.0 would deemphasize a word/phrase, and a variation coefficient of 1.0 would neither emphasize nor deemphasize a word/phrase. *Schmandt* ¶¶178-180.

As discussed in §VIII.A.1.c(v), “South” is identified as different from the corresponding word “North” and therefore, the calculated variation degree of “South” is 1.0. Thus, a POSITA would have understood that *Matsumoto*'s system calculates variation coefficients (C1, C2, C3) for “South” as above 1.0. *Schmandt* ¶¶181-188 (discussing the calculation of the volume variation coefficient for “South”). Accordingly “South” (i.e., one first portion of at least one token) is emphasized. *Schmandt* ¶193.

As also discussed in §VIII.A.1.c(v), in the example weather report discussed in this ground, “Carolina” in the second sentence (a(2, 6)) would be assigned a variation degree of “0,” because it is the same as “Carolina” in the first sentence (a(1, 6)). When the variation degree is calculated to be “0,” the variation coefficient is also calculated to be “0.” *Schmandt* ¶¶189-192 (discussing the calculation of the volume variation coefficient for “Carolina”). Therefore, the user-set minimum

variation coefficient is used, which in the case of figure 4, is 0.85. *Schmandt* ¶191. In the specific example weather report discussed herein, 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. *Schmandt* ¶193.

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 at least one second portion of the at least one token,” as recited in claim limitation 1[c] and understood in the context of the ’291 patent. *Schmandt* ¶194. The ’291 patent describes “contrastive stress” as ***increasing*** the prosody for the words/phrases meant to contrast with each other. See ’291Pat, 9:22-27 (“[S]ynthesis 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.”), 12:44-48; *Schmandt* ¶195. The ’291 patent does not disclose or otherwise suggest that decreasing the pitch, amplitude, and/or duration of words/phrases (e.g., to deemphasize a word/phrase) constitutes applying contrastive stress. *Schmandt* ¶196.

Additionally, during prosecution of U.S. Patent 8,447,610 (“the ’610 patent”), Applicant stated that “[c]ontrastive stress’ is a term of art, whose meaning is

explained in detail in the specification, e.g., at pages 11-12.”⁸ ’610FH, 83. The portion of the specification cited by the Applicant describes contrastive stress as “a particular emphasis pattern [] applied in speech to words or syllables that are *meant to contrast with each other*.” ’610FH, 147; ’291Pat, 8:18-20; ’610Pat, 8:12-15. That portion of the specification further states that “[h]uman speakers naturally apply contrastive stress to *emphasize* a word or syllable contrary to its normal

⁸ The ’610 patent is in the same family as the ’291 patent, and is the parent to U.S. Patent 8,682,671 and the grand parent to U.S. Patent 8,825,486 (which Petitioners are challenging in IPR2025-00458 and IPR2025-00459, respectively). *See* Ex. 1022. The Board can consider the file history of the ’610 patent even though it is earlier in time than ’291 patent because the two patents share the same specification (*see* Ex. 1019) and the Applicant’s statements pertain to the same claim term in both patents. *See Microsoft Corp. v. Multi-Tech. Sys., Inc.*, 357 F.3d 1340, 1350 (Fed. Cir. 2004). To be clear, Petitioners do not rely on these cases to argue that Applicant disclaimed claim scope, but for the proposition that Applicant’s statements about the scope of a term in an earlier, but related, patent can be relevant to the construction of the same term in a later patent in the same family.

accentuation, in order to contrast it with an alternative word or syllable or to focus attention on it.” ’610FH, 147; ’291Pat, 8:20-23; ’610Pat, 8:15-18.

Moreover, Applicant explicitly rejected the Examiner’s suggestion that contrastive stress could be read on any form of stress that differs from one word and any other word:

The Examiner questioned whether the term “contrastive stress” could be read on any form of stress that differs between one word and any other word. The Assignee’s representatives pointed out that such an interpretation would not be consistent with how the term “contrastive stress” is generally understood in the art. ***Contrastive stress is a particular type of stress, and not every type of stress is contrastive stress.***

’610FH, 83. Accordingly, both the specification and file history of the ’291 patent are consistent with a POSITA’s understanding that deemphasizing words that are the same is not an application of contrastive stress. *Schmandt* ¶200. Therefore, a POSITA would have understood that when *Matsumoto* deemphasizes words/phrases that are the same, *Matsumoto* does not assign “contrastive stress” because *Matsumoto* is not applying a stress to words/phrases ***meant to contrast*** with each other. *Schmandt* ¶201.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches assigning stress by increasing the speed/duration, pitch, and volume using variation coefficients C1, C2, and C3 to the identified first portion of the at least one token (e.g., “South” of “SC”) of a weather report (“assigning contrastive stress to the identified at least one first portion of the at least one token”) but not increasing the speed/duration, pitch, and volume for the second portion of the at least one token (e.g., “Carolina” of “SC”) (“but not to the identified at least one second portion of the at least one token”). *Schmandt* ¶202.

e. 1[d]

Walker in combination with *Matsumoto* and *Malsheen* teaches this feature. *Schmandt* ¶¶203-211.

Matsumoto discloses a speech synthesizer that generates a synthesized speech output for a given text input by adjusting the speed/duration, pitch, and volume of words or phrases of a new sentence based on assignment of contrastive stress using coefficients C1 (speed/duration), C2 (pitch), and C3 (volume). *See Matsumoto* ¶¶[0103]-[0105]. Specifically, *Matsumoto* discloses the phoneme string with changes to speed/duration, pitch and volume is generated by the “waveform generation unit.” *Id.*, ¶[0107]; *see also id.*, ¶[0043] (“The waveform generation unit 8 generates a synthetic speech from the phoneme length, the phoneme string, the

pitch pattern and the volume information by a predetermined method, and outputs the synthetic speech.”).

Each of *Walker*, *Matsumoto*, and *Malsheen* disclose that their inventions are performed using a computer-based system. *Schmandt* ¶206. For example, *Walker* discloses that its “TTS engines ... synthesize the written words of a text document into **computer generated** speech understandable by humans.” *Walker* ¶[0003]; *see also id.* ¶¶[0019] (disclosing that *Walker* is “critical for wider acceptance of **voice computer interfaces**”), [0024] (*Walker*’s TTS engine “converts or synthesizes ... text from voice application 16 into **computer generated audio speech**.”). Similarly, *Matsumoto* discloses that “[t]he present invention can be actualized as a program that **makes a computer function as the speech synthesizer**.” *Matsumoto* ¶[0018]; *see also id.*, ¶¶[0036], [0093]. *Malsheen* discloses that “the present invention is a **computer system** for converting a text message into an audio representation.” *Malsheen*, 2:62-65; *see also id.*, 3:51-4:5, 4:17-35. Thus, a POSITA would have understood that the *Walker-Matsumoto-Malsheen* combination’s functionality, including generating synthesized speech with appropriate contrastive stress, would also be implemented using a computer system. *Schmandt* ¶210.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches using a computer system to generate audio speech output of a text weather report

(“generating, using at least one computer system, an audio speech output corresponding to at least a portion of the text input”) with changes to speed/duration, pitch and volume using the assigned coefficients C1, C2, and C3 for a differing portion of a token (e.g., “South” of “SC”) of a new sentence (n=2) (“the audio speech output comprising at least one portion carrying contrastive stress corresponding to the at least one first portion of the at least one token of the text input”) but not adjusting the speed/duration, pitch, and volume for a same portion of a token (e.g., “North” of “NC”) in a prior sentence (n=1) (“to contrast with at least one other portion of the audio speech output corresponding to the at least one corresponding first portion of the at least one other token of the text input”). *Schmandt* ¶211.

f. 1[e]

Walker in combination with *Matsumoto* and *Malsheen* teaches this feature. *Schmandt* ¶¶212-216.

Walker discloses that “TTS engine 18 ... converts the cleaned text into speech and provides the speech which is labeled ‘Synthesized Audio’ in FIG. 1 to voice application 16.” *Walker* ¶[0034]. As discussed in §VIII.A.1.c(ii), a POSITA would have found it obvious to incorporate *Matsumoto*’s TTS system into *Walker*’s TTS engine farm as the TTS engine for synthesizing text weather reports into speech. As

discussed in §VIII.A.1.a, a POSITA would have understood that voice application 16 is a speech-enabled application.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches a TTS engine providing synthesized speech to a voice application (“providing the audio speech output for the speech-enabled application”). *Schmandt* ¶216.

2. Independent Claim 8

a. 8[pre]

To the extent the preamble is limiting, *Walker* teaches this feature for the reasons discussed in §VIII.A.1.a. *Schmandt* ¶217.

b. 8[a]-8[b]

To the extent the preamble is limiting, *Walker* in combination with *Matsumoto* and *Malsheen*, teaches these features. *Schmandt* ¶¶218-222.

As discussed in §VIII.A.1.e, each of *Walker*, *Matsumoto*, and *Malsheen* disclose that their inventions are performed using a computer-based system. As discussed in §§VIII.A.1.c(ii) and VIII.A.1.c(iv), a POSITA would have found it obvious to combine the teachings of *Walker*, *Matsumoto*, and *Malsheen*.

A POSITA would have found it obvious for the *Walker-Matsumoto-Malsheen* combination to include a memory for storing a plurality of processor-executable instructions, and to also include a processor, operatively couple to the memory, to

execute the instructions because at least each of *Matsumoto* and *Malsheen* explicitly disclose such an arrangement. See *Matsumoto* ¶¶[0018] (disclosing that its TTS system “can be actualized as a program that makes a computer function as the speech synthesizer ... and as a *storage medium storing this program.*”), [0036] (“the configuration described above is a function actualized in such a way that *the CPU loads the program stored in the hard disc into the memory and executes this program.*”), [0093], FIG. 10; *Malsheen*, 3:59-4:5 (“The computer system 100 includes *a central processing unit 102* The *computer memory 106 stores a number of different programs, sometimes herein called procedures, and data structures.* These programs and data structures, which are discussed in more detail below, include ... a text reader module 128.”), FIG. 2; *Schmandt* ¶220.

A POSITA would have also found it obvious to combine the teachings of *Walker*, *Matsumoto*, and *Malsheen* because it amounted to use of a known technique (*Matsumoto* and *Malsheen* each teach a computer having a processor operatively coupled to a non-transitory memory storing instructions that are executed by the processor) to improve a similar device (the *Walker-Matsumoto-Malsheen* combination is implemented using a computer) in the same way (incorporating a processor coupled to non-transitory memory storing instructions that are executed

by the processor in the *Walker-Matsumoto-Malsheen* combination's computer). *Schmandt* ¶221.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches a computer that includes a hard disk or other memory that stores processor-executable instructions (“a memory storing a plurality of processor-executable instructions”) to be carried out by the computer's processor (“at least one processor ... that executes the instructions”), wherein the processor reads the instructions from the memory (“operatively coupled to the memory”). *Schmandt* ¶222.

c. 8[c]-[g]

Walker in combination with *Matsumoto* and *Malsheen* teaches Elements 8[c]-8[g] for the same reasons discussed in §§VIII.A.1.b-VIII.A.1.f, according to the table below. *Schmandt* ¶223.

Element	Section
8[c]	VIII.A.1.b
8[d]	VIII.A.1.c(i) through VIII.A.1.c(v)
8[e]	VIII.A.1.d
8[f]	VIII.A.1.e
8[g]	VIII.A.1.f

3. Independent Claim 14

a. 14[Pre]

To the extent the preamble is limiting, *Walker* alone or in combination with *Matsumoto* and *Malsheen*, teaches this feature. *Schmandt* ¶¶224-234.

For the reasons discussed in §VIII.A.1.a, *Walker* teaches “a method for providing speech output for a speech-enabled application.”

Walker discloses that “[t]ext-to speech (TTS) engines are **computing devices** which convert written text into audible **computer generated speech**.” *Walker* ¶[0003]; *see also id.*, ¶¶[0019] (disclosing that *Walker* is “critical for wider acceptance of **voice computer interfaces**”), [0024] (*Walker*’s TTS engine “converts or synthesizes ... text from voice application 16 into **computer generated audio speech**.”). Accordingly, a POSITA would have understood that *Walker*’s method is performed using software (i.e., computer readable instructions) running on a computer. *Schmandt* ¶227.

This understanding is confirmed, for example, by the fact that the provisional application from which *Walker* claims priority refers to the context detector and text cleaner as “software components.” Ex. 1015, 5; *Schmandt* ¶228. A POSITA would have further understood that *Walker*’s computer comprises at least one processor for executing the software to implement *Walker*’s method. *Schmandt* ¶229 (describing

Bellegarda). Additionally, a POSITA would have understood that such software components/instructions would be stored in a non-transitory computer-readable storage medium to be available for execution (e.g., so as to not be lost when the computer is powered off). *Schmandt* ¶230.

To the extent PO argues and the Board finds that *Walker* alone does not teach a non-transitory computer-readable storage medium for storing the software components/instructions executed by *Walker*'s computing device, this feature would have been obvious based on *Walker* in combination with *Matsumoto* and *Malsheen*. *Schmandt* ¶231.

For example, *Matsumoto* discloses that its TTS system “can be actualized as a program that makes a computer function as the speech synthesizer ... and as a ***storage medium storing this program.***” *Matsumoto* ¶[0018]; see also *id.*, ¶¶[0036] (“[T]he configuration described above is a function actualized in such a way that ***the CPU loads the program stored in the hard disc into the memory and executes this program.***”), [0093], FIG. 10.

Thus, a POSITA would have found it obvious to implement the functionality of the *Walker-Matsumoto* combination as computer-executable instructions stored on a non-transitory computer-readable medium because *Matsumoto* expressly teaches to do so. *Schmandt* ¶232. A POSITA would have also found it obvious to do

so because it amounted to use of a known technique (*Matsumoto* teaches a non-transitory memory for storing program instructions executed by a computer) to improve a similar device (*Matsumoto* teaches methods performed by a computer) in the same way (incorporating a non-transitory memory in the *Walker-Matsumoto* combination to store the program instruction executed by the computer). *Schmandt* ¶233.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches a system that includes a hard disk or other non-transitory memory (“at least one non-transitory computer readable medium”) that stores computer-executable code to be carried out by a computer’s processor (“storing instructions that, when executed by at least one processor”) to implement the conversion of text from a voice application into speech (“that, when executed, perform a method for use with a speech-enabled application”). *Schmandt* ¶234.

b. 14[a]-[e]

Walker in combination with *Matsumoto* and *Malsheen* teaches Elements 14[a]-14[e] for the same reasons discussed in §§VIII.A.1.b-VIII.A.1.f, according to the table below. *Schmandt* ¶235.

Element	Section
14[a]	VIII.A.1.b

Element	Section
14[b]	VIII.A.1.c(i) through VIII.A.1.c(v)
14[c]	VIII.A.1.d
14[d]	VIII.A.1.e
14[e]	VIII.A.1.f

4. Dependent Claims 2-4, 6-7, 9-11, 13, 15-17, and 19-20

a. Claim 2

(i) 2[a]

Walker in combination with *Matsumoto* and *Malsheen* teaches this feature. *Schmandt* ¶¶236-246.

The '291 patent describes “the same text normalization type” as encompassing the same word type, including state names. *See '291Pat*, 24:45-51 (“Examples of text normalization types of text input fields to which synthesis system 200 may apply contrastive stress patterns include, but are not limited to ... state name ...”).

As discussed in §VIII.A.1.c(iv), *Malsheen* discloses a “text classifier procedure” that classifies whether a word is a state name. Classifying an abbreviation based on type (e.g., state name) is consistent with how the '291 patent describes identifying the same text normalization type of the input text. *'291Pat*,

32:7-16 (“A tokenizer of synthesis system 200 may ... analyze the tokens that it parses to infer their text normalization type ... tokenizer of synthesis system 200 may be programmed to determine that ‘9:42 a.m.’, ‘10:30 a.m.’, ‘11:30 a.m.’ and ‘12:30 p.m.’ in text input 350 are of the ‘time’ text normalization type”). *Schmandt* ¶238.

For the reasons discussed in §§VIII.A.1.c(i) and VIII.A.1.c(iv), a POSITA would have understood the abbreviations “NC” and “SC” in the example weather report discussed in this ground to comprise “tokens.” As also discussed in §VIII.A.1.c(iv), in the *Walker-Matsumoto-Malsheen* combination, the context cleaner would recognize each of “NC” and “SC” to correspond to an abbreviation of a state name and would expand the abbreviations to the full name of the state (e.g., “North Carolina” and “South Carolina”). Additionally, as discussed in §VIII.A.1.d, the *Walker-Matsumoto-Malsheen* combination assigns a contrastive stress pattern to “South,” which is a portion of the token “NC.”

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches identifying that each of the abbreviations “NC” and “SC” (“identifying a plurality of tokens”) represent state names (“of a same text normalization type”) and assigns contrastive stress pattern to “South” which is portion of the expanded token “SC” (“for which a contrastive stress pattern is to be applied”). *Schmandt* ¶242.

Further, as described in §VIII.A.1.c(v) and §VIII.A.1.d, *Matsumoto's* collation unit operates on the expanded state names “North Carolina” and “South Carolina” (i.e., tokens that have undergone text normalization) and identifies the word “South” is different from “North” and therefore, “South” is assigned contrastive stress.

The '291 patent contemplates that applying a “contrastive stress pattern” to a plurality of tokens includes assigning contrastive stress to one of two tokens, and in some embodiments, a portion of one of two tokens. *See '291Pat*, 25:31-34 (“In some other embodiments, synthesis system 200 may be programmed to apply a contrastive stress pattern in which ***only the second of two contrasting fields (i.e., '1145A') is rendered with stress.***”); 23:3-5. Thus, a POSTIA would have understood *Matsumoto's* collation unit 14 identifies one token (i.e., “SC”) of the plurality of tokens (i.e., “NC” or “SC”) of the same text normalization type (i.e., state name) to receive contrastive stress (i.e., “South” receives prosodic adjustment) when synthesized into speech. *Schmandt* ¶245.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches identifying a same text normalization type (i.e., state names) for a plurality of tokens (i.e., “NC” and “SC”) (“wherein the identifying comprises: identifying a plurality of tokens of the text input of a same text normalization type for which a contrastive

stress pattern is to be applied”) and identifying one token (i.e., “SC”) to receive contrastive stress from the plurality of tokens (i.e., “NC” or “SC”) of the same text normalization type (i.e., state names) (“and identifying the at least one token to be rendered with contrastive stress from among the plurality of tokens of the same text normalization type”). *Schmandt* ¶246.

(ii) 2[b]

Walker in combination with *Matsumoto* and *Malsheen* teaches this feature. *Schmandt* ¶¶247-252.

The ’291 patent contemplates that applying a “contrastive stress pattern” to a plurality of tokens includes assigning contrastive stress to one of two tokens, and in some embodiments, a portion of one of two tokens. *See* ’291*Pat*, 25:31-34 (“In some other embodiments, synthesis system 200 may be programmed to apply a *contrastive stress pattern in which only the second of two contrasting fields (i.e., ‘1145A’) is rendered with stress.*”); 23:3-5 (“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)). Thus, a POSTIA would have understood *Matsumoto*’s collation unit 14 identifies one token (i.e., “SC”) of the plurality of tokens (i.e., “NC” or “SC”) of the same text normalization type (i.e.,

state name) to receive contrastive stress (i.e., “South” receives prosodic adjustment) when synthesized into speech. *Schmandt* ¶249.

For the reasons discussed in §§VIII.A.1.c(v) and VIII.A.1.d, the *Walker-Matsumoto-Malsheen* combination teaches identifying “South” as the portion of the abbreviation “SC” after it has been expanded to “South Carolina” to which contrastive stress is assigned (“identifying the at least one token to be rendered with contrastive stress”) from among “NC” and “SC” both of which represent abbreviations of state names (“from among the plurality of tokens of the same text normalization type”). *Schmandt* ¶250. This is consistent with how the ’291 patent describes assigning contrastive stress to a portion of a token after it has undergone normalized orthography. *See* ’291*Pat.*, 27:66-28:5 (“In some embodiments, ***synthesis system 200 may identify the specific sub-portion(s) of one or more fields or tokens that should carry contrastive stress by comparing the normalized orthography of the field(s) or token(s)*** identified to be rendered with contrastive stress to the normalized orthography of the other contrasting field(s) or token(s), and determining which specific portion or portions differ.”), 28:27-32 (“[I]n the example of FIG. 3A, ***synthesis system 200 may compare the normalized orthography ‘eleven_forty_five_a_m’ to the normalized orthography ‘ten_forty_five_a_m_’ to determine that “eleven” is the portion that differs.*** As a result, synthesis system 200

may assign the word “eleven” to carry contrastive stress.”), 28:62-67 (“*Identifying specific portions of one or more fields or tokens to be rendered to carry contrastive stress through a comparison of normalized orthography* in some embodiments may provide the advantage that the portions that differ can be readily identified in terms of their spoken word forms as they will be rendered in the speech output.”); *Schmandt* ¶251.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches identifying “South” as the portion of the abbreviation “SC” after it has been expanded to “South Carolina” to which contrastive stress is assigned (“identifying the at least one token to be rendered with contrastive stress”) from among “NC” and “SC” both of which represent abbreviations of state names (“from among the plurality of tokens of the same text normalization type”). *Schmandt* ¶252.

b. Claim 3

For the same reasons discussed in §VIII.A.4.a, *Walker-Matsumoto-Malsheen* combination teaches this claim, particularly that “the same text normalization type is a text normalization type selected from the group consisting of: ... a state name type.” *Schmandt* ¶253.

c. Claim 4

Walker in combination with *Matsumoto* and *Malsheen* teaches this claim. *Schmandt* ¶¶254-257.

As discussed in §§VIII.A.1.c-VIII.A.1.d, *Matsumoto* discloses comparing words in corresponding “b” positions of a current sentence and a prior sentence to identify words that are different and then adjusting the pitch, speed, and volume of the words that are different (i.e., assigns “contrastive stress” to the words that are different). Comparing corresponding words (i.e., in the same “b” position) is consistent with how the ’291 patent describes indications in the text input of contrastive stress. *See* ’291*Pat*, 26:52-55 (“In some embodiments, fields and/or tokens to be rendered with contrastive stress may also or alternatively be identified based on part-of-speech sequences that establish a repeated pattern with one element different.”). *Schmandt* ¶256.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches identifying the abbreviations “NC” and “SC” representing state names (“wherein the plurality of tokens are identified”), expanding the abbreviations to the full state name and assigning stress to the word “South” in the new sentence based on the fact that it is different from the word “North” in the same “b” position in the previous sentence (“based at least in part on at least one indication in the text input that the

contrastive stress pattern is desired in association with the plurality of tokens”).
Schmandt ¶257.

d. Claim 6

(i) 6[a]

Walker in combination with *Matsumoto* and *Malsheen* teaches this feature.
Schmandt ¶¶258-270.

A POSITA would have understood the term “tokenizing” or “tokenization” to refer to breaking up input text into its component parts (e.g., words/tokens). *Schmandt* ¶¶259-262 (discussing *Jurafsky*, *Addison*, *Bennett*). The ’291 patent uses the term “tokenizing” consistent with this understanding. *See* ’291*Pat*, 31:56-59 (“Synthesis system 200 may be programmed, e.g., through a tokenizer implemented as part of front-end 250, to parse text input 350 into a sequence of ***individual tokens on the order of single words.***”), 36:23-25 (“At act 410, the text input may be ***tokenized, i.e., parsed into individual tokens on the order of single words.***”); *see also id.*, 32:2-6. *Schmandt* ¶263.

Walker discloses that “context detector 26 processes the raw text to determine the context of the text by locating in the text key words and phrases associated with each context detection rule set stored in context detection rules database 21” *Walker* ¶[0030]; *see also id.*, ¶[0031]. A POSITA would have understood that for *Walker*’s

context detector to compare words in the raw text input with words in a context detection rules databases, the raw text input would have been broken up into its constituent words for analysis. *Schmandt* ¶265. In other words, a POSITA reading *Walker* would have understood *Walker* to teach breaking the raw text input into its individual words, and then comparing those individual words to words in the context detection rules database. *Schmandt* ¶265.

To the extent that the Board finds that *Walker* alone does not teach tokenizing the text input, doing so would have been obvious based on *Malsheen*'s teachings. *Schmandt* ¶266. For the reasons discussed in §VIII.A.1.c(iv), it would have been obvious to a POSITA to incorporate *Malsheen*'s text classifier procedure that determines and classifies the syntax aspects of each word including making determinations as to whether a word is a member of a particular class (e.g., geographic names) into the *Walker-Matsumoto* combination. *Schmandt* ¶266.

As part of implementing its text classifier procedure, *Malsheen* teaches “generat[ing] eleven arrays 138 of classification values, each of which stores a single element ***for each ‘word’ in the paragraph being processed.*** A word, for purposes of the enhanced text reader 128 is any chunk of text that requires separate processing. For instance, in strings having special punctuation such as slashes, hyphens, and colons, each text element separated by such punctuation is treated as a separate word.

Thus the string ‘3/12/95’ is treated as three words: 1) ‘3/’, 2) ‘12/’ and 3) ‘95’.” *Malsheen*, 5:20-29. A POSITA would have understood that for *Malsheen*’s text classifier procedure to generate the arrays “*for each word*” in the paragraph of text, *Malsheen*’s text classifier would have had to separate the text into its constituent words. *Schmand* ¶268. Indeed, *Malsheen* provides the example of the date 3/12/95 being separated into and treated as three words: 3/, 12/, and 95. *Schmandt* ¶269. In other words, a POSITA reading *Malsheen* would have understood *Malsheen* to teach that its text classifier procedure breaks the text input into its individual words, and then further processes each word to determine its classification. *Schmandt* ¶269.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches breaking the raw text input into its constituent words (“tokenizing the text input”). *Schmandt* ¶270.

(ii) 6[b]-6[c]

Walker in combination with *Matsumoto* and *Malsheen* teaches Elements 6[b]-6[c] for the same reasons discussed in §§VIII.A.4.a(i)-VIII.A.4.a(ii), according to the table below. *Schmandt* ¶271.

Element	Section
6[b]	VIII.A.4.a(i)
6[c]	VIII.A.4.a(ii)

For example, for the same reasons discussed in §VIII.A.4.a(i), a POSITA would have understood the *Walker-Matsumoto-Malsheen* combination to teach a computer identifying tokens “SC” and “NC” as abbreviations for state names (“automatically identifying the text normalization type of the plurality of tokens”). *Schmandt* ¶272.

Further, as discussed in §§VIII.A.1.c-VIII.A.1.d, *Matsumoto*’s collation unit 14 compares the words in corresponding “b” positions to determine how to assign variation coefficients for pitch, speed/duration, and volume to different words. Thus, a POSITA would have understood the *Walker-Matsumoto-Malsheen* combination to teach automatically determining that a contrastive stress pattern is to be applied to the plurality of tokens (i.e., “NC” and “SC”) (“automatically determining that the contrastive stress pattern is to be applied for the plurality of tokens.”). *Schmandt* ¶273.

e. Claim 7

Walker in combination with *Matsumoto* and *Malsheen* teaches this feature. *Schmandt* ¶¶274-277.

As described in §VIII.A.1.c(v), *Matsumoto* discloses that collation unit 14 collates two sets of a(n, b) such that words in the same “b” position are compared. The “b” position is determined by the order of words in a sentence. *See Matsumoto*

¶[0073]. Indeed, using the example weather report described herein, *Matsumoto*'s collation unit determines whether words in the second sentence (n=2)—the second occurrence of words in the same “b” position—are the same as, or different from, corresponding words in the first sentence (n=1), and therefore, whether words in the second sentence (n=2) are assigned contrastive stress. See §VIII.A.1.c(v) (application of the *Walker-Matsumoto-Malsheen* combination to the example text weather report). *Schmandt* ¶275.

As also described in §VIII.A.1.c(v), a POSITA would have understood that collation unit 14 compares corresponding portions of two tokens because “North” and “South” both appear in the b=5 position and “Carolina” and “Carolina” both appear in the b=6 position. Thus, a POSITA would have understood *Matsumoto*'s collation unit 14 calculates the variation degree based on the order of a word within a sentence. *Schmandt* ¶276.

Accordingly, the *Walker-Matsumoto-Malsheen* combination teaches identifying words of a weather report to be the same or different between two or more sentences (“wherein the at least one token to be rendered with contrastive stress is identified”) based on the order of how words appear in a sentence (“based at least in part on an order of the plurality of tokens in the text input”). *Schmandt* ¶277.

f. Claims 9-11, 13

A POSITA would have understood the *Walker-Matsumoto-Malsheen* combination to teach an apparatus performing the functions recited in claims 9-11 and 13 for the same reasons as described in §§VIII.A.4.a-VIII.A.4.d, according to the table below. *Schmandt* ¶278.

Claim	Section
9	VIII.A.4.a(i) and VIII.A.4.a(ii)
10	VIII.A.4.b
11	VIII.A.4.c
13	VIII.A.4.d(i) and VIII.A.4.d(ii)

g. Claims 15-17 and 19-20

A POSITA would have understood the *Walker-Matsumoto-Malsheen* combination to teach the at least one software module performing the functions recited in claims 15-17, and 19-20 for the same reasons as described in §§VIII.A.4.a-VIII.A.4.e, according to the table below. *Schmandt* ¶279.

Claim	Section
15	VIII.A.4.a(i) and VIII.A.4.a(ii)
16	VIII.A.4.b
17	VIII.A.4.c

Claim	Section
19	VIII.A.4.d(i) and VIII.A.4.d(ii)
20	VIII.A.4.e

B. Ground 1B: *Walker* in Combination With *Matsumoto*, *Malsheen*, and *Bellegarda* Renders Obvious Claims 1-4, 6-11, 13-17, and 19-20

1. 1[c], 8[e], 14[c]

The *Walker-Matsumoto-Malsheen* combination teaches elements 1[c], 8[e], and 14[c] for the reasons discussed in §§VIII.A.1.d (limitation 1[c]), VIII.A.2.c (limitation 8[e]), and VIII.A.3.b (limitation 14[c]), respectively. *Schmandt* ¶280.

To the extent that PO argues and the Board agrees that deemphasizing a word/phrase constitutes assigning contrastive stress—and, therefore, *Walker-Matsumoto-Malsheen* does not meet the portions of elements 1[c], 8[e], and 14[c] that require not assigning contrastive stress to the identified at least one second portion of the at least one token—these claim limitations would nevertheless have been obvious to a POSITA based on *Bellegarda*’s teachings to configure *Walker-Matsumoto-Malsheen*’s TTS engine such that words that are the same (e.g., “Carolina” at a(2, 6) as compared to “Carolina” at a(1, 6)) are not deemphasized. *Schmandt* ¶281.

For example, *Bellegarda* discloses a “speech synthesis system 100 [that] receives a text input 104” (*Bellegarda*, 5:8-9) and incorporates a “word prominence specification system 200” that “applies word prominence” (*id.*, 5:21-26). *See also id.*, FIG. 1. In one embodiment, *Bellegarda* discloses a speech synthesizer that stores and tracks previously spoken words in a buffer such that newly synthesized words that are the same as previous words are not emphasized and receive no change when spoken. *See id.*, 5:60-67 (“For example, in one alternative embodiment, each new word is considered a candidate for prominence, and a list of previously spoken words is maintained in a FIFO (first-in-first-out) buffer having a specified depth. ***If a current word is already in the FIFO buffer, no accent is applied to the word when spoken***, but if the word is not in the buffer (i.e., the current word is a “new” word), prominence is applied to the word.”).

As noted in §VIII.A.1.d, *Matsumoto*’s maximum and minimum variation coefficient values are user-settable. Thus, it would have been obvious for a POSITA to set the minimum variation coefficient value to a value above 0, including 1.0. *Schmandt* ¶283. As further noted in §VIII.A.1.d, if the minimum variation coefficient is set to 1.0, there would be no change to the variation coefficients for words/phrases determined to be the same. Therefore, a POSITA would have understood that the minimum variation coefficient value could be set to a value of

1.0. *Schmandt* ¶284. 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. *Schmandt* ¶285.

Only adjusting the prosody of words determined to be different between two sentences 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 sentences) for another (*Matsumoto* teaches deemphasizing corresponding words that are the same between two sentences) to obtain predictable results (in the *Walker-Matsumoto-Malsheen-Bellegarda* combination corresponding words that different between two sentences would continue to receive contrastive stress). *Schmandt* ¶286.

A POSITA would have also recognized that not adjusting the prosody of corresponding words that are the same between two sentences would have reduced the computations required by the *Walker-Matsumoto-Malsheen-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 sentences. *Schmandt* ¶287. Thus, the combination would have also been obvious to a POSITA because it amounted to use of a known technique

(*Bellegarda* teaches not adjusting the prosody of words that are the same between two sentences) to improve similar devices (to reduce the computational requirements of *Walker-Matsumoto-Malsheen*'s TTS engine by) in the same way (by not adjusting the prosody of words that are the same between the two sentences in the *Walker-Matsumoto-Malsheen-Bellegarda* combination). *Schmandt* ¶288.

Accordingly, the *Walker-Matsumoto-Malsheen-Bellegarda* combination teaches assigning stress by increasing the speed/duration, pitch, and volume using variation coefficients C1, C2, and C3 to the identified first portion of the at least one token (e.g., “South” of “SC”)—(“ assigning contrastive stress to the identified at least one first portion of the at least one token”) but not changing the speed/duration, pitch, and volume for the second portion of the at least one token (e.g., “Carolina” of “SC”) —(“but not to the identified at least one second portion of the at least one token”). *Schmandt* ¶289.

2. All Other Features of Claims 1-4, 6-11, 13-17, and 19-20

A POSITA would have understood the *Walker-Matsumoto-Malsheen-Bellegarda* combination to teach all other features of claims 1-4, 6-11, 13-17, and 19-20 for the reasons previously discussed according to the table below. *Schmandt* ¶290.

Element/Claim	Section
1[pre]	VIII.A.1.a
1[a]	VIII.A.1.b
1[b]	VIII.A.1.c(i) through VIII.A.1.c(v)
1[d]	VIII.A.1.e
1[e]	VIII.A.1.f
2[a]	VIII.A.4.a(i)
2[b]	VIII.A.4.a(ii)
3	VIII.A.4.b
4	VIII.A.4.c
6	VIII.A.4.d(i) through VIII.A.4.d(ii)
7	VIII.A.4.e
8[pre]	VIII.A.2.a
8[a]	VIII.A.2.b
8[b]	VIII.A.2.b
8[c]	VIII.A.1.b
8[d]	VIII.A.1.c(i) through VIII.A.1.c(v)
8[f]	VIII.A.1.e
8[g]	VIII.A.1.f

Element/Claim	Section
9	VIII.A.4.a(i) and VIII.A.4.a(ii)
10	VIII.A.4.b
11	VIII.A.4.c
13	VIII.A.4.d(i) and VIII.A.4.d(ii)
14[pre]	VIII.A.3.a
14[a]	VIII.A.1.b
14[b]	VIII.A.1.c(i) through VIII.A.1.c(v)
14[d]	VIII.A.1.e
14[e]	VIII.A.1.f
15	VIII.A.4.a(i) and VIII.A.4.a(ii)
16	VIII.A.4.b
17	VIII.A.4.c
19	VIII.A.4.d(i) and VIII.A.4.d(ii)
20	VIII.A.4.e

IX. SECONDARY CONSIDERATIONS

There is no evidence in the '291 patent's prosecution history or elsewhere supporting any secondary considerations arguments, or evidence of nexus of such alleged evidence to the Challenged Claims. *See generally* '291FH. To the extent PO

asserts the existence of any secondary considerations in its responses, Petitioners reserve the right to address any such evidence.

X. THE DISCRETIONARY FACTORS FAVOR INSTITUTING TRIAL

A. 35 U.S.C. §314(a)

To the extent the Patent Owner asks the Board to exercise its discretion to deny institution despite the strong invalidity showing on the merits, the Board should decline to do so because the weight of the factors articulated in *Apple Inc. v. Fintiv, Inc.*, IPR2020-00019, Paper 11 at 6 (PTAB Mar. 20, 2020) (precedential) favors institution.⁹

1. Stay

Factor 1 is neutral. Petitioners have not requested a stay but intend to do so. The PTAB has explained that it will not speculate on how any such motion would be resolved, before one is filed. *Google LLC v. Parus Holdings, Inc.*, IPR2020-00847, Paper 9 at 12 (PTAB Oct. 21, 2020); *see also Hulu LLC v. SITO Mobile R&D IP, LLC*, IPR2021-00298, Paper 11 at 10-11 (PTAB May 19, 2021) (concluding that

⁹ The *Fintiv* framework should not be followed because it is legally invalid as (1) exceeding the Director's authority, (2) arbitrary and capricious, and (3) adopted without notice-and-comment rulemaking.

without evidence of a requested stay or consideration by the district court in the parallel litigation, this factor does not significantly impact the Board's exercising discretion to deny institution of the IPR).

2. Trial Date

The docket control order in the parallel litigation currently sets jury selection for January 5, 2026, about 12 months away. Ex. 1020. Any comparison of a projected FWD date against the scheduled trial date is speculative. *See Dish Network v. Broadband iTV*, IPR2020-01280, Paper 17 at 16 (PTAB Feb. 4, 2021) (“We cannot ignore the fact that the currently scheduled trial date is more than nine months away and much can change during this time”).

Setting that aside, this Petition is filed on January 21, 2025, so a FWD would be expected by July 2026. Although this is approximately six months after the currently scheduled (speculative) trial date, this factor is not determinative or considered in isolation. *Facebook, Inc. v. USC IP P'Ship*, IPR2021-00034, Paper 13 at 11 (PTAB April 13, 2021) (“[T]his factor is not considered in isolation, but holistically along with other factors”) (citation omitted). Moreover, the Board has instituted IPR and found, on similar facts, that this factor weighs only minimally in favor of denial. *See, e.g., Google LLC, et al. v. Multimodal Media LLC*, IPR2024-00056, Paper 9 at 8 (PTAB Apr. 12, 2024) (a period of about six months between

trial and the expected FWD “weighs only marginally in favor” of denial); *NetNut Ltd. v. Bright Data Ltd.*, IPR2021-01492, Paper 12 at 9-16 (PTAB Mar. 21, 2022) (instituting IPR without stipulation and copending trial date six months before FWD); *Facebook*, IPR2021-00034, Paper 13 at 11 (a period of five months between trial and expected FWD “slightly favors denial”); *Coolit Systems, Inc. v. Asetek Danmark A/S*, IPR2021-01195, Paper 10 at 11 (PTAB Dec. 28, 2021) (a period of five months between trial and expected FWD “weighs slightly in favor” of denial); *Equipmentsshare.com Inc. v. Ahern Rentals, Inc.*, IPR2021-00834, Paper 19 at (PTAB Nov. 16, 2021) (a period of seven months between trial and expected FWD “weighs somewhat in favor” of denial).

3. Diligence/Investment

Factor 3 weighs strongly against discretionary denial. No substantive orders have been issued by the court in the underlying litigation, and investment in the parallel litigation against Petitioner Samsung will have remained low at the time of institution. *See, e.g., Hulu LLC v. SITO Mobile R&D IP, LLC*, IPR2021-00298, Paper 11 at 12-14 (PTAB May 19, 2021) (holding that this factor supports instituting IPR given the early stage of the district court proceedings, the lack of substantial discovery related to invalidity claims, and the petitioner’s diligent filing of the petition after receiving preliminary infringement contentions). Assuming that a

Decision on Institution is issued by July 2025, much work in district court will remain. *Id.* Claim construction briefing does not begin until May 23, 2025, and the *Markman* hearing is scheduled for July 11, 2025. Opening expert reports are due August 11, 2025, and dispositive motions are due September 29, 2025. *Id.*

4. Overlap

Factor 4 weighs against discretionary denial. Petitioners challenge claims 1-4, 6-11, 13-17, and 19-20 in this Petition, whereas Patent Owner has alleged infringement of only claims 1-2, 4, 6, and 7. *See* Ex. 1021. Accordingly, a material number of the challenged claims will not be addressed by the district court. *See, e.g., Precision Planting LLC v. Maschio Gaspardo S.p.A.*, IPR2024-00008, Paper 12 at 18 (PTAB Mar. 26, 2024) (holding that despite substantial overlap in issues between the IPR petition and parallel district court action, the inclusion of claims in the petition not contested in court argues against discretionary denial due to incomplete overlap).

5. Parties

Petitioners and PO are parties to the parallel litigation. However, with respect to Factor 5, the *Fintiv* decision “says nothing about situations in which the petitioner is the same as, or is related to, the district court defendant.” *Cisco Systems, Inc. v. Ramot at Tel Aviv University Ltd.*, IPR2020-00122, Paper 15, 10 (Crumbley,

dissenting); *Fintiv*, 13-14. Accordingly, Petitioners submit that Factor 5 should, at worst, be neutral.

6. Other considerations.

Even if the *Fintiv* factors favor discretionary denial (which they do not), the merits are compelling here. Petitioners therefore respectfully submit that the *Fintiv* factors favor institution and that discretionary denial of this Petition would be neither appropriate nor equitable.

B. 35 U.S.C. §325(d)

None of the references presented in the grounds herein were previously considered by the Office. Accordingly, the Board should not exercise its discretion under §325(d) to deny institution.

XI. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8

A. Real Parties-in-Interest

Petitioners identifies themselves as the real parties-in-interest.

1. Related Matters

To the best of Petitioners' knowledge, the '291 patent has only been involved in the following district court litigation: *Cerence Operating Co. v. Samsung Electronics Co., Ltd. et al.*, 2:24-cv-00181-JRG-RSP (E.D. Tex.), filed March 15, 2024.

To the best of Petitioners' knowledge, the '29 1 patent has not been challenged in any other *inter partes* review or post-grant review prior to this proceeding.

Petitioners are concurrently filing IPR2025-00458 directed to U.S. Patent No. 8,682,671, and IPR2025-00459 directed to U.S. Patent No. 8,825,486, both of which are related to the '291 patent. *See* Ex. 1022.

B. Lead and Backup Counsel

Lead Counsel	Back-Up Counsel
Ali R. Sharifahmadian, Reg. No. 48,202 (ali.sharifahmadian@arnoldporter.com) Arnold & Porter Kaye Scholer LLP 601 Massachusetts Ave., NW Washington, DC 20001-3743 Tel: 202-942-5000 Fax: 202-942-5999	Patrick Reidy (Reg. No. 72,148) patrick.reidy@arnoldporter.com Olivia Mathews (Reg. No. 80,801) olivia.mathews@arnoldporter.com Arnold & Porter Kaye Scholer LLP 70 West Madison Street Suite 4200 Chicago, IL 60602-4231 Tel: 312-583-2300 Fax: 312-583-2360 Douglas L. Clark (Reg. No. 68,443) douglas.clark@arnoldporter.com Arnold & Porter Kaye Scholer LLP 777 South Figueroa St., 44th Floor Los Angeles, CA 900017-5844 Tel: 213-243-4008 Fax: 213-243-4199

C. Service Information

Please address all correspondence to lead and back-up counsel at the addresses shown above. Petitioner consents to electronic service by e-mail at the following addresses:

ali.sharifahmadian@arnoldporter.com
patrick.reidy@arnoldporter.com
douglas.clark@arnoldporter.com
olivia.mathews@arnoldporter.com
xSamsungCerence2AP@arnoldporter.com

D. Power of Attorney

A power of attorney is filed herewith according to 37 C.F.R. §42.10(b).

XII. FEES

Petitioners concurrently electronically submits the required fees for this Petition. The Board is authorized to charge Arnold & Porter Kaye Scholer LLP's deposit account, No. 50-2387, for any fee deficiency.

Date: January 21, 2025

Respectfully submitted,

/ Ali R. Sharifahmadian /
Ali R. Sharifahmadian (Reg. No. 48,202)
Counsel for Petitioner

CERTIFICATE OF COMPLIANCE

The undersigned hereby certifies that the foregoing Petition for *Inter Partes* Review contains 12,908 words, excluding those portions identified in 37 C.F.R. § 42.24(a), as measured by the word-processing system used to prepare this paper.

/Ali R. Sharifahmadian/

Ali R. Sharifahmadian (Reg. No. 48,202)
Counsel for Petitioner

CERTIFICATE OF SERVICE

I certify that on January 21, 2025, I caused a true and correct copy of the foregoing Petition for *Inter Partes* Review of U.S. Patent No. 8,914,291 and supporting exhibits to be served via overnight delivery on the Patent Owner at the following correspondence address of record as listed on PAIR:

Douglas Wolf
Nuance Communications, Inc.
One Wayside Road
Burlington, MA 01803

A courtesy copy was also sent via electronic mail to Patent Owner's litigation counsel listed below:

Qi Tong: ptong@raklaw.com
James Tsuei jtsuei@raklaw.com
Benjamin Wang: bwang@raklaw.com
Andrew Weiss: aweiss@raklaw.com
Daniel Kolko: dkolko@raklaw.com
Minna Chan: mchan@raklaw.com
Shani Williams: swilliams@raklaw.com
Paul Kroeger: pkroeger@raklaw.com

/Ali R. Sharifahmadian/

Ali R. Sharifahmadian (Reg. No. 48,202)
Counsel for Petitioner