

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

BUFFALO PATENTS, LLC,
Patent Owner.

IPR2023-01386
Patent 6,904,405 B2

Before JAMESON LEE, MAHSHID D. SAADAT, and
CHRISTOPHER L. OGDEN, *Administrative Patent Judges*.

LEE, *Administrative Patent Judge*.

DECISION
Granting Institution of *Inter Partes* Review
35 U.S.C. § 314

I. INTRODUCTION

Google LLC (“Petitioner”) filed a Petition requesting an *inter partes* review of claims 1–12 (“challenged claims”) of U.S. Patent No. 6,904,405 B2 (Ex. 1001, “the ’405 patent”). Paper 1. Buffalo Patents LLC (“Patent Owner”) has not filed a preliminary response.

We have authority to determine whether to institute review under 35 U.S.C. § 314 and 37 C.F.R. § 42.4. Upon consideration of the contentions and the evidence of record before us, we conclude Petitioner has shown a reasonable likelihood that it would prevail in establishing unpatentability of at least one challenged claim of the ’405 patent. Accordingly, we institute an *inter partes* review of all challenged claims of the ’405 patent on all grounds set forth in the Petition.

II. BACKGROUND

A. *Real Parties in Interest*

Petitioner identifies itself and possibly Motorola Mobility LLC as real parties in interest. Pet. 73. Patent Owner identifies itself as the real party in interest. Paper 4, 1.

B. *Related Matters*

The parties identify the following four district court litigations as related to the ’405 patent (Pet. 74; Paper 4, 2):

1. *Buffalo Patents, LLC v. ZTE Corp.*, Case No. 6-22-cv-00423 (W.D. Tex.);
2. *Buffalo Patents, LLC v. TCL Tech. Grp. Corp.*, Case No. 6-22-cv-00422 (W.D. Tex.);
3. *Buffalo Patents, LLC v. OnePlus Tech. (Shenzhen) Co.*, Case No. 6-22-cv-00421 (W.D. Tex.);

4. *Buffalo Patents, LLC v. Motorola Mobility LLC*, Case
No. 1-22-cv-04540 (N.D. Ill.).

C. *The '405 Patent*

The '405 patent discloses methods and systems which perform multiple different types of message recognition using a shared language model. Ex. 1001, code (57). Message recognition of a first type is performed responsive to a first type of message input, such as speech, to provide text data according to both the shared language model and a first model specific to the first type of message recognition, such as an acoustic model. *Id.* Message recognition of a second type is performed responsive to a second type of message input, such as handwriting, to provide text data according to both the shared language model and a second model specific to the second type of message recognition, such as a model that determines basic units of handwriting conveyed by freehand input. *Id.* “Accuracy of both such message recognizers can be improved by user correction of misrecognition by either one of them.” *Id.*

Figure 3B of the '405 patent is reproduced below:

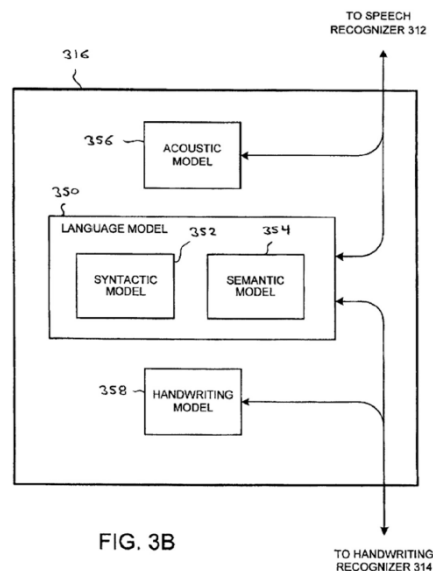


Figure 3B is a functional block diagram of a message model in a domain controller of the '405 patent. Ex. 1001, 5:16–17. The '405 patent describes:

Local message model 316 includes a language model 350, an acoustic model 356, and a handwriting model 358. An acoustic model is any model that may be used to determine basic units of speech (e.g., phonemes, allophones, etc.) responsive to suitably transformed voice input (e.g., by conventional acoustic feature extractor). A handwriting model is any model that may be used to determine basic units of handwriting conveyed by freehand input (e.g., strokes of printed characters or cursive words) responsive to suitably transformed freehand input, for example as disclosed in U.S. Pat. No. 5,903,668 to Beemink (incorporated by reference above).

Speech recognizer 312 generates text data in accordance with acoustic model 316 and shared language model 350. Handwriting recognizer 314 generates text data in accordance with handwriting model 358 and shared language model 350. Advantageously, user training of language model 350 can improve accuracy of both speech recognizer 312 and handwriting recognizer 314. For example, user correction of misrecognition by speech recognizer 312 adapts syntactic model 352 of shared language model 350. Both speech recognizer 312 and handwriting recognizer 314 then operate in accordance with the adapted syntactic model.

Ex. 1001, 22:53–23:8.

D. Illustrative Claims

Claims 1 and 7 are independent. Claims 2–6 each depend, directly or indirectly, from claim 1. Claims 8–12 each depend, directly or indirectly, from claim 7. Claims 1 and 7 are reproduced below:

1. [1P] A system for generating text responsive to message input of a first type and a second type, the system comprising:
 - [1a] (a) a unified message model including:
 - (1) a shared language model;

(2) a first model specific to a first type of message recognition; and

(3) a second model specific to a second type of message recognition;

[1b] (b) a first message recognizer, responsive to the first type of message input to provide text data in accordance with both the shared language model and the first model; and

[1c] (c) a second message recognizer, responsive to the second type of message input to provide text data in accordance with both the shared language model and the second model; wherein

[1d] (d) the shared language model is trainable responsive to user correction of misrecognition by either of the first and second message recognizers, thereby improving accuracy of each of the first and second message recognizers.

Ex. 1001, 39:21–43.¹

7. [7P] A method for performing message recognition with a shared language model, the method comprising:

[7a] (a) performing message recognition of a first type, responsive to a first type of message input, to provide text data in accordance with both the shared language model and a first model specific to the first type of message recognition;

[7b] (b) performing message recognition of a second type, responsive to a second type of message input, to provide text data in accordance with both the shared language model and a second model specific to the second type of message recognition; and

[7c] (c) training the shared language model responsive to user correction of error in message recognition of either of the first and second types, thereby improving accuracy of each of the first [and] second types of message recognition.

¹ The bracketed heading labels correspond to those used by Petitioner to reference the claim elements. *See* Pet. 24–25, 33–35. We use the same labels here for ease of reference, understanding, and consistency.

Ex. 1001, 40:5–22.²

E. Prior Art and Declaration Evidence

Petitioner relies on the following evidence:³

Waibel et al., *Multimodal Interfaces for Multimedia Information Agents*, 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 167–170 (April 21–24, 1997) (Munich, Germany) (Ex. 1005) (“Waibel”);

Suhm et al., *JANUS: Towards Multilingual Spoken Language Translation*, Proc. ARPA Spoken Language Systems Technology Workshop (Jan. 22–25, 1995) (Austin, Texas) pp. 221–226 (Ex. 1008) (“Suhm”);

Vo et al., *Building and Application Framework for Speech and Pen Input Integration in Multimodal Learning Interfaces*, Proc. 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, Vol. VI, pp. 3545–3548 (May 7–10, 1996) (Atlanta, GA) (Ex. 1009) (“Vo”);

S. Manke et al., *The Use of Dynamic Writing Information in a Connectionist On-Line Cursive Handwriting Recognition System*, Advances in Neural Information Processing Systems 7 (1995), pp. 1093–1100 (Ex. 1010) (“Manke”);

U.S. Patent No. 5,027,406, issued June 25, 1991, filed December 6, 1988 (Ex. 1007) (“Roberts”);

² The bracketed heading labels correspond to those used by Petitioner to reference the claim elements. See Pet. 45–46. We use the same labels here for ease of reference, understanding, and consistency. Also, the word “and” in limitation [7c] appears to have been misspelled as “arid.” Ex. 1001, 40:21. We read and understand the word as “and” shown in bracket above.

³ The ’405 patent issued from Application No. 10/061,052, filed Jan. 28, 2002, which is a continuation of Application No. 09/616,157, and claims priority to Provisional Application No. 60/144,481, filed July 17, 1999. Ex. 1001, codes (21), (22), (60), (63).

EP 0622724 A2, published Nov. 2, 1994 (Ex. 1011)
 (“Verrier”).

Petitioner also relies on the Declaration of Christopher Schmandt (Ex. 1004) and the Declarations of Shauna Wiest (Ex. 1019–1022). Patent Owner has not submitted any declaration testimony for consideration.

F. The Asserted Grounds of Unpatentability

Petitioner asserts that the challenged claims of the ’405 patent are unpatentable based on the following grounds (Pet. 23, 48, 53, 65, 69):

Claims Challenged	35 U.S.C. §	References/Basis
1–3, 6–9, 12	103(a) ⁴	Waibel, Roberts
1–3, 6–9, 12	103(a)	Waibel, Suhm, Vo, Manke, Roberts
4, 5, 10, 11	103(a)	Waibel, Roberts, Verrier
4, 5, 10, 11	103(a)	Waibel, Suhm, Vo, Manke, Roberts, Verrier
1–3, 7–9	103(a)	Roth, Roberts
4, 5, 10, 11	103(a)	Roth, Roberts, Verrier
6, 12	103(a)	Roth, Roberts, Vo

⁴ The Leahy-Smith America Invents Act (“AIA”), Pub. L. No. 112–29, 125 Stat. 284, 287–88 (2011), amended 35 U.S.C. §§ 102 and 103. Because the ’405 patent claims priority to an application filed before March 16, 2013 (the effective date of the relevant amendments), the pre-AIA versions of §§ 102 and 103 apply.

III. ANALYSIS

A. *Legal Standards*

A claim is unpatentable under 35 U.S.C. § 103(a) if the differences between the claimed subject matter and the prior art are such that the subject matter, as a whole, would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. *See KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 406 (2007).

The question of obviousness is resolved on the basis of underlying factual determinations, including (1) the scope and content of the prior art; (2) any differences between the claimed subject matter and the prior art; (3) the level of skill in the art; and (4) where in evidence, so-called secondary considerations.⁵ *See Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966).

B. *Level of Ordinary Skill in the Art*

Citing testimony from Christopher Schmandt, Petitioner contends a person of ordinary skill in the art “would have at least a bachelor’s degree in computer engineering, computer science, electrical engineering, or a similar field, and approximately three years of industry or academic experience in a field related to acoustics, speech recognition, speech detection, handwriting recognition, or signal processing.” Pet. 22–23 (citing Ex. 1004 ¶¶ 33–37). Petitioner further notes that “[w]ork experience can substitute for formal education and additional formal education can substitute for work experience.” *Id.* at 23.

At this preliminary stage and on this record, we adopt Petitioner’s statement of the level of ordinary skill in the art. It is supported by the

⁵ The present record does not include any evidence of secondary considerations of nonobviousness.

testimony of Christopher Schmandt and not disputed by Patent Owner. Further, it appears consistent with what is reflected by the content of the applied prior art references. *Cf. Okajima v. Bourdeau*, 261 F.3d 1350, 1354–55 (Fed. Cir. 2001) (the applied prior art may reflect an appropriate level of skill).

C. Claim Interpretation

We use the same claim construction standard that would be used to construe a claim in a civil action under 35 U.S.C. § 282(b), including construing the claim in accordance with the ordinary and customary meaning of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent. 37 C.F.R. § 42.100(b) (2022). The claim construction standard set forth in *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc) is applicable.

Claim terms are generally given their ordinary and customary meaning as would be understood by one with ordinary skill in the art in the context of the specification, the prosecution history, other claims, and extrinsic evidence including expert and inventor testimony, dictionaries, and learned treatises, although extrinsic evidence is less significant than the intrinsic record. *Phillips*, 415 F.3d at 1312–17. Usually, the specification is dispositive, and it is the single best guide to the meaning of a disputed term. *Id.* at 1315.

The specification may reveal a special definition given to a claim term by the patentee, or the specification or prosecution history may reveal an intentional disclaimer or disavowal of claim scope by the inventor. *Id.* at 1316. “There are only two exceptions to this general rule: 1) when a patentee sets out a definition and acts as his own lexicographer, or 2) when

the patentee disavows the full scope of a claim term either in the specification or during prosecution.” *Thorner v. Sony Comput. Entm’t Am. LLC*, 669 F.3d 1362, 1365 (Fed. Cir. 2012).

If an inventor acts as his or her own lexicographer, the definition must be set forth in the specification with reasonable clarity, deliberateness, and precision. *Renishaw PLC v. Marposs Societa’ per Azioni*, 158 F.3d 1243, 1249 (Fed. Cir. 1998). The disavowal, if any, can be effectuated by language in the specification or the prosecution history. *Poly-Am., L.P. v. API Indus., Inc.*, 839 F.3d 1131, 1136 (Fed. Cir. 2016).

Only those claim terms that are in controversy need to be construed, and only to the extent necessary to resolve the controversy. *Realtime Data, LLC v. Iancu*, 912 F.3d 1368 (Fed. Cir. 2019) (“The Board is required to construe ‘only those terms . . . that are in controversy, and only to the extent necessary to resolve the controversy.’” (quoting *Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999))).

Petitioner states that “no terms require construction because the claims encompass the prior-art mappings under any reasonable construction consistent with the *Phillips* standard.” Pet. 23 (citing Ex. 1004 ¶¶ 38–42).

On the current record, we agree with Petitioner’s contention, not disputed by Patent Owner, that it is not necessary to conduct express claim construction for any claim term.

D. Alleged Obviousness of Claims 1–3, 6–9, and 12 over Waibel (incorporating by reference Suhm, Vo, and Manke), and Roberts (Petitioner’s Ground 1) and Roberts and Alleged Obviousness Claims 1–3, 6–9, and 12 over Waibel, Suhm, Vo, Manke, and Roberts (Petitioner’s Ground 2)

1. Ground 1 (claims 1–3, 6–9, 12 over Waibel and Roberts)

For its Ground 1, alleged obviousness over Waibel and Roberts, Petitioner asserts that Waibel incorporates by reference each of Suhm, Vo, and Manke. Pet. 7–8, 23. For reasons discussed below, we disagree that any of Suhm, Vo, and Manke has been incorporated by reference into Waibel. Because the basis of Ground 1 is merely Waibel and Roberts, and because Petitioner’s explanation of obviousness in Ground 1 relies on and applies teachings from Suhm, Vo, and Manke (Pet. 23–47), Petitioner has not shown a reasonable likelihood that it would prevail in establishing obviousness of any of claims 1–3, 6–9, and 12 over Waibel and Roberts.

Regarding the alleged incorporation by reference of each of Suhm, Vo, and Manke into Waibel, Petitioner asserts:

In addition to its broad teachings, *Waibel* incorporates three papers: *Suhm*, *Vo*, and *Manke*. [Ex. 1005], 167-68, 170. All four arise from the same university lab, from around the same time period, and share a number of the same authors. Decl. ¶¶53-54. It was common practice in academia at the time to publish separate papers, with each describing a component of one multi-component system. *Id.*

Suhm, *Vo*, and *Manke* are identified in *Waibel* and explicitly identified as describing, in detail, *Waibel*’s speech recognizer, joint recognizer, and handwriting recognizer, respectively. *Waibel*, 167-68, 170. Because *Waibel* identifies the three papers in their entirety (not just portions of their disclosures), and where they can be found (*id.*, 170), such broad and unequivocal language sufficiently incorporates all three papers in their entirety. See *Paice LLC v. Ford Motor Co.*, 881 F.3d 894, 907 (Fed. Cir. 2018) (such “broad and

unambiguous” language “plainly sufficient” to incorporate patent in its entirety); *Harari v. Lee*, 656 F.3d 1331, 1335 (Fed. Cir. 2011) (incorporated in their entirety based on “broad and unequivocal language”). As such, everything in *Suhm*, *Vo*, and *Manke* is “effectively part of” Waibel, as if their disclosures were “explicitly contained therein.” *See Paice*, 881 F.3d at 906.

Pet. 7–8.

The circumstance here, however, is distinguishable from both *Paice* and *Harari*, where patent specifications incorporated material by reference in compliance with 37 C.F.R. § 1.57(c), which states:

(c) Except as provided in paragraph (a) or (b) of this section, an incorporation by reference must be set forth in the specification and must:

- (1) Express a clear intent to incorporate by reference by using the root words “incorporat(e)” and “reference” (e.g., “incorporate by reference”); and
- (2) Clearly identify the referenced patent, application, or publication.

37 C.F.R. § 1.57(c).

Here, Petitioner has identified no language in Waibel that satisfies the requirement of 37 C.F.R. § 1.57(c) (1), i.e., using root words “incorporat(e)” and “reference,” with respect to any one of *Suhm*, *Vo*, and *Manke*. Also, the words in Waibel referencing *Suhm*, *Vo*, and *Manke* are much more general and merely reflect referencing or identification without incorporation.

For example, Waibel states: “Our speech recognition system *is based on* the recognition front-end of the JANUS speech translation system [reference number of bibliography entry of *Suhm*] which is capable of processing speaker-independent, spontaneous speech.” Ex. 1005 § 2.1 (emphasis added). Waibel also states: “Our MS-TDNN-based handwriting

recognizer [reference number of bibliography entry of Manke] is capable of processing writer-independent, continuous (cursive) handwriting at a recognition rate of 94% on a 20,000-word vocabulary.” *Id.* Waibel further states: “Recognition output from the modality processors are parsed into partially filled frames that are merged together to produce the combined interpretation as described in [reference number of bibliography entry of Vo]. *Id.* § 2.2. None of such descriptions uses the root words “incorporat[e]” and “reference,” and we determine a reader would not discern a clear intent to incorporate the actual text of the referenced publications, literally, into Waibel, notwithstanding that Waibel gives guidance to readers to look to the referenced publications for additional content.

We recognize that 37 C.F.R. § 1.57(c) pertains to how to determine whether other material has been incorporated by reference into a patent specification, and that Waibel is not a patent but a published article. However, Petitioner articulates no reason that references to other material in non-patent publications should be treated differently than patent specifications, and we see none. Indeed, Petitioner even cites to *Paice* and *Harari*, both of which relate to incorporation by reference into a patent specification. We think a broader or looser standard for incorporation into published articles would lead to chaos and confusion, as all references to something described elsewhere being useful, beneficial, or applicable may be argued as incorporated by reference. Furthermore, the referenced article may discuss other subject matter unrelated to what is at issue in the referencing prior art. In that circumstance, it would make little sense to

deem the referenced article as having been incorporated by reference, without seeing or identifying express language to that effect.

For the foregoing reasons, none of Suhm, Vo, and Manke has been incorporated by reference into Waibel. Petitioner has not shown a reasonable likelihood that it would prevail in establishing obviousness of any of claims 1–3, 6–9, and 12 over Waibel and Roberts.

2. *Ground 2 (Waibel, Suhm, Vo, Manke, and Roberts)*

a. *Overview of Waibel (Ex. 1005)*

Waibel is a paper titled “Multimodal Interfaces for Multimedia Information Agents.” Ex. 1005, 167. Waibel includes the following abstract:

When human communicate they take advantage of a rich spectrum of cues. Some are verbal and acoustic. Some are non-verbal and non-acoustic. Signal processing technology has devoted much attention to the recognition of speech, as a single human communication signal. Most other complementary communication cues, however, remain unexplored and unused in human-computer interaction. In this paper we show that the addition of non-acoustic or non-verbal cues can significantly enhance robustness, flexibility, naturalness and performance of human-computer interaction. We demonstrate computer agents that use speech, gesture, handwriting, pointing, and spelling *jointly* for more robust, natural and flexible human-computer interaction in the various tasks of an information worker: information creation, access, manipulation or dissemination.

Id. at 167. Waibel further describes:

In this paper we describe information agents that use multimodal interfaces to access, manipulate, create and disseminate information. In particular, we seek effective methods by which human users can easily interact with a computer by speaking, pointing, drawing, spelling, etc. *We show how cross-modal cues can be used effectively to recover from errors or miscommunications, since confusions in one modality*

usually are unambiguous in another. We describe how multiple modalities can cooperate and complement each other in various phases of the preparation of multimedia documents. Moreover, the added flexibility and robustness significantly increases reliability and naturalness, and with it the usability and acceptance of information systems. We present QuickDoc as a prototypical application that combines our multimodal subsystems in a simple, but powerful way.

Id. § 1 (emphasis added).

Waibel implements recognizers and processing modules for various input modalities, including speech, pen input, and face tracking, and these modality processors constitute the basic components of all of Waibel's multimodal systems. Ex. 1005 § 2.1. Waibel's speech recognition subsystem is based on the recognition front end of JANUS speech translation system which is capable of processing speaker-independent, spontaneous speech. *Id.* § 1 (citing Suhm (Ex. 1008)). Waibel's approach to pen-based gesture recognition is to decompose pen strokes into sequences of basic shapes such as line, arc, arrow, circle, and cross, and the same gesture shape may mean different things depending on the surrounding context. *Id.* (citing Vo (Ex. 1009)). Waibel's handwriting recognizer is capable of processing writer-independent, continuous (cursive) handwriting at a recognition rate of 94% on a 20,000-word vocabulary, and employs simple heuristics to decide when to invoke handwriting recognition on pen input, such as when the gesture recognizer cannot identify the input strokes as basic shapes. *Id.* (citing Manke (Ex. 1010)).

Waibel further describes that in order to make sense of input from all available sources, it needs a multimodal interpreter capable of producing an interpretation of user intent from the output of the modality processors.

Ex. 1005 § 2.2. Waibel explains that recognition outputs from the modality

processors are parsed into partially filled frames that are merged together to produce the combined interpretation. *Id.* (citing Vo).

Section 3.1 of Waibel is titled “Multimodal Interactive Error Repair and reads as follows:

Our approach is to have the user collaborate with the system in recovering from interpretation errors. First, errors have to be identified, either initiated by the system, for instance based on some confidence measure, or by the user. If a graphical user interface (GUI) is available, the user can simply highlight erroneous words (the reparandum) in the recognition hypothesis displayed. For error correction, the user provides additional input, choosing among different correction methods: repair by repeating the reparandum by respeaking, spelling out loud, or handwriting; repair by paraphrasing the reparandum; repair by pen gestures (e.g. to delete or insert), in addition to the standard repair by typing or selecting among N-best alternatives.

The rationale for this multimodal approach to error recovery is twofold. First, it exploits the fact that different input modalities are orthogonal: words which are confusable in one modality, can be disambiguated in a different modality (e.g. “road” and “rote” spoken versus spelled). Secondly, recent studies show that switching modality after system misinterpretation alleviates user frustration.

Ex. 1005 § 3.1 (citations omitted).

b. Overview of Roberts (Ex. 1007)

Roberts relates to training models for use in speech recognition.

Ex. 1007, 1:6–8. Roberts describes that its speech recognition system improves the speed and accuracy with which users can dictate desired text, and that the system does so by making it easier to correct errors and enter new words into the vocabulary. *Id.* at 37:18–22. Roberts describes:

[A] continuous speech recognition system is provided which indicates one or more best scoring word sequences selected by a recognition means. When a plurality of best scoring sequences

are indicated for a given portion of speech to be recognized, the system provides means for enabling the user to select one of the indicated best matching word sequences for use as an output, including means for that selection by voice. The system also enables the user to select an individual word from a given indicated best matching word sequence, and to correct that word sequence by correcting the selected word. *Preferably the word to be corrected can be selected by voice. Once the word has been selected for correction, the system provides a choice of alternate words to replace the word to be corrected by performing speech recognition against the speech corresponding to the word to be corrected.* The user can restrict the choice of alternate words by entering filtering information specifying a subset of words to which the desired word belongs. After the word to be corrected is replaced, the system uses the replaced word to help recognize the rest of the portion of speech.

Id. at 4:7–30 (emphasis added).

c. Overview of Suhm (Exhibit 1008)

Suhm is a paper titled “JANUS: TOWARDS MULTILINGUAL SPOKEN LANGUAGE TRANSLATION.” Ex. 1008, 8. It refers to a speech recognition system named “JANUS” which is still undergoing development. *Id.* § 1. Suhm identifies a previous version of JANUS as “JANUS-1” and the current and still developing version of JANUS as “JANUS-2.” *Id.* JANUS-2 in its current state can translate English and German spoken input into English, German, Spanish, Japanese, or Korean output. *Id.* Suhm states: “This paper reports on the current status of the system and ongoing efforts to extend and improve the recognition component. Then, we describe our new approach to robust translation of spoken language.” *Id.*

As system overview, Suhm describes:

The main system modules are speech recognition, parsing, discourse processing, and generation. Each module is language-

independent in the sense that it consists of a general processor that applies independently specified knowledge about different languages.

The recognition module decodes the speech in the source language into a list of sentence candidates, represented either as a word lattice or Nbest list. At the core of the machine translation components is a language independent representation of the meaning, which is extracted from the recognizer output by the parsing module. As last step, the final language independent representation is sent to the generator to be translated in any of the target languages. Figure 1 shows the system architecture.

After parsing, a discourse processor can be used to put the current utterance in the context of previous utterances, opening possibilities to integrate the speech and natural language processing components of the system to resolve parsing ambiguities and dynamically adapt the vocabulary and language model of the recognizer based on the current discourse state.

Ex. 1008 § 2.2.

Suhm describes that it is experimenting with “TDNN,” “MS-TDNN,” “MLP,” “LVQ,” and “HMM’s” for acoustic modeling, and “n-grams, word clustering, and automatic phase detection” for language modeling. Ex. 1008 § 2.2. Suhm further describes that the baseline JANUS-2 recognizer uses “LVQ-2 or phonetically tied SCHMM” for acoustic modeling, and “standard word bigram language model” for a decoder with word lattice output. *Id.*

§ 2.3. Suhm states: “We describe efforts to improve the recognition component along its major knowledge sources acoustic models, dictionary and language models.” *Id.* § 3 (citations omitted).

For dictionary learning, Suhm describes: “Modifications of the dictionary should not lead to higher phonetic confusability after retraining. Therefore we have proposed a data-driven approach to improve existing

dictionaries and automatically add new words and pronunciation variants whenever needed.” Ex. 1008 § 3.2 (citation omitted).

Section 3.3 of Suhm is titled “Morpheme Based Language Models.”

Suhm describes:

Based on our scheduling databases we noticed that in morphologically rich languages such as German and Spanish, dictionaries grow much faster with increasing database size, compared to English (cf. figure 2). This is due to the large number of inflections and compound words. One way to limit this growth with increasing dictionary sizes is to use other base units than words.

Ex. 1008 § 3.3. Suhm applied and compared three different decomposition methods: (1) strictly *morpheme* based decomposition; (2) decomposition in *root forms*; and (3) composition of strictly morpheme based decomposition and root forms. *Id.* Suhm determined that all three decomposition methods “significantly reduce vocabulary size and perplexity.” *Id.*

d. Overview of Vo (Exhibit 1009)

Vo is a paper titled “BUILDING AN APPLICATION FRAMEWORK FOR SPEECH AND PEN INPUT INTEGRATION IN MULTIMODAL LEARNING INTERFACES.” Ex. 1009, 3545. Vo states: “In this paper we present a multimodal interface capable of jointly interpreting speech, pen-based gestures, and handwriting in the context of an appointment scheduling application.” *Id.* at Abstr. In particular, “[a] multimodal interpretation engine jointly interprets information from all input sources by merging semantic frames. A domain-independent dialog processor maintains context information across input events.” *Id.* § 1.

Vo describes that individual modality processor components, i.e., speech, gesture, and handwriting recognizers, are separate modules that can

be replaced with ease. Ex. 1009 § 2. Vo states that its “speech recognition subsystem is based on the recognition front-end of the JANUS speech translation system.” *Id.* § 2.2. The “text string output” of the speech recognizer is then processed by “the Phoenix semantic parser” which “compiles a grammar specifying semantically meaningful fragments of text into an efficient concurrent transition network that identifies these fragments in the recognized utterance.” *Id.* § 3.1. Vo describes that the gesture/handwriting recognizers produce a sequence of gesture/handwriting components, and that the same gesture shape may mean different things depending in the context. *Id.* § 3.1. Vo further describes that in the present implementation, “gestures and handwritten words are encoded into frames according to simple hard-coded heuristics.” *Id.*

Vo explains:

The multimodal interpreter is responsible for producing an interpretation of user intent (i.e., a command to send to the calendar interface) from the output of the modality processors. We represent this interpretation as a frame consisting of slots specifying pieces of information such as the action to carry out or the date and time of a meeting. Recognition output from the modality processors are parsed into partially filled frames that are merged together to produce the combined interpretation.

Ex. 1009 § 3. Vo further explains:

Frames are merged by taking the union of the sets of values filling each slots and adding corresponding scores. Embedded frames are merged recursively. This produces an aggregate frame encoding all alternative interpretation hypotheses. The multimodal interpreter extracts these hypotheses in descending order of aggregate scores and sends the best ones to the dialog processor.

Id. § 3.2

If the dialog processor fails to find an acceptable hypothesis with a high enough score, it constructs an appropriate feedback message to guide the user. *Id.* § 3.3. As soon as a good hypothesis is found, the dialog processor sends it to the domain interface as the operation to carry out and starts a new dialog, and the user has the option of undoing the operation just performed if it is incorrect. *Id.*

e. Overview of Manke (Exhibit 1010)

Manke is a paper titled “The Use of Dynamic Writing Information in a Connectionist On-Line Cursive Handwriting Recognition System.” Ex. 1010, 1093. It describes a connectionist system NPen⁺⁺ for writer independent, large vocabulary on-line cursive handwriting recognition. *Id.* at Abstract. Manke states:

This system combines a robust input representation, which preserves the dynamic writing information, with a neural network architecture, a so called Multi-State Time Delay Neural Network (MS-TDNN), which integrates recognition and segmentation in a single framework. Our preprocessing transforms the original coordinate sequence into a (still temporal) sequence of feature vectors, which combine strictly local features, like curvature or writing direction, with a bitmap-like representation of the coordinate’s proximity.

Id.

Manke describes that its handwriting recognizer uses Hidden Markov Models (“HMM”) where “each character c_j itself is modelled by a three state hidden Markov model.” Ex. 1010, 1096. “The idea of using three states per character is to model explicitly the initial, middle and final section of the characters.” *Id.* Manke also describes that its handwriting recognizer, a so-called Multi-State Time Delay Neural Network (MS-TDNN), “integrates recognition and segmentation of words into a single network architecture.”

Id. § 1. “The MS-TDNN, which was originally proposed for continuous speech recognition tasks, combines shift-invariant, high accuracy pattern recognition capabilities of a TDNN with a non-linear alignment procedure for aligning strokes into character sequences.” *Id.* (citations omitted).

Manke states that the final goal is to integrate the speech recognizer JANUS with its proposed handwriting recognizer NPen⁺⁺ system into a multi-modal system. Ex. 1010, 1099.

f. Claim 1

(1) Preamble [1P]

The preamble of claim 1 recites: “A system for generating text responsive to message input of a first type and a second type, the system comprising.” We need not determine whether the preamble is limiting, because Petitioner sufficiently establishes that Waibel discloses a system for generating text responsive to message input of a first type and a second type.

Petitioner asserts that Waibel discloses a system including a speech recognizer receiving message input in the form of speech and a handwriting recognizer receiving message input in the form of handwriting. Pet. 24 (citing Ex. 1005, 167–168). Petitioner asserts that the system outputs an “HTML report,” which a person of ordinary skill in the art would understand includes text. *Id.* (citing Ex. 1005, 169–170; Ex. 1004 ¶ 74).

Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner. We determine that Waibel discloses the preamble recitation of claim 1.

(2) Limitation [1a]

Limitation [1a] recites:

(a) a unified message model including:

(1) a shared language model;

(2) a first model specific to a first type of message recognition; and

(3) a second model specific to a second type of message recognition.

Ex. 1001, 39:24–29.

Petitioner puts forth arguments and evidence asserting that Waibel, with teachings from Suhm, Vo, and Manke discloses *or* would have rendered obvious limitation [1a]. Pet. 25–33. We understand Petitioner to be arguing two positions, one based on incorporation by reference of Suhm, Vo, and Manke into Waibel such that Waibel itself discloses limitation [1a], and another based on Waibel in combination with teachings from Suhm, Vo, and Manke. Because we determined above that none of Suhm, Vo, and Manke is incorporated by reference into Waibel, we are not sufficiently persuaded that Waibel itself discloses limitation [1a]. However, Petitioner’s assertion that these limitations would have been rendered obvious by the combined teachings of Waibel, Suhm, Vo, and Manke, is supported by the cited evidence and not disputed by Patent Owner, as we discuss below. We are sufficiently persuaded that limitation [1a] would have been reasonably suggested by the combined teachings of Waibel, Suhm, Vo, and Manke.

The motivation to apply Suhm’s teaching in Waibel stems from this statement in Waibel: “Our speech recognition subsystem is based on the recognition front-end of the JANUS speech translation system [5] which is capable of processing speaker-independent, spontaneous speech.” Ex. 1005 § 2.1 (citation [5] refers to Suhm). The motivation to apply Vo’s teaching in Waibel stems from these statements in Waibel: “Our approach to pen-based gesture recognition is to decompose pen strokes into sequences of basic shapes such as line, arc, arrow, circle, cross . . . [6].” *Id.* (citation [6] refers

to Vo); “Recognition output from the modality processors are parsed into partially filled frames that are merged together to produce the combined interpretation as described in [6].” *Id.* § 2.2 (citation [6] refers to Vo). The motivation to apply Manke’s teachings in Waibel stems from this statement in Waibel: “Our MS-TDNN-based handwriting recognizer [7] is capable of processing writer-independent, continuous (cursive) handwriting at a recognition rate of 94% on a 20,000-word vocabulary.” *Id.* § 2.1 (citation [7] refers to Manke).

At the outset, the recitation of a unified message model is met if the applied prior art discloses a message model that “includes” the three individual components expressly required by the unified message model. That is the position taken by Petitioner and not disputed by Patent Owner.

As for the requirement of a “shared language model,” it is undisputed that language models are used in Waibel for all of its recognizers including the speech recognizer and the handwriting recognizer. The Petition specifically identifies the bigram language model described in Suhm as being used for Waibel’s speech recognizer. Pet. 26. The Petition also identifies the dictionary used in Manke’s handwriting recognition algorithm, used as Waibel’s handwriting algorithm, as a language model. *Id.* at 27–28. These assertions are supported by the cited evidence and not disputed by Patent Owner.

The ’405 patent states that a “shared language model” is “any language model that may be shared between two types of subordinate message recognizers included within a message recognizer.” Ex. 1001, 22:38–41. Petitioner asserts that one of ordinary skill in the art “would have understood the language model in Waibel’s speech and/or handwriting

recognizers is shared or would at least have found the use of a shared language model obvious.” Pet. 26 (citing Ex. 1004 ¶¶ 77–87). Petitioner explains that the sharing of a language model between different recognizers was “well-known at the time.” *Id.* (citing Ex. 1004 ¶ 77; Ex. 1007, 103, 111, 104, 180; Ex. 1018, 6:27–46; Ex. 1029, 296). These assertions are supported by the cited evidence and are also not disputed by Patent Owner.

Petitioner offers additional reasons why Waibel in light of Suhm, Vo, and Manke reasonably would have suggested to one of ordinary skill in the art sharing Waibel’s language models between different types of recognizers. *Id.* at 27–31. These explanations are supported by the cited evidence and not disputed by Patent Owner.

As for “a first model specific to a first type of message recognition,” Petitioner asserts that “*Waibel* discloses an acoustic model for speech recognition.” Pet. 32 (citing Ex. 1005, 167). With regard to speech recognition, Waibel describes: “The recognizer can be adapted to any task domain by retraining the language models and possibly tuning the acoustic models if the domain involves special vocabulary.” Ex. 1005, 167. Petitioner also notes that Suhm’s speech recognizer, usable in Waibel, makes use of an acoustic model using TDNN, MS-TDNN, MLP, LVQ, and HMM’s for acoustic modeling. Pet. 32 (citing Ex. 1008, 222; Ex. 1004 ¶ 88). Petitioner’s assertion is supported by the cited evidence and not disputed by Patent Owner.

As for “a second model specific to a second type of message recognition,” Petitioner explains that Manke discloses several handwriting-specific models, usable in Waibel’s handwriting recognizer, including a “15-dimensional feature vector,” the three-state hidden Markov model, a “Viterbi

approximation,” and scores that are “derived from the word models.” Pet. 32–33 (citing Ex. 1010, 1095–97, Fig. 1, Fig. 2; Ex. 1004 ¶ 89). Petitioner’s assertion is supported by the cited evidence and not disputed by Patent Owner.

(3) *Limitation [1b]*

Limitation [1b] recites:

(b) a first message recognizer, responsive to the first type of message input to provide text data in accordance with both the shared language model and the first model.

Ex. 1001, 39:30–33. We discussed above in the context of limitation [1a] how a “shared language model” is met by the applied prior art. Petitioner identifies Waibel’s speech recognizer together with Waibel’s multimodal interpreter as a first message recognizer, responsive to speech according to a shared language model and an acoustic model to provide text data output. Pet. 34 (citing Ex. 1005, 167, 169–170; Ex. 1005 ¶¶ 90–91; Ex. 1009, 3456). Petitioner’s stated position here with regard to limitation [1b] is supported by the cited evidence and not disputed by Patent Owner.

(4) *Limitation [1c]*

Limitation [1c] recites:

(c) a second message recognizer, responsive to the second type of message input to provide text data in accordance with both the shared language model and the second model; wherein

Ex. 1001, 39:34–37. We discussed above in the context of limitation [1a] how a “shared language model” is met by the applied prior art. Petitioner identifies Waibel’s handwriting recognizer together with Waibel’s multimodal interpreter as a second message recognizer, responsive to user’s handwriting according to a shared language model and a handwriting model

to provide text data output. Pet. 34–35 (citing Ex. 1005, 167, 169–170; Ex. 1004 ¶¶92–93; Ex. 1009, 3455). Petitioner’s stated position here with regard to limitation [1c] is supported by the cited evidence and not disputed by Patent Owner.

(5) *Limitation [1d]*

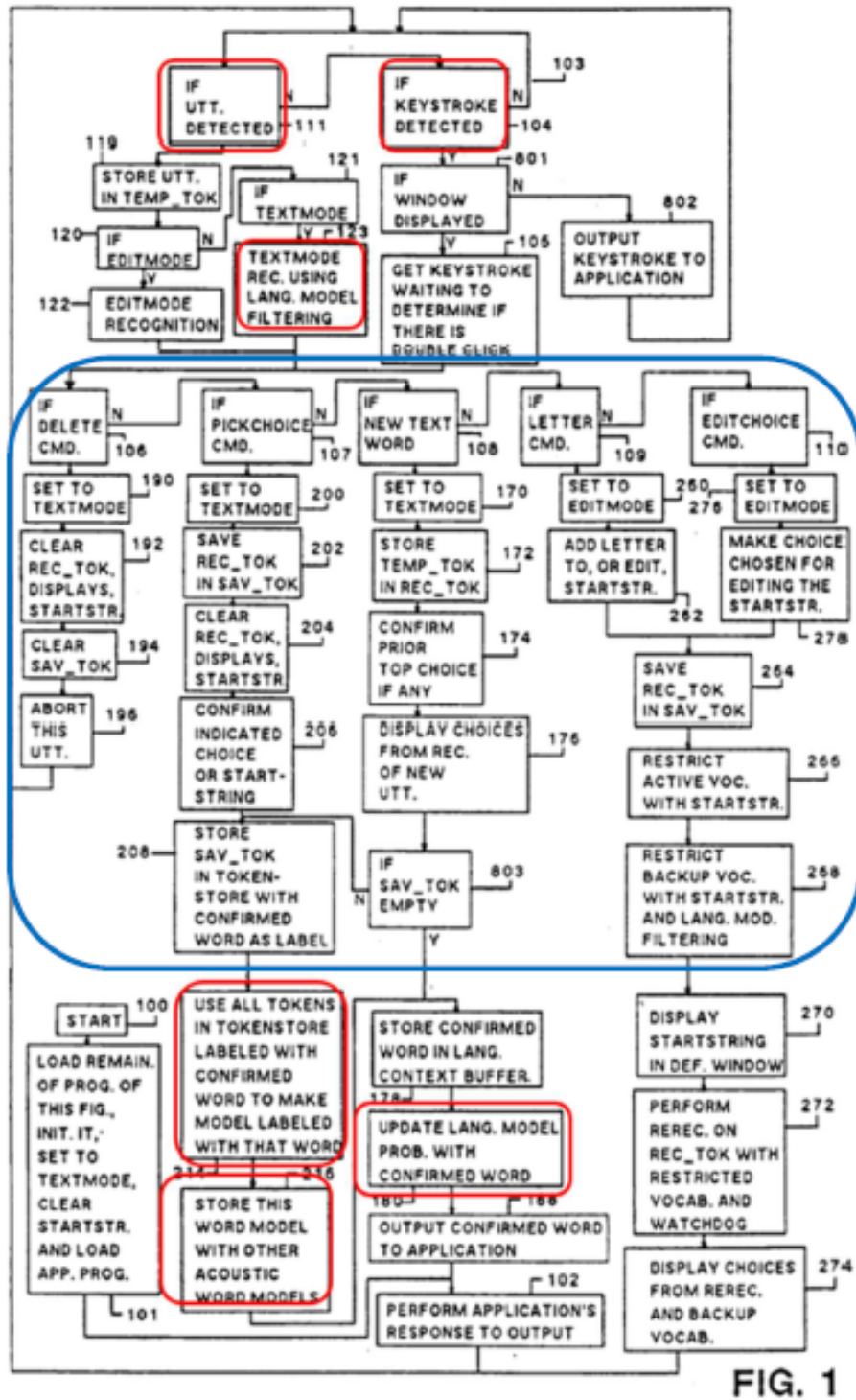
Limitation [1d] recites:

(d) the shared language model is trainable responsive to user correction of misrecognition by either of the first and second message recognizers, thereby improving accuracy of each of the first and second message recognizers.

Ex. 1001, 39:39–43.

We discussed above in the context of limitation [1a] how a “shared language model” is met by the applied prior art. According to Petitioner, having a “trainable” shared language model in Waibel would have been suggested by Roberts. Pet. 35. Petitioner asserts that “*Roberts* discloses a method for speech recognition and training of a language model in response to user correction of misrecognition by a first (speech) or second (keystroke input) message recognizer.” Pet. 35 (citing Ex. 1004 ¶¶ 96–97). In that regard, Petitioner further discusses the applicable portion of Roberts. *Id.* at 35–36 (citing Ex. 1007, code (57), 12:13–14:54, 16:11–17, 16:22–28, 15:58–16:28, 23:59–25:31. 27:30–33).

Petitioner produces an annotated version of Figure 1 of Roberts, reproduced below:



Pet. 37. Figure 1 of Roberts is a schematic block diagram of the functional steps of a discrete utterance speech recognition system according to Roberts. Ex. 1007, 5:1–3. Petitioner explains:

As shown [in Figure 1], *Roberts* discloses receiving speech and keystroke input (boxes 111 and 104) and using shared language model in speech recognition (box 123). *Roberts* also discloses training the shared model responsive to user correction of misrecognition (e.g., blue box) by the speech recognizer (e.g., boxes 214, 216, 180). *Id.*, 12:13-14:54 (describing error correction); Decl. ¶¶96-97; *Roberts*, 13:44-50 (“step 180 uses the confirmed word, if any, to update the language model”), 19:23-31, 7:55-56, 16:7-22 (“saving of tokens associated with confirmed words and the use of multiple tokens saved in conjunction with the confirmed word performed by the program of FIG. 1 provide an effective and efficient means of adaptive speech recognition.”), 29:55-30:53 (“[W]hen the user speaks a multiword phrase [and] the system fails to correctly recognize it, he has two major ways to get the correct phrase [F]irst is to enter all the characters of the phrase with letter commands, either by typing them or by speaking them The second . . . is to use the phrase edit routine.”), claim 48.

Pet. 38. Petitioner further notes that *Roberts* also discloses that its training language models improves the accuracy of the speech recognizer. *Id.* at 39 (citing Ex. 1007, 2:5-7, 2:17-22, code (57)). Regarding a motivation to combine the teachings of *Waibel* and *Roberts*, Petitioner explains:

A POSITA would have been motivated to use *Roberts*’s method of training a language model in *Waibel*’s system so that *Waibel* can similarly train its shared language model responsive to user correction of misrecognition in the speech or handwriting recognizer. Decl. ¶¶98-101. As explained in *Roberts*, this method allows a user to “begin productive work on large vocabulary, natural language dictation tasks with little or no advance training of word models” (*Roberts*, 2:8-12), “build acoustic models from tokens . . . collected during actual use of the system” (*id.*, 2:17-22), “easily correct errors” (*id.*, 2:23-25), and “develop a language model for use in speech recognition system . . . customized to his or her particular dictation, without requiring much additional effort to obtain such customization” (*id.*, 2:26-30). Accordingly, a POSITA would have appreciated

and expected these benefits using such a training method in *Waibel*'s system. Decl. ¶¶98-101.

Pet. 38–39. Petitioner's assertions are supported by the cited evidence and not disputed by Patent Owner.

(6) *Conclusion*

For the foregoing reasons, Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claim 1 over *Waibel*, *Suhm*, *Vo*, *Manke*, and *Roberts*. However, Petitioner has not shown a reasonable likelihood that it would prevail in establishing obviousness of claim 1 over *Waibel* and *Roberts* on a theory of incorporation by reference of *Suhm*, *Vo*, and *Manke* into *Waibel*.

g. *Claim 7*

Claim 7 is directed to a method but its steps substantially corresponds to the functionalities of the components recited in claim 1. Ex. 1001, 40:5–22. For claim 7's preamble, Petitioner relies on the arguments and evidence it provides for the preamble of claim 1 and limitation [1a]. Pet. 45. For limitation [7a], Petitioner relies on the arguments and evidence it provides for limitations [1a] and [1b]. *Id.* For limitation [7b], Petitioner relies on the arguments and evidence it provides for limitations [1a] and [1c]. *Id.* at 46. For limitation [7c], Petitioner relies on the arguments and evidence it provides for limitations [1a] and [1d]. *Id.*

Except for its assertions regarding alleged incorporation by reference of *Suhm*, *Vo*, and *Manke* into *Waibel*, Petitioner's assertions are supported by the cited evidence and not disputed by Patent Owner. Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claim 7 over *Waibel*, *Suhm*, *Vo*, *Manke*, and *Roberts*. However, Petitioner has not shown a reasonable likelihood that it would

prevail in establishing obviousness of claim 7 over Waibel and Roberts on a theory of incorporation by reference of Suhm, Vo, and Manke into Waibel.

h. Claims 2, 3, 8, and 9

Claim 2 depends from claim 1 and claim 3 depends from claim 2. Ex. 1001, 39:44–58. Claim 8 depends from claim 7 and claim 9 depends from claim 8. Ex. 1001, 40:23–37.

For claims 2 and 8, Petitioner puts forth argument and evidence asserting that Waibel discloses the additional limitations of claims 2 and 8. Pet. 41–42, 46–47. The assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Waibel itself discloses the additional features recited in claims 2 and 8 relative to the independent claims on which they depend.

For claims 3 and 9, Petitioner puts forth arguments and evidence asserting that Waibel discloses or would have rendered obvious the additional limitations of claims 3 and 9. Pet. 42–43, 47. Alternatively, Petitioner puts forth arguments and evidence asserting that a proper combination of Waibel and Roberts teaches the additional limitations of claims 3 and 9. *Id.* Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Waibel itself discloses the additional features recited in claims 3 and 9 relative to claims 2 and 8 from which they respectively depend. We are also sufficiently persuaded that a proper combination of Waibel and Roberts teaches the additional limitations of claims 3 and 9.

Accordingly, Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claims 2, 3, 8, and 9 over Waibel, Suhm, Vo, Manke, and Roberts. However, Petitioner has not shown

a reasonable likelihood that it would prevail in establishing obviousness of claims 2, 3, 8, and 9 over Waibel and Roberts on a theory of incorporation by reference of Suhm, Vo, and Manke into Waibel.

i. Claims 6 and 12

Claim 6 depends from claim 1 and further recites: “wherein the shared language model comprises a syntactic model and a semantic model.” Ex. 1001, 40:3–4. Claim 12 depends from claim 7 and further recites: “wherein performing message recognition in accordance with the shared language model comprises performing message recognition in accordance with a syntactic model and a semantic model.” *Id.* at 40:47–50.

Petitioner puts forth arguments and evidence asserting that Waibel, with teachings from Suhm and from Vo, discloses *or* would have rendered obvious the additional limitations of claims 6 and 12. Pet. 44–45, 47. We understand Petitioner to be arguing two positions, one based on incorporation by reference of Suhm and Vo into Waibel such that Waibel itself discloses the additional limitations of claims 6 and 12, and another based on Waibel in combination with teachings from Suhm and Vo.

Because we determined above that neither Suhm nor Vo is incorporated by reference into Waibel, we are not sufficiently persuaded that Waibel itself discloses the limitations added by claims 6 and 12. However, Petitioner’s assertion that these limitations would have been rendered obvious by the combined teachings of Waibel, Suhm, and Vo is supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that the limitations added by claims 6 and 12 would have been reasonably suggested by the combined teachings of Waibel, Suhm, and Vo.

Specifically, Suhm discloses a standard word bigram language model usable in Waibel's speech recognizer. Pet. 44; Ex. 1005 § 2.1; Ex. 1008 § 2.3. Petitioner establishes that the bigram language model is a syntactic model. Pet. 44 (citing Ex. 1004 ¶ 108). In its discussion of limitation [1a], Petitioner sufficiently establishes that Vo's dialog frames in its multimodal interpreter, which is identified in Waibel as usable in Waibel (Ex. 1005 § 2.2), is a shared language model. *Id.* at 30 (citing Ex. 1004 ¶¶ 77–87). In that regard, Dr. Schmandt also testifies that “frames were a well-known natural language processing construct reflecting both syntax and semantics.” Ex. 1004 ¶ 86 (cited at Pet. 30). In that regard, Vo states: “The text string output from the speech recognizer is processed by the Phoenix semantic parser developed by Ward.” Ex. 1009, 3546 (citation omitted).

For the foregoing reasons, Petitioner sufficiently establishes that Waibel in combination with Suhm and Vo discloses a shared language model including a syntactic model component and a semantic model component. Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claims 6 and 12 over Waibel, Suhm, Vo, Manke, and Roberts. However, Petitioner has not shown a reasonable likelihood that it would prevail in establishing obviousness of claims 6 and 12 over Waibel and Roberts on a theory of incorporation by reference of Suhm, Vo, and Manke into Waibel.

3. *Ground 3 (claims 4, 5, 10, 11 over Waibel, Roberts, Verrier)*

Claim 4 depends from claim 3. Ex. 1001, 39:59–67. Claim 5 depends from claim 4. *Id.* at 40:1–2. Claim 10 depends from claim 9. *Id.* at 40:38–44. Claim 11 depends from claim 10. *Id.* at 40:45–46. For its Ground 3, alleged obviousness of claims 4, 5, 10, and 11 over Waibel and Roberts,

Petitioner asserts that Waibel incorporates by reference each of Suhm, Vo, and Manke. Pet. 48, 49, 53. We determined above in Section III.D.1 that none of Suhm, Vo, and Manke has been incorporated by reference into the body of Waibel. Because the basis of Ground 3 is merely Waibel and Roberts, and because Petitioner’s explanation of obviousness in Ground 3 relies on and applies teachings from Suhm, Vo, and Manke (Pet. 24–41, 48, 49, 53), Petitioner has not shown a reasonable likelihood that it would prevail in establishing obviousness of any of claims 4, 5, 10, and 11 over Waibel, Roberts, and Verrier.

4. *Ground 4 (claims 4, 5, 10, 11 over Waibel, Suhm, Vo, Manke, Roberts, Verrier)*

a. *Overview of Verrier (Exhibit 1011)*

Verrier discloses a voice communication feature of an untethered personal stylus used with a digitizing display in a pen-based computer system. Ex. 1011, code (57). “A microphone and speaker are included in an untethered stylus in conjunction with a mechanical contact and radiative pickup mechanisms, to enable multimedia communication between the user and the pen-based computer system, through a digitizing display.” *Id.*

Verrier’s stylus in a pen-based computer system includes a contact detection mechanism such as a pressure transducer connected to a multiplexer inside the stylus, and a position detector such as a radiative pickup mechanism and signal strength detector also connected to the multiplexer inside the stylus. Ex. 1011, 4:8–17. A multiplexed signal including information from both inputs is then transmitted by the stylus to a digitizing display. *Id.* at 4:17–21. Verrier’s stylus also accommodates voice communication. *Id.* at 4:22–24. The multiplexer in the stylus has an input connecting to a microphone in the stylus and an output connecting to a

speaker in the stylus. *Id.* at 4:24–36. Audio signal from the microphone, through the multiplexer, may be transmitted to the digitizing display or the pen-based computer system. *Id.* at 4:27–31. Audio output, through the multiplexer, may be produced by the speaker in the stylus. *Id.* at 4:31–36.

Figure 2 of Verrier is reproduced below:

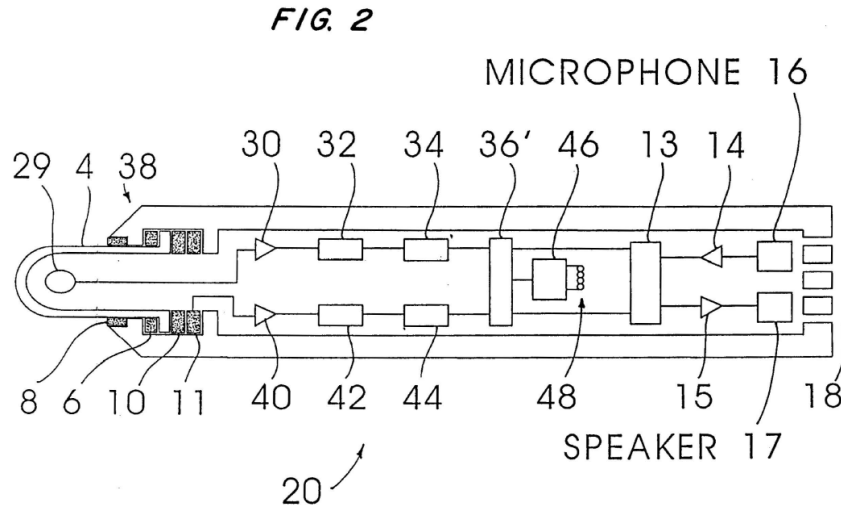


Figure 2 is a side view of an example of Verrier's stylus. Ex. 1011, 6:5–9.

Stylus 20 includes mechanical contacting mechanism 38 comprising pressure transducer 10. Ex. 1011, 7:27–32. When tip 4 of the stylus is pressed, transducer 10 is compressed and its electrical resistance changes. *Id.* at 7:47–57. Stylus 20 also includes pickup antenna 29. *Id.* at 8:4–7. The output of pickup antenna 29 is coupled through amplifier 30 to signal strength detector 32. *Id.* at 8:12–14. The output of signal strength detector 32 is applied to analog-to-digital converter 34. *Id.* at 8:24–25. The output of analog-to-digital converter 34 is a digital representation of the position of the stylus over a tablet. *Id.* at 8:24–29. Figure 2 also shows that stylus 20 includes microphone 16 and speaker 17.

b. Claims 4 and 10

Claim 4 recites:

4. The system of claim 3 wherein:
 - (a) the stylus includes appointing interface and a microphone;
 - (b) the system further comprises a tablet responsive to the pointing interface;
 - (c) the voice input is from the microphone; and
 - (d) the freehand input is from the pointing interface being manually manipulated on the tablet.

Ex. 1001, 39:59–67.

Claim 10 recites:

10. The method of claim 9 wherein:
 - (a) performing speech recognition responsive to voice input includes responding to voice input from a microphone in the stylus; and
 - (b) performing handwriting recognition responsive to freehand input includes responding to manual manipulation of the stylus on a tablet.

Id. at 40:38–44.

A stylus is already accounted for by Petitioner in analyzing claim 3 and 9. Pet. 42–43, 47. Petitioner asserts that one of ordinary skill in the art would have understood that such a stylus would include a pointing interface to interact with the computer receiving the input. *Id.* at 48 (citing Ex. 1003 ¶¶ 116–121). With regard to the added feature of “the stylus includes . . . a microphone” and “voice input is from the microphone,” Petitioner asserts that these limitations were well-known stylus features at the time. *Id.* at 49 (citing Ex. 1003 ¶¶ 116–121). Petitioner asserts:

Indeed, *Verrier* discloses a stylus with a microphone for receiving voice input. *Verrier*, Abstract (noting “[a] microphone and speaker are included in an untethered stylus . . . to enable multimedia communication between the user and the pen-based computer system, through a digitizing display”), 18:35–40, Fig. 2.

Id. Petitioner further explains:

It would have been obvious to a POSITA [one of ordinary skill in the art] to incorporate a microphone in *Waibel’s* pen and receive voice input at the microphone for the acquisition of high-quality speech in a pen-based system to accommodate voice communications at the user-stylus interface. *Verrier*, 3:27–35; Decl. ¶¶116-21. Indeed, a POSITA would have known that placing the microphone in the stylus improves the quality of the audio signal due to the user’s proximity to the stylus while dictating regardless of the location of the computer device. Decl. ¶¶116-21. And a POSITA would have been so motivated to minimize overall hardware and power consumption while increasing portability. The ’405 patent admits “[i]t has long been known to incorporate a microphone into a stylus for the acquisition of high-quality speech in a pen-based” system. ’405 patent, 2:54-64. A POSITA would have known incorporating such a microphone into *Waibel’s* stylus would have been a simple substitution of known elements according to known functions to achieve predictable results. *Id.* A POSITA would have had a reasonable expectation of success in doing so. *Id.*

Id. at 50.

Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that the combined teachings of the applied prior art reasonably would have suggested the added features of the stylus including a microphone and the voice input being from the microphone.

With regard to the added feature of the tablet being responsive to the pointing interface, Petitioner notes that the ’405 patent states

that a tablet can be “any suitable hardware and/or software for performing functions of, inter alia, a speech recognizer and an editor.”

Pet. 50 (citing Ex. 1001, 6:11–14). Petitioner asserts that Waibel discloses such a tablet. *Id.* at 50–51 (citing Ex. 1004 ¶¶ 116–121).

Petitioner explains:

Further, *Waibel* teaches its tablet is responsive to the pointing interface of the stylus. *Waibel*, 167. *Waibel* discloses uses can interact with the computer by “speaking, pointing, drawing, spelling,” and the handwriting recognizer can “invoke handwriting recognition on pen input” into the computer. *Id.*

Id. at 51. Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Waibel’s tablet is responsive to a pointing interface.

With regard to the freehand input for handwriting recognition being from manual manipulation of the stylus (claim 10) or from manual manipulation of the pointing interface (claim 4), Petitioner asserts: “*Waibel* discloses using a pen for providing handwriting input on a tablet, as discussed above. A POSITA would have understood handwriting input includes freehand input from a pointing interface of the pen when it is manually manipulated on the tablet.” Pet. 51 (citing Ex. 1002 ¶¶ 116–21). Petitioner’s assertion is supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Waibel discloses the freehand input for handwriting recognition being from manual manipulation of the stylus (claim 10) or from manual manipulation of the pointing interface (claim 4).

Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claims 4 and 10 over Waibel, Suhm, Vo, Manke, Roberts, and Verrier.

c. Claims 5 and 11

Claim 5 depends from claim 4 and further recites: “wherein the visual indicia of dictated text appears on the tablet.” Ex. 1001, 40:1–2. Claim 11 depends from claim 10 and adds the same to claim 10. *Id.* at 40:45–46.

Petitioner asserts that Waibel discloses the use of a tablet as a human-computer interface and display of visual indicia of dictated text on that tablet. Pet. 51 (citing Ex. 1005, 167–168; Ex. 1003 ¶¶ 122–24). Petitioner further asserts that one of ordinary skill in the art would have understood that the display for Waibel’s navigation system responds to the pen and displays the dictated text. Pet. 52 (citing Ex. 1003 ¶¶ 122–24). Petitioner alternatively asserts:

At the very least, a POSITA would have found doing so [the display responding to the pen and displaying the dictated text] obvious. The fast and commonplace development of touch sensitive screens in portable devices, like tablets, would have made the integration of various functions like touch input and display on one integrated device obvious. Doing so would be desirable for user-portability, power efficiency, and manufacturing costs. *Id.* And a POSITA would have a reasonable expectation of successfully doing so. *Id.*

Pet. 52–53 (citing Ex. 1004 ¶¶ 122–124).

Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner at this stage of the proceeding. We are sufficiently persuaded that Waibel both discloses and also would have suggested the limitation added by claims 5 and 11.

Petitioner has shown a reasonable likelihood that it would establish obviousness of claims 5 and 11 over Waibel, Suhm, Vo, Manke, Roberts, and Verrier.

5. *Ground 5 (Claims 1–3 and 7–9 over Roth and Roberts)*

a. *Overview of Roth (Ex. 1006)*

Roth is a published European Patent Application titled “AUDIO-AUGMENTED HANDWRITING RECOGNITION.” Ex. 1006, code (54).

Roth relates to an apparatus for receiving handwriting input and for augmenting the handwritten input with audio or spoken input. Ex. 1006,

1:1–4. In summary, Roth describes:

In accordance with the invention there is provided an improved data-entry apparatus for entering handwritten data. The user enters a speech portion (a word, a part of a word, or multi-word phrase) by writing it on a tablet and speaking it into a microphone. Features are extracted from the spoken information. The touchpad data are indicative of any of a number of candidate words, and a multiplicity of speech portion templates are generated, each template indicative of one of the candidate words. The apparatus evaluates the correlation between the extracted features and the features of each generated speech portion template, and the speech portion template having the highest correlation with the extracted features determines the recognized word. Optionally, a speech synthesizer speaks the word, or the recognition result is provided to the user on a visual display, and an opportunity is provided so the user may confirm the correctness of the recognition.

Ex. 1006, 2:33–3:15. Figure 1 of Roth is reproduced below:

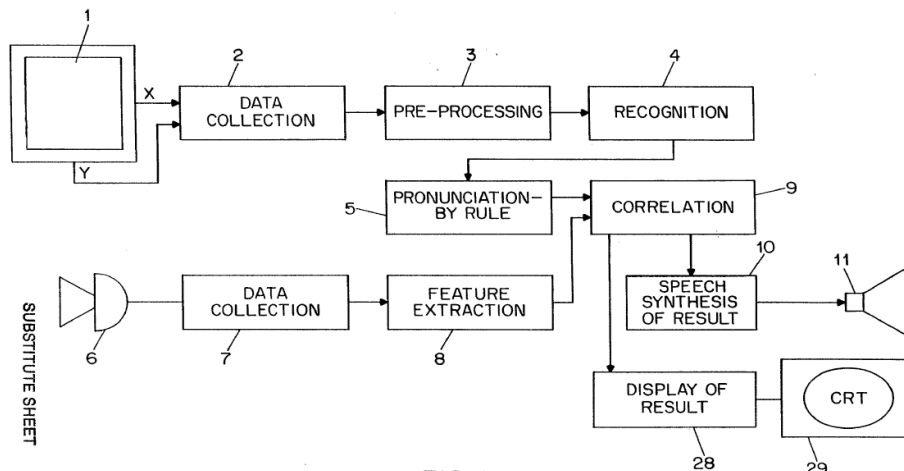


FIG. 1

Figure 1 is a functional block diagram for an embodiment of Roth's system. *Id.* at 4:4–6.

As shown in Figure 1, handwritten input is provided by the user at touchpad 1, and data regarding stylus position when providing handwritten input is collected at block 2 as a function of time, preferably at least as often as 100 samples per second. Ex. 1006, 4:12–16. “Once the data are collected, the data points are configured as indicative of the strokes making up the character to be recognized.” *Id.* at 4:29–31. Then, various preprocessing such as smoothing out of strokes and correcting wild points, etc., are performed at block 3. *Id.* at 4:31–5:5.

At block 4, character recognition is performed and a number of possible matches are produced. Ex. 1006, 5:35–6:2. Roth describes that where letters of an alphabet are being recognized, “the number of candidates can be half a dozen or so.” *Id.* at 6:2–4. At block 5, a pronunciation for each candidate is produced through pronunciation-by-rule methods. *Id.* at 6:8–10. “At the same time, the user is asked to speak the character into a microphone [6].” *Id.* at 6:23–24. Features are extracted from the spoken input by known means. *Id.* at 6:29–31. Then, at block 9, “[t]he extracted features of the spoken information are compared one by one with the features of the candidates [produced by handwriting input] mentioned before.” *Id.* at 6:33–7:4.

Roth describes that optionally each feature match is followed by audio synthesis of the matched character. Ex. 1006, 8:3–5. Roth states:

The synthesized speech is replayed for the user, and assuming the user agrees with the match, the process continues with the next handwriting to be recognized. In the (hopefully rare) case that the match was incorrect, the user will so indicate, either by touchpad input or by speech input to the system.

Instead of or in addition to an audio synthesis of the match character, there may be a visual display of the match character as indicate by block 28. Assuming the user agrees with the match, the process continues with the next handwriting to be recognized. As with the audio synthesis technique just described, in the (hopefully rare) case that the match was incorrect, the user will so indicate, either by touchpad input or by speech input to the system.

Id. at 8:3–19.

Roth describes that its system “may be optionally embodied with the opposite sequence, namely (1) analyzing the spoken input to derive a number of candidate speech portions (e.g. words) and (2) receiving written input to disambiguate (e.g. narrow down the candidates to one result).” Ex. 1006, 11:16–21. Roth additionally describes that “[a] system could, in fact, apply both approaches top a set of inputs, and choose the result with the higher degree of confidence.” *Id.* at 11:30–32.

b. Claim 1

(1) Preamble [1P]

The preamble of claim 1 recites: “A system for generating text responsive to message input of a first type and a second type, the system comprising.” We need not determine whether the preamble is limiting, because Petitioner sufficiently establishes that Roth discloses a system for generating text responsive to message input of a first type and a second type.

Petitioner explains that in response to speech and handwriting input Roth’s system generates a visual display of the matched character. Pet. 54 (citing Ex. 1006, 6:26–29, 8:3–19, Fig. 1). Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner. We determine that Roth discloses the preamble recitation of claim 1.

(2) *Limitation [1a]*

Limitation [1a] recites:

(a) a unified message model including:

- (1) a shared language model;
- (2) a first model specific to a first type of message recognition; and
- (3) a second model specific to a second type of message recognition;

Ex. 1001, 39:24–29.

Petitioner asserts that Roth discloses a “unified message model” by using both handwriting input (boxes 2–4 in Figure 1) and speech input (boxes 7–8 in Figure 1) to generate a text match (box 9 in Figure 1). Pet. 55 (citing Ex. 1004 ¶¶ 129, 130; Ex. 1006, Fig. 1). Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner. We determine that Roth discloses a “unified message model.”

Regarding a shared language model, Petitioner explains:

Roth discloses a shared language model in elements 5 and 9 (pronunciation-by-rule algorithm and correlation boxes), where *Roth* performs correlations and rankings using thresholds and margins based on features extracted from speech and handwriting. Decl. ¶¶ 131-32. Specifically, *Roth* discloses a language model including a pronunciation-by-rule algorithm that determines the pronunciation of a word by a rule, followed by a correlation of that written word to its pronunciation as uttered by a user. *Id.*; *Roth*, 3:22-24, 6:8-19, 10:13-25. A POSITA would have understood such a pronunciation-by-rule algorithm is a shared language model, for instead of being used only for handwriting or speech, it is used as a higher-level model to measure the similarity of words input by both the speech and handwriting recognizers. Decl. ¶¶ 131-32. This is performed independent of the underlying acoustic or pen stroke data, and before determining a best match from the shared input of word

hypotheses independently produced by the speech and pen recognizers. *Id.*

Pet. 55–56. Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Roth discloses a shared language model within the unified message model,

Regarding “a first model specific to a first type of message recognition,” Petitioner asserts that Roth discloses this limitation. Pet. 57. Petitioner explains:

Roth discloses using speech models based on American and British templates, as well as men and women templates. *Roth*, 6:19-22. *Roth* also teaches extraction of features from speech input, like a “determination of the amount of speech energy within various frequency ranges as a function of time.” *Id.*, 6:23-31. A POSITA would have understood the purpose of such extracted features would be for correlation with a speech model. *Id.*, 6:29-31; Decl. ¶134.

Id. at 57. Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Roth reasonably would have suggested “a first model specific to a first type of message recognition.” In the context of Roth, that first type of message recognition is speech recognition.

Regarding “a second model specific to a second type of message recognition,” Petitioner asserts that Roth either discloses or would have rendered obvious this limitation. Pet. 57 (citing Ex. 1004 § 135). Petitioner explains:

Roth discusses handwriting models used in existing handwriting recognizers. *Roth* 1:18-29, 2:9-15. Because it does not suggest the models in its handwriting recognizer deviate from existing models and does not otherwise provide detail on the models in its handwriting recognizer (element 2 (data collection), element 3 (handwriting pre-processing), element 4 (handwriting

recognition)), a POSITA would have understood similar handwriting models are used in the handwriting recognizer of *Roth* to decipher handwriting. *See, e.g., id.*, 4:12-6:22. Indeed, using handwriting models in recognizers was well known. '405 patent, 6:22-34 (citing *Beemink*, U.S. Patent No. 5,903,668, 10:11-32); Decl. ¶ 135.

Id. at 57–58. Petitioner's assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Roth reasonably would have suggested “a second model specific to a second type of message recognition.” In the context of Roth, that second type of message recognition is handwriting recognition.

(3) *Limitation [1b]*

Limitation [1b] recites:

(b) a first message recognizer, responsive to the first type of message input to provide text data in accordance with both the shared language model and the first model;

Ex. 1001, 39:30–33. We discussed above in the context of limitation [1a] how Roth discloses a “shared language model.” Petitioner asserts that Roth either discloses or would have rendered obvious limitation [1b]. Pet. 58 (citing Ex. 1004 ¶¶ 136–138). Petitioner explains:

Roth discloses a first message recognizer (element 6 (microphone), 7 (data collection), 8 (feature extraction), 9 (correlation)) responsive to a first type of message input, e.g., speech, to provide text data (element 28 (display)), in accordance with the shared language model (e.g., element 9 (correlation)), and first model (e.g., 7 (speech data collection) and 8 (speech feature extraction)). *Id.*; *Roth*, 6:23-31, 11:33-13:35, 6:33-8:2, 10:13-11:32, 14:26-17:2, 18:14-21.

Id.

Petitioner's assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Roth reasonably would have suggested limitation [1b].

(4) *Limitation [1c]*

Limitation [1c] recites:

(c) a second message recognizer, responsive to the second type of message input to provide text data in accordance with both the shared language model and the second model; wherein

Ex. 1001, 39:34–37. We discussed above in the context of limitation [1a] how a “shared language model” is met by the applied prior art. Petitioner asserts that Roth either discloses or would have rendered obvious limitation [1c]. Pet. 58 (citing Ex. 1004 ¶¶ 139–141). Petitioner explains:

Roth discloses a second message recognizer (elements 1 (touchpad), 2 (data collection), 3 (pre-processing), 4 (recognition), 5 (pronunciation by rule), and 9 (correlation)) responsive to a second type of message input, e.g., handwriting, to provide text data (element 28), in accordance with the shared language model (e.g., element 9 (correlation)), and a second model (e.g., element 2 (handwriting data collection), 3 (pre-processing), 4 (recognition), 5 (pronunciation by rule)). *Id.*; *Roth*, 4:29-5:5, 7:5-8:2, 10:13-11:32, 14:26-17:2, 18:14-21.

Id. at 58–59.

Petitioner's assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Roth reasonably would have suggested limitation [1c].

(5) *Limitation [1d]*

Limitation [1d] recites:

(d) the shared language model is trainable responsive to user correction of misrecognition by either of the first and

second message recognizers, thereby improving accuracy of each of the first and second message recognizers.

Ex. 1001, 39:39–43.

Petitioner asserts that Roth discloses or at least renders obvious limitation [1d]. Pet. 59. Alternatively, Petitioner asserts that Roth and Roberts in combination would have rendered obvious limitation [1d]. *Id.*

We have read the portion of the Petition addressing limitation [1d], i.e., pages 59–61, and do not find that Petitioner has presented reasoning as to why Roth itself discloses or would have rendered obvious limitation [1d]. All of the reasoning presented on these pages involve combined teachings of Roth and Roberts.

Regarding the disclosure of Roberts, Petitioner explains:

As explained with limitation [1d] in Grounds 1 and 2, *Roberts* discloses a shared language model trainable responsive to user correction of misrecognition by a first (e.g., speech) or a second message (e.g., keystroke) recognizer, and a “statistical language model” and “word models” from a phonemic dictionary that generates words to match the user’s dictation. *Roberts*, Abstract, 7:55-57, 12:7-22, 13:44-50, 16:11-17, 16:22-28, 19:23-31, 28:61-29:32, Fig. 1. When the user chooses the correct word by “keyboard or voice command,” the language model and acoustic model are updated. *Id.* This training improves the accuracy of the speech recognizer. *Id.*; see *supra* Section VI.A.1.e; Decl. ¶143.

Pet. 59. Petitioner also provides two separate reasons to combine the teachings of Roth and Roberts. Pet. 60–61.

First, Petitioner contends that one of ordinary skill in the art would have found it obvious “to use *Roth*’s method of user correction to train shared language models (feature templates in pronunciation-by-rule correlation method or the speech templates in the template store) to improve

the accuracy of its recognizers in view of *Roberts*'s method of updating language models through user correction." Pet. 60 (citing Ex. 1004 ¶¶ 152–153). Petitioner notes as an example that "when a user corrects a recognition output (*Roth*, 8:16-19), the system may use correction data to change pronunciation rules to account for differences in regional spelling or pronunciation. *Id.*" *Id.* These assertions are supported by the cited evidence and not disputed by Patent Owner.

Second, Petitioner asserts that one of ordinary skill in the art also would have found it obvious to use *Roberts*'s trainable bigram language model in *Roth*'s language model to achieve the same benefits. Pet. 60 (citing Ex. 1004 ¶¶ 144–151). Petitioner explains:

[B]igram models were known to effectively reduce errors in recognizers by predicting the likelihood that a second word follows the first word. *Id.*; *Roberts*, 13:44-60. Because this prediction is predicated on the accuracy of the first word, the bigram model would have worked well in *Roth*'s system because the first word was already recognized with a high degree of certainty by the pronunciation-by-rule algorithm and through a user's optional correction of a misrecognition. Decl. ¶¶ 144-51; *Roth*, 3:22-24, 6:8-19, 8:17-29, 10:13-25. This further reduction of errors would reduce the time it takes to provide a correct result. Decl. ¶¶ 144-51.

Pet. 60. Petitioner provides additional explanation on pages 60–61 of the Petition. These assertions of Petitioner are supported by the cited evidence and not disputed by Patent Owner.

For the foregoing reasons, we are sufficiently persuaded that *Roth* and *Roberts* in combination reasonably would have suggested limitation [1d].

(6) Conclusion

Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claim 1 over *Roth* and *Roberts*.

c. Claim 7

Claim 7 is directed to a method but its steps substantially corresponds to the functionalities of the components recited in claim 1. Ex. 1001, 40:5–22. For claim 7’s preamble, Petitioner relies on the arguments and evidence it provides for the preamble of claim 1 and limitation [1a]. Pet. 64. For limitation [7a], Petitioner relies on the arguments and evidence it provides for limitations [1a] and [1b]. *Id.* For limitation [7b], Petitioner relies on the arguments and evidence it provides for limitations [1a] and [1c]. *Id.* For limitation [7c], Petitioner relies on the arguments and evidence it provides for limitations [1a] and [1d]. *Id.*

Petitioner’s assertions are supported by the cited evidence and not disputed by Patent Owner. Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claim 7 over Roth and Roberts.

d. Claims 2, 3, 8, and 9

Claim 2 depends from claim 1 and claim 3 depends from claim 2. Ex. 1001, 39:44–58. Claim 8 depends from claim 7 and claim 9 depends from claim 8. Ex. 1001, 40:23–37.

For claims 2 and 8, Petitioner puts forth argument and evidence asserting that Roth discloses the additional limitations of claims 2 and 8. Pet. 41–42, 46–47. The assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Roth discloses the additional features recited in claims 2 and 8 relative to the independent claims on which they depend.

For claims 3 and 9, Petitioner puts forth arguments and evidence asserting that Roth would have rendered obvious the additional limitations

of claims 3 and 9. Pet. 62. Alternatively, Petitioner puts forth arguments and evidence asserting that a proper combination of Roth and Roberts teaches the additional limitations of claims 3 and 9. *Id.* at 63. Petitioner's assertions are supported by the cited evidence and not disputed by Patent Owner. We are sufficiently persuaded that Roth would have reasonably suggested the additional features recited in claims 3 and 9 relative to claims 2 and 8 from which they respectively depend. We are also sufficiently persuaded that a proper combination of Roth and Roberts teaches the additional limitations of claims 3 and 9.

Accordingly, Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claims 2, 3, 8, and 9 over Roth and Roberts.

6. *Ground 6 (claims 4, 5, 10, and 11 over Roth, Roberts, and Verrier)*
 - a. *Claims 4 and 10*

Claim 4 recites:

4. The system of claim 3 wherein:
 - (a) the stylus includes a pointing interface and a microphone;
 - (b) the system further comprises a tablet responsive to the pointing interface;
 - (c) the voice input is from the microphone; and
 - (d) the freehand input is from the pointing interface being manually manipulated on the tablet.

Ex. 1001, 39:59–67.

Claim 10 recites:

10. The method of claim 9 wherein:
 - (a) performing speech recognition responsive to voice input includes responding to voice input from a microphone in the stylus; and
 - (b) performing handwriting recognition responsive to freehand input includes responding to manual manipulation of the stylus on a tablet.

Ex. 1001, 40:38–44.

With regard to the features of a stylus including a pointing interface and freehand input being from the pointing interface being manually manipulated on a tablet, Petitioner asserts that Roth discloses using a stylus for providing freehand input on a touchpad.” Pet. 65 (citing Ex. 1006, 1:7–10). Petitioner also asserts that one of ordinary skill in the art would have understood that that disclosure means the freehand input is from the pointing interface of a stylus being manually manipulated on a tablet. *Id.* (citing Ex. 1004 ¶¶ 165–168).

With regard to the features of a microphone and the voice input being from the microphone, Petitioner asserts that incorporating a microphone in Roth’s stylus and providing voice input from the microphone would have been obvious to one of ordinary skill in the art in light of Verrier. Pet. 65–66. Petitioner notes that Verrier discloses a stylus with a microphone for receiving voice input. *Id.* at 66 (citing Ex. 1011, code (57), 18:35–40, Fig. 2). Petitioner explains:

It would have been obvious to a POSITA to incorporate a microphone in *Roth*’s pen and receive voice input from the microphone for the acquisition of high-quality speech in pen-based system to accommodate voice communications at the user-stylus interface. *Verrier*, 3:27–35; Decl. ¶¶165–68. Indeed, a POSITA would have known that placing the microphone in the

stylus improves the quality of the audio signal due to the user's proximity to the stylus while dictating, regardless of the location of the computer device. *Id.* Even the '405 patent admits "[i]t has long been known to incorporate a microphone into a stylus for the acquisition of high-quality speech in a pen-based" system. '405 patent, 2:54–64. A POSITA would have been motivated to do so to minimize hardware and power consumption, while increasing portability. A POSITA would have known incorporating a microphone into *Roth*'s stylus would have been the simple substitution of known elements according to known functions to achieve predictable results. Decl. ¶¶165–68. As such, a POSITA would have had a reasonable expectation of success doing so. *Id.*

Pet. 66–67. These assertions are supported by the cited evidence and not disputed by Patent Owner.

With regard to the requirement of a tablet responsive to the pointing interface, Petitioner asserts that Roth discloses this feature. Pet. 67 (citing Ex. 1006, 2:35–3:2, 8:30–33; Ex. 1004 ¶ 169). The assertion is supported by the cited evidence and not disputed by Patent Owner.

For the foregoing reasons, Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claims 4 and 10 over Roth, Roberts, and Verrier.

b. Claims 5 and 11

Claim 5 depends from claim 4 and further recites: "wherein the visual indicia of dictated text appears on the tablet." Ex. 1001, 40:1–2. Claim 11 depends from claim 10 and adds the same to claim 10. *Id.* at 40:45–46.

Petitioner asserts that Roth, Roberts, and Verrier would have rendered this additional limitation obvious. Pet. 67. Petitioner asserts that "*Roth* teaches the use of a tablet as a human-computer interface and the display of visual indicia of dictated text on that tablet." *Id.* at 68 (citing Ex. 1004

¶¶ 170–171). But then the explanation that follows is directed to how it would have been obvious to one of ordinary skill in the art to integrate touch input and display on one device. Also, Dr. Schmandt’s testimony is not that Roth discloses the use of a tablet as a human-computer interface and the display of visual indicia of dictated text *on that tablet*, but “*Roth discloses the use of a tablet as a human-computer interface and its display of the visual indicia of dictated text.*” Ex. 1004 ¶ 170.

Thus, Petitioner’s assertion that Roth teaches the use of a tablet as a human-computer interface and the display of visual indicia of dictated text *on that tablet* is not supported by the cited evidence. Nevertheless, we are sufficiently persuaded by the following reasoning provided by Petitioner and based on obviousness:

The fast and commonplace development of touch sensitive screens in portable devices (e.g., tablets) was well-known at that time, and it would have been obvious to a POSITA to integrate various functions, including touch input and display, into a single integrated device, rather than two devices, to improve user-portability, power efficiency, and manufacturing cost. Decl. ¶¶170-71. And a POSITA would have had a reasonable expectation of success in doing so.

Pet. 68 (citing Ex. 1004 ¶¶ 170–171).

For the foregoing reasons, Petitioner has shown a reasonable likelihood that it would prevail in establishing obviousness of claims 5 and 11 over Roth, Roberts, and Verrier.

7. *Ground 7 (Claims 6 and 12 over Roth, Roberts, and Vo)*

Claim 6 depends from claim 1 and further recites: “wherein the shared language model comprises a syntactic model and a semantic model.” Ex. 1001, 40:3–4. Claim 12 depends from claim 7 and further recites: “wherein performing message recognition in accordance with the shared

language model comprises performing message recognition in accordance with a syntactic model and a semantic model.” *Id.* at 40:47–50.

Petitioner asserts that “[a]lthough *Roth* and *Roberts* do not expressly disclose this model also comprises a ‘*semantic model*,’ using a semantic model with a syntactic model was well known in the art to improve recognition.” Pet. 69 (citing Ex. 1004 ¶ 174; Ex. 1026, 172-175). Petitioner further asserts that even the ’405 patent itself “admits the use of semantic and syntactic models in language models in recognizers was well known.” *Id.* (citing Ex. 1001, 26:11–21). None of the cited sources, however, support the assertion that using syntactic and semantic models at the same time was well known in the art. For instance, Dr. Schmandt only testifies that “the use of a semantic model was well known in the art to improve recognition.” Ex. 1004 ¶ 174. The cited portion of the Specification of the ’405 patent does not convey that using both syntactic and semantic model together was well known. Ex. 1001, 26:11–21.

Nonetheless, Petitioner does cite to Vo, and states:

Vo, for example, discloses a semantic model in a language model. Decl. ¶174; *Vo*, 3546 (“output from the speech recognizer is processed by the Phoenix semantic parser” which “compiles a grammar specifying semantically meaningful fragments of text into an efficient recurrent transition network that identifies these fragments in the recognized utterance”);

Pet. 69. In discussing limitation [1d] above, we already determined that it would have been obvious in light of *Roberts* to use *Roberts*’s trainable bigram language model in *Roth*’s language model. These assertions are supported by the cited evidence and not disputed by Patent Owner.

Petitioner has made a sufficient showing that in light of *Vo*’s teachings noted above, it would have been obvious to one of ordinary skill in the art to

add a semantic model to the bigram model usable in Roth.⁶ Accordingly, the added limitation of claims 6 and 12 reasonably would have been suggested by Roth, Roberts, and Vo.

For the foregoing reasons, Petitioner has shown a reasonable likelihood that it would establish claims 6 and 12 as obvious over Roth, Roberts, and Vo.

IV. CONCLUSION

For the foregoing reasons, we conclude that the information presented in the Petition establishes a reasonable likelihood that Petitioner would prevail in establishing that at least one of claims 1–12 of the '405 patent is unpatentable.⁷

V. ORDER

In consideration of the foregoing, it is

ORDERED that *inter partes* review of all challenged claims, i.e., claims 1–12 of the '405 patent, is instituted on all grounds stated in the Petition; and

FURTHER ORDERED that pursuant to 35 U.S.C. § 314(c) and 37 C.F.R. § 42.4, notice is hereby given of the institution of a trial; the trial will commence on the entry date of this decision.

⁶ Petitioner asserts that the bigram model is a syntactic model. Pet. 69 (citing Ex. 1004 ¶ 174). The assertion is supported by the cited evidence and not disputed by Patent Owner.

⁷ Patent Owner has not filed a preliminary response.

IPR2023-01386
Patent 6,904,405 B2

For PETITIONER:

Erika H. Arner
Sydney R. Kestle
Jia Lu
Finnegan, Henderson, Farabow, Garrett & Dunner, LLP
erika.arner@finnegan.com
sydney.kestle@finnegan.com
jia.lu@finnegan.com

For PATENT OWNER:

Zachariah S. Harrington
Larry D. Thompson
Matthew J. Antonelli
Rehan M. Safiullah
ANTONELLI, HARRINGTON & THOMPSON LLP
zac@ahtlawfirm.com
larry@ahtlawfirm.com
matt@ahtlawfirm.com
rehan@ahtlawfirm.com