

Filed: September 12, 2023

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,

Petitioner,

v.

BUFFALO PATENTS, LLC,

Patent Owner.

Case No. IPR2023-01386

U.S. Patent No. 6,904,405

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 6,904,405**

TABLE OF CONTENTS

I.	Preliminary Statement	1
II.	The '405 Patent.....	3
A.	Overview of the '405 Patent.....	3
B.	Earliest Available Priority Date	5
III.	Overview of Prior Art.....	5
A.	<i>Waibel</i>	5
1.	Incorporation of <i>Suhm</i> , <i>Vo</i> , and <i>Manke</i>	7
a.	<i>Suhm</i>	8
b.	<i>Vo</i>	10
c.	<i>Manke</i>	14
2.	<i>Waibel</i> , <i>Suhm</i> , <i>Vo</i> , and <i>Manke</i> Are Prior Art.....	15
a.	<i>Waibel</i>	15
b.	<i>Suhm</i>	16
c.	<i>Vo</i>	17
d.	<i>Manke</i>	18
B.	<i>Roberts</i>	19
C.	<i>Verrier</i>	20
D.	<i>Roth</i>	21
IV.	Level of Ordinary Skill.....	22
V.	Claim Construction.....	23
VI.	Claims 1-12 Are Unpatentable	23

A.	Ground 1: Claims 1-3, 6-9, and 12 Would Have Been Obvious over <i>Waibel</i> (incorporating <i>Suhm</i> , <i>Vo</i> , and <i>Manke</i>) and <i>Roberts</i> ; or Ground 2: <i>Waibel</i> in view of <i>Suhm</i> , <i>Vo</i> , <i>Manke</i> , and <i>Roberts</i>	23
1.	Independent Claim 1	24
a.	[1P] “ <i>A system for generating text responsive to message input of a first type and a second type, the system comprising:</i> ”	24
b.	[1a] “ <i>(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;</i> ”	25
i.	“ <i>(1) a shared language model</i> ”	26
ii.	“ <i>(2) a first model specific to a first type of message recognition</i> ”	32
iii.	“ <i>(3) a second model specific to a second type of message recognition</i> ”	32
c.	[1b] “ <i>(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</i> ”	33
d.	[1c] “ <i>(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</i> ”	34
e.	[1d] “ <i>(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</i> ”	35
2.	Claim 2	41
3.	Claim 3	42

4.	Claim 6.....	44
5.	Independent Claim 7	45
a.	[7P] “ <i>A method for performing message recognition with a shared language model, the method comprising:</i> ”	45
b.	[7a] “ <i>(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;</i> ”	45
c.	[7b] “ <i>(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</i> ”	46
d.	[7c] “ <i>(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.</i> ”	46
6.	Claim 8.....	46
7.	Claim 9.....	47
8.	Claim 12	47
B.	Ground 3: Claims 4, 5, 10, and 11 Would Have Been Obvious over <i>Waibel</i> (incorporating <i>Suhm, Vo, Manke</i>), <i>Roberts</i> , and <i>Verrier</i> ; or Ground 4: <i>Waibel, Suhm, Vo, Manke, Roberts</i> , and <i>Verrier</i>	48
1.	Claim 4.....	48
2.	Claim 5	51
3.	Claim 10.....	53

4.	Claim 11	53
C.	Ground 5: <i>Roth</i> and <i>Roberts</i> Render Obvious Claims 1-3 and 7-9	53
1.	Independent Claim 1	53
a.	[1P].....	54
b.	[1a].....	55
i.	“(1) a shared language model”	55
ii.	“(2) a first model specific to a first type of message recognition”	57
iii.	“(3) a second model specific to a second type of message recognition”	57
c.	[1b].....	58
d.	[1c].....	58
e.	[1d].....	59
2.	Claim 2	61
3.	Claim 3	62
4.	Independent Claim 7	64
a.	[7P].....	64
b.	[7a].....	64
c.	[7b].....	64
d.	[7c].....	64
5.	Claim 8	65
6.	Claim 9	65
D.	Ground 6: <i>Roth</i> , <i>Roberts</i> , and <i>Verrier</i> Render Obvious Claims 4, 5, 10, and 11	65

1.	Claim 4.....	65
2.	Claim 5.....	67
3.	Claim 10.....	68
4.	Claim 11.....	69
E.	Ground 7: <i>Roth, Roberts, and Vo</i> Render Obvious Claims 6 and 12.....	69
1.	Claim 6.....	69
2.	Claim 12.....	70
VII.	Noninstitution Would Be Improper Under 35 U.S.C. § 314.....	70
VIII.	Noninstitution Would Be Improper Under 35 U.S.C. § 325(d).....	71
IX.	Mandatory Notices.....	73
A.	Real Party-in-Interest.....	73
B.	Related Matters.....	74
C.	Lead and Back-Up Counsel, and Service Information.....	74
X.	Grounds for Standing.....	75
XI.	Conclusion.....	75

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Advanced Bionics, LLC v. MED-EL Elektromedizinische Geräte GmbH,</i> IPR2019-01469, Paper 6 (P.T.A.B. Feb. 13, 2020).....	71
<i>Commonwealth Sci. & Indus. Rsch. Org. v. Buffalo Tech. (USA), Inc.,</i> 542 F.3d 1363 (Fed. Cir. 2008)	24
<i>Ericsson Inc. v. Intellectual Ventures I LLC,</i> IPR2014-00527, Paper 41 (PTAB May 18, 2015)	15, 16, 17, 18
<i>Harari v. Lee,</i> 656 F.3d 1331 (Fed. Cir. 2011)	7
<i>Intel Corp. v. FG SCR LLC,</i> IPR2020-01449, Paper 53 (PTAB Mar. 1, 2022).....	15, 16, 18
<i>Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.,</i> 868 F.3d 1013 (Fed. Cir. 2017)	23
<i>Paice LLC v. Ford Motor Co.,</i> 881 F.3d 894 (Fed. Cir. 2018)	7, 8
<i>Phillips v. AWH Corp.,</i> 415 F.3d 1303 (Fed. Cir. 2005) (en banc)	23
<i>SharkNinja Operating LLC v. iRobot Corp.,</i> IPR2020-00734, Paper 11 (PTAB Oct. 6, 2020).....	73
<i>Unified Patents, LLC v. MemoryWeb, LLC,</i> Decision Granting Director Review Vacating-in-part the Final Written Decision and Vacating Board Order, IPR2021-01413, Paper 76 (PTAB May 22, 2023)	73
<i>Versa Prods. v. Varidesk, LLC,</i> IPR2020-00387, Paper 13 (PTAB, July 10, 2020).....	71

Statutes

35 U.S.C. § 102(b)	1, 2
35 U.S.C. § 103	1
35 U.S.C. § 311	1
35 U.S.C. § 314.....	70, 71
35 U.S.C. § 315	73
35 U.S.C. § 325(d)	71

Regulations

37 C.F.R. § 42.100(b)	23
-----------------------------	----

TABLE OF EXHIBITS

Exhibit	Description
Ex1001	U.S. Patent No. 6,904,405 to Suominen (“the ’405 patent”)
Ex1002	Prosecution History of U.S. Patent No. 6,904,405
Ex1003	Curriculum Vitae of Christopher Schmandt
Ex1004	Declaration of Christopher Schmandt (“Decl.”)
Ex1005	A. Waibel et al., “Multimodal Interfaces for Multimedia Information Agents,” IEEE (1997) (“ <i>Waibel</i> ”)
Ex1006	WO 92/05517 to Roth (“ <i>Roth</i> ”)
Ex1007	U.S. Patent No. 5,027,406 to Roberts et al. (“ <i>Roberts</i> ”)
Ex1008	B. Suhm et al., “JANUS: Towards Multilingual Spoken Language Translation,” Proc. ARPA SLT Workshop 95 (Austin, Texas) (“ <i>Suhm</i> ”)
Ex1009	Vo et al., “Building and Application Framework for Speech and Pen Input Integration in Multimodal Learning Interfaces,” Proc. ICASSP ’96 (Atlanta, GA) (“ <i>Vo</i> ”)
Ex1010	S. Manke et al., “The Use of Dynamic Writing Information in a Connectionist On-Line Cursive Handwriting Recognition System,” <i>Advances in Neural Information Processing Systems</i> 6, Morgan Kaufmann, 1994 (“ <i>Manke</i> ”)
Ex1011	EP 0622724 to Verrier (“ <i>Verrier</i> ”)
Ex1012	Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,” Proceedings of the IEEE, Vol. 77, No. 2, Feb. 1989 (“ <i>Rabiner</i> ”)
Ex1013	Klevans et al., “Voice Recognition,” Artech House 1997, pp. 38-45 (“ <i>Klevans</i> ”)

Exhibit	Description
Ex1014	Jelinek, “Statistical Methods for Speech Recognition,” Copyright 1997 Massachusetts Institute of Technology, Third Printing 2001, chapter 3 “The Acoustic Model,” pp. 39-54 (“ <i>Jelinek</i> ”)
Ex1015	Rabiner and Juang, “Fundamentals of Speech Recognition,” Chapter 6, pp. 321-386, Prentice Hall PTR, 1993 (“ <i>Rabiner & Juang</i> ”)
Ex1016	Bellegarda et al. “A Fast Statistical Mixture Algorithm for On-Line Handwriting Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 12, 1994, 1227-1233
Ex1017	U.S. Patent No. 5,459,798 to Bailey et al. (“ <i>Bailey</i> ”)
Ex1018	U.S. Patent No. 5,855,000 to Waibel et al. (“’000 Patent”)
Ex1019	Declaration of Shauna Wiest Regarding <i>Waibel</i> and Appendices A-C
Ex1020	Declaration of Shauna Wiest Regarding <i>Suhm</i> and Appendices A-C
Ex1021	Declaration of Shauna Wiest Regarding <i>Vo</i> and Appendices A-G
Ex1022	Declaration of Shauna Wiest Regarding <i>Manke</i> and Appendices A-C
Ex1023	Starner et al. “On-Line Cursive Handwriting Recognition Using Speech Recognition Methods” Proceedings of ICASSP ’94. IEEE International Conference on Acoustics, Speech and Signal Processing, 1994 (“ <i>Starner</i> ”)
Ex1024	Order Granting Motorola Mobility’s Motion to Stay in <i>Buffalo Patents, LLC v. Motorola Mobility LLC</i> , Case 1:22-cv-04540, Dkt. 30 (Dec. 13, 2022).
Ex1025	LegalMetric Time to Trial Report Northern District of Illinois, Patent Cases, January 2013 to April 2023.
Ex1026	Bellegarda, et al., “A Novel Word Clustering Algorithm Based on Latent Semantic Analysis,” IEEE, 1996 (“ <i>Bellegarda</i> ”)
Ex1027	Declaration of Shauna Wiest Regarding <i>Bellegarda</i>

Exhibit	Description
Ex1028	Proof of Service in <i>Buffalo Patents, LLC v. Motorola Mobility LLC</i> , Case 1:22-cv-04540, Dkt. 18 (Sept. 13, 2022)
Ex1029	Moran et al., “Multimodal User Interfaces in the Open Agent Architecture,” Knowledge-Based Systems Vol. 10 295-303, 1998 (“ <i>Moran</i> ”)

I. Preliminary Statement

Petitioner requests review under 35 U.S.C. § 311 of claims 1-12 of U.S. Patent No. 6,904,405 (“the ’405 patent”) and cancellation of the challenged claims under pre-AIA 35 U.S.C. § 103.

Exhibit	Prior Art
Ex1005	<i>Waibel</i> , a publication entitled “Multimodal Interfaces for Multimedia Information Agents”; published in 1997; prior art under at least pre-AIA 35 U.S.C. § 102(b).
Ex1006	<i>Roth</i> , WO 92/05517; published April 2, 1992; prior art under at least pre-AIA 35 U.S.C. § 102(b).
Ex1007	<i>Roberts</i> , U.S. Patent No. 5,027,406; issued June 25, 1991; prior art under at least pre-AIA 35 U.S.C. § 102(b).
Ex1008	<i>Suhm</i> , a publication entitled “JANUS: Towards Multilingual Spoken Language Translation”; published in 1995; prior art under at least pre-AIA 35 U.S.C. § 102(b).
Ex1009	<i>Vo</i> , a publication entitled “Building and Application Framework for Speech and Pen Input Integration in Multimodal Learning Interfaces”; published in 1996; prior art under at least pre-AIA 35 U.S.C. § 102(b).

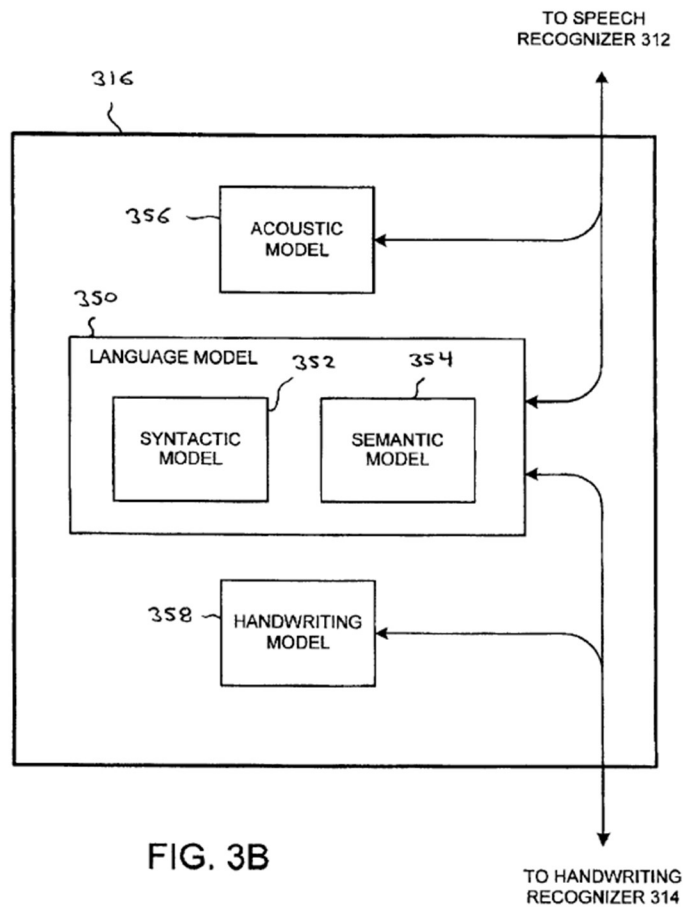
Exhibit	Prior Art
Ex1010	<i>Manke</i> , a publication entitled “The Use of Dynamic Writing Information in a Connectionist On-Line Cursive Handwriting Recognition System”; published in 1994; prior art under at least pre-AIA 35 U.S.C. § 102(b).
Ex1011	<i>Verrier</i> , EP 0622724A2; published November 2, 1994; prior art under at least pre-AIA 35 U.S.C. § 102(b).

Grounds of Unpatentability	
1, 2	<i>Waibel</i> (incorporating <i>Suhm</i> , <i>Vo</i> , and <i>Manke</i>) and <i>Roberts</i> (Ground 1) or <i>Waibel</i> in view of <i>Suhm</i> , <i>Vo</i> , <i>Manke</i> , and <i>Roberts</i> (Ground 2) render obvious claims 1-3, 6-9, and 12
3, 4	<i>Waibel</i> (incorporating <i>Suhm</i> , <i>Vo</i> , and <i>Manke</i>), <i>Roberts</i> , and <i>Verrier</i> (Ground 3) or <i>Waibel</i> , <i>Suhm</i> , <i>Vo</i> , <i>Manke</i> , <i>Roberts</i> , and <i>Verrier</i> (Ground 4) render obvious claims 4, 5, 10, and 11
5	<i>Roth</i> and <i>Roberts</i> render obvious claims 1-3 and 7-9
6	<i>Roth</i> , <i>Roberts</i> , and <i>Verrier</i> render obvious claims 4, 5, 10, and 11
7	<i>Roth</i> , <i>Roberts</i> , <i>Vo</i> , and <i>Verrier</i> render obvious claims 6 and 12

II. The '405 Patent

A. Overview of the '405 Patent

The '405 patent discloses methods and systems to perform “multiple different types of message recognition using a shared language model.” '405 patent, Abstract. It explains a shared language model is “any language model that may be shared between two types of subordinate message recognizers included within a message recognizer.” *Id.*, 22:38-51. In one embodiment, the specification discloses an acoustic model 356 for use by a speech recognizer, a handwriting model 358 for use by a handwriting recognizer, and a language model 350 shared between the speech recognizer and handwriting recognizer. *Id.*, 22:52-23:8, Fig. 3B. These are shown in Figure 3B:



Id., Fig. 3B; *id.*, 22:39-23:8. According to the '405 patent, the speech recognizer provides text data in response to speech input using the acoustic model and shared language model, and the handwriting recognizer provides text data in response to handwriting input using the handwriting model and shared language model. *Id.*, 22:65-23:2, Fig. 3B. The '405 patent explains the shared language model can be trained by a user making corrections to any misrecognition by the speech or the handwriting recognizer. *Id.*, 23:2-8, claim 1.

B. Earliest Available Priority Date

The application that matured into the '405 patent was filed January 28, 2002, and claims priority to U.S. Provisional Application No. 60/144,481, filed July 17, 1999.

III. Overview of Prior Art

A. *Waibel*

According to *Waibel*, existing human-computer interfaces were “limited and inflexible,” failing to take advantage of the “communication channels humans use to communicate.” *Waibel*, 167. To address this, and build computer systems “that are more intuitive” and that “take advantage of a rich spectrum of cues,” *Waibel* describes a multimodal system and method of generating text; specifically, doing so jointly in response to inputs from multiple modalities (speech or pen). *Waibel*, Abstract, 167. This system includes:

- 1) A speech recognition subsystem “based on the recognition front-end of the JANUS speech translation system [5]”; and
- 2) An “MS-TDNN-based handwriting recognizer [7]” “capable of processing writer-independent, continuous (cursive) handwriting.”

Id., 167; *see also id.*, 167, 170 (incorporated reference [5] included as Ex1008; incorporated reference [7] included as Ex1010).

Waibel teaches the output from the processors of its speech and handwriting recognizers are input into a multimodal interpreter “capable of producing an interpretation of user intent . . . from the output of the modality processors.” *Id.*, 167-68. *Waibel* calls this “joint interpretation.” *Id.* And *Waibel* discloses:

In our joint interpretation scheme, the user intent is represented by a frame consisting of slots specifying pieces of information such as the action to carry out or the parameters for that action. Recognition output from the modality processors are parsed into partially filled frames . . . merged together to produce the combined interpretation as described in [6]. This technique leads to uniform handling of high-level information from all input sources, which is very important for modularity and extensibility.

Id., 168; *see also id.*, 168, 170 (incorporated reference [6] included as Ex1009).

Waibel also discloses Multimodal Error Correction. *Id.*, 168. The potential for error in these interfaces can “be balanced by the redundant and alternate ability to enter information or queries by spelled, handwritten or typed input.” *Id.* *Waibel* explains:

Our approach is to have the user collaborate with the system in recovering from interpretation errors [10]. 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.

Id., 168. Decl. ¶¶46-52.

1. Incorporation of *Suhm*, *Vo*, and *Manke*

In addition to its broad teachings, *Waibel* incorporates three papers: *Suhm*, *Vo*, and *Manke*. *Id.*, 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.

a. *Suhm*

Suhm describes the speech recognizer in *Waibel*, namely, the JANUS speech translation system where spoken input is translated into text output. *Waibel*, 167, 170; *Suhm*, 221. To “tackle the difficulty of spontaneous human-human dialogs,” *Suhm* “improved the JANUS-2 recognizer along its three knowledge sources[:] acoustic models, dictionary and language models.” *Suhm*, 221. Its system “uses semantic analysis and as is “suited to processing spontaneous speech.” *Id.* The modules for the recognizer are “speech recognition, parsing, discourse processing, and generation.” *Id.* Each is “language-independent” because each “consists of a general processor that applies independently specified knowledge about different languages.” *Id.*, 221-22.

Suhn explains its speech recognition module will decode speech from the source language “into a list of sentence candidates” represented “as a word lattice or Nbest list.” *Id.*, 222. The final language independent representation is then “sent to the generator to be translated.” *Id.* The system architecture is shown in Figure 1:

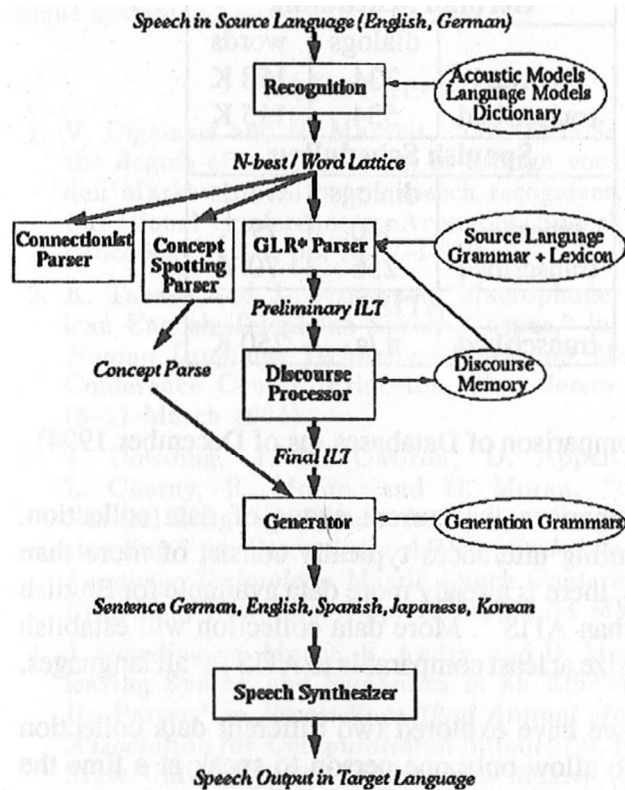


Figure 1: System Diagram

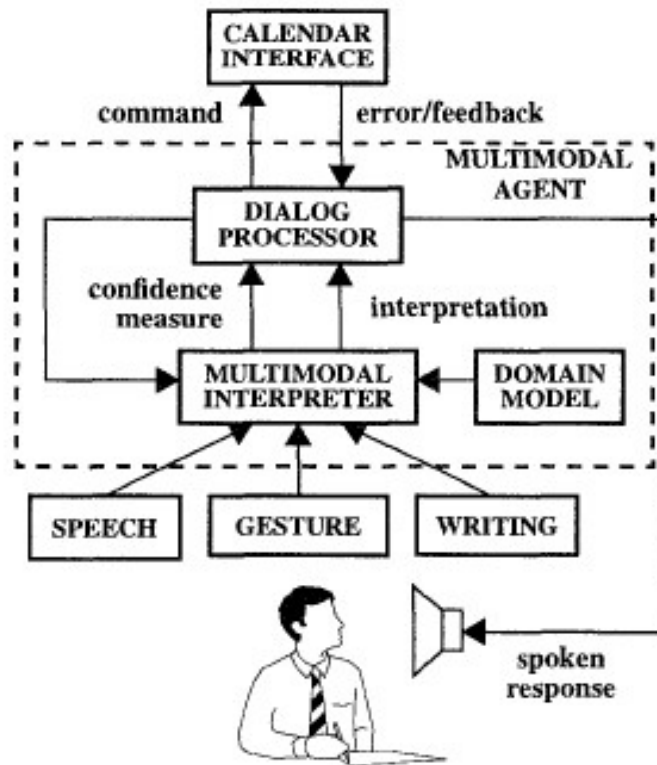
Id., 222 & Fig. 1. After parsing, “a discourse processor can be used to put the current utterance in context” and permit the system to “dynamically adapt the vocabulary and language model of the recognizer based on the current discourse state.” *Id.*, 222.

To improve the speech recognition component of its system, *Suhm* discloses a semi-continuous, phonetically tied Hidden Markov Model (HMM), and adapts the codebook size of each phoneme to the amount and distribution of training data. *Id.*, 223. *Suhm* also discloses a “standard word bigram language model” and “data-driven approach to improve existing dictionaries and automatically add new words

and pronunciation variants whenever needed.” *Id.*, 222-23. Its “learning algorithm requires transcripts for the whole training set and a phoneme confusability matrix of the speech recognizer.” *Id.*, 223. It may also use morpheme-based language models to limit the growth rate for “morphologically rich languages.” *Id.*, 223-24. Decl. ¶¶55-57.

b. *Vo*

Vo describes the joint interpreter in *Waibel*, which jointly interprets speech, pen-based gestures, and handwriting information from different sources based on semantic frame merging. *Waibel*, 168, 170; *Vo*, 3545. The “multimodal interpreter” produces an “interpretation of user intent” based on outputs from the modality processors. *Id.*, 3546. Specifically, these outputs are “parsed into partially filled frames that are merged together to produce the combined speech interpretation.” *Id.* Figure 1 is a block diagram of *Vo*’s system:



Id., 3545-46 & Fig. 1. According to *Vo*, individual modality components (speech, gesture, and handwriting recognizers) may be separate modules and replaced with ease. *Id.*, 3545.¹

As to the speech recognizer, *Vo* explains it may capture inputs by a speech recognizer. *Id.*, 3546. The “text string output” is then processed by the Phoenix semantic parser, which “compiles a grammar specifying semantically meaningful

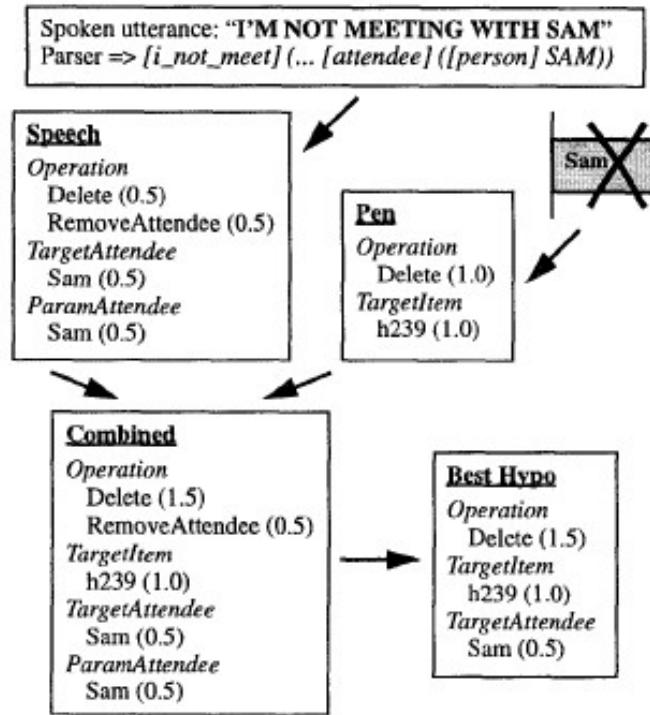
¹ *Vo* also explains its speech recognition subsystem is based on the JANUS speech translation system, which was discussed above with respect to *Suhm*. *Vo*, 3546. Its handwriting recognizer is based on the MS-TDNN. *Id.*

fragment of text” into a “network that identifies these fragments in the recognized utterance.” *Id.* The parser can “skip[] unknown words and unmatched fragments” and “deal gracefully with ungrammatical sentences.” *Id.*

As to the handwriting recognizer, *Vo* explains it captures inputs by a “touch-sensitive screen.” *Id.*, 3546. It also “produce[s] a sequence of gesture/handwriting components” that “may mean different things depending on the calendar context,” and words are “encoded into frames according to simple hard-coded heuristics.” *Id.*

Vo explains “[f]rames are merged by taking the union of the sets of values filling each slot[] and adding corresponding scores.” *Id.*, 3547. This “produces an aggregate frame encoding all alternative interpretation hypotheses.” *Id.* Figure 2 shows “interpretation using frames”:

FIGURE 2: Example of interpretation using frames



Id., Fig. 2. The “multimodal interpreter extracts these hypotheses in descending order of aggregate scores and sends the best ones to the dialog processor.” *Id.*, 3547. The “dialog processor maintains a single frame per dialog to hold all [the] information obtained since the start of the dialog; new information from each additional input event is merged into this frame and the scores are weighted appropriately so that frames produced by different input events in the same dialog contribute equally to the end result.” *Id.* It will also “filter out hypotheses that would cause errors.” *Id.* Once a “good hypothesis” is found, the dialog processor will send it to the “domain interface,” where a “user has the option of undoing the operation just performed if it is incorrect.” *Id.*; Decl. ¶¶58-61.

c. *Manke*

Manke describes the handwriting interpreter in *Waibel*, namely, the NPen⁺⁺ system for writer-independent, large vocabulary, on-line cursive handwriting recognition. *Waibel*, 167, 170; *Manke*, Abstract. *Manke*'s system combines an input representation that "preserves the dynamic writing information" with "a neural network architecture," MS-TDNN, which "integrates recognition and segmentation in a single framework." *Id.*

During preprocessing, "a temporal sequence of *N*-dimensional feature vectors is computed from the original coordinate sequence" and recorded. *Id.*, 1094-95. The recognizer employs Hidden Markov Models (HMM) where "each character c_j itself is modelled by a three state hidden markov model." *Id.*, 1096.

During recognition processing, *Manke*'s system integrates word recognition and segmentation into one network architecture using a model (MS-TDNN) that was "originally proposed for continuous speech recognition tasks." *Id.*, 1094. This "combines shift-invariant, high accuracy pattern recognition capabilities of a TDNN" with "a non-linear alignment procedure for aligning strokes into character sequences." *Id.*

Manke tested its model with "CRT (Conference Registration Task) and WSJ (ARPA Wall Street Journal Task)" vocabularies. *Id.*, 1099. These two vocabularies were chosen because of the "synergy effects between handwriting recognition and

speech recognition,” since the “goal is to integrate our speech recognizer” and the NPen⁺⁺ system “into a multi-modal system.” *Id.* Decl. ¶¶62-65.

2. *Waibel, Suhm, Vo, and Manke* Are Prior Art

a. *Waibel*

Waibel is prior art because it was publicly available before the '405 priority date, and persons interested and ordinarily skilled in the art exercising reasonable diligence would have located it.

First, *Waibel* includes a 1997 Copyright stamp and price information. Ex1019, pp.28 (Appx. A). It was included in an IEEE publication from 1997, which also includes a 1997 Copyright stamp, as well as IEEE catalog and order numbers, ISBN numbers, a Library of Congress identification number, a date stamp from the Library of Congress showing the library received the publication in 1997, and instructions for ordering additional copies. Ex1019, ¶¶9-13, pp.16 (Appx. A). These indicia establish *Waibel* was publicly available as of its 1997 publication date. *See Ericsson Inc. v. Intellectual Ventures I LLC*, IPR2014-00527, Paper 41, at 10-12 (PTAB May 18, 2015).

Second, interested persons would have been able to locate it exercising reasonable diligence. Based on standard IEEE practices, *Waibel* would have been available to the public no later than the last day of the conference, i.e., April 24, 1996. Ex1019, pp.16 (Appendix A); *Intel Corp. v. FG SCR LLC*, IPR2020-01449,

Paper 53 at 29, 34-36, Exhibit 1027 (PTAB Mar. 1, 2022). *Waibel* was also received and catalogued at the Library of Congress on July 7, 1997, and accessible to the public within weeks of that date. Ex1019 ¶¶5-18. *Waibel*'s cataloguing is confirmed by the Library of Congress public catalog record and MARC record for the IEEE publication. *Id.* Those skilled in the art exercising reasonable diligence would have been able to locate *Waibel*, a publication covering speech recognition and language modeling.

b. *Suhm*

Suhm is prior art because it was publicly available before the '405 priority date, and persons interested and ordinarily skilled in the art exercising reasonable diligence would have located it.

First, *Suhm* was included in an ARPA publication from 1995. Ex1020, ¶12, Appx. A. The ARPA publication includes an ISBN Number. Ex1020, pp.13. These indicia establish that *Suhm* was publicly available as of 1995. *See Ericsson*, IPR2014-00527, Paper 41, at 10-12.

Second, interested persons would have been able to locate it exercising reasonable diligence. Based on standard practices, *Suhm* would have been made available to the public no later than the last day of the conference, i.e., January 25, 1995. Ex1020, pp.12; *Intel*, Paper 53 at 29, 34-36. Indeed, *Suhm* is cited by *Waibel*, showing interested persons skilled in the art actually located *Suhm* before

July 7, 1997. *Suhm* was also received and catalogued at the Iowa State University Library on October 24, 1995 or, at the very latest, April 22, 1997, and made accessible to the public within weeks of receipt. Ex1020 ¶¶5-21. *Suhm*'s cataloguing is confirmed by the Library of Congress public catalog record and the MARC record for the 1997 IEEE publication. *Id.* ¶¶16-19. Those skilled in the art exercising reasonable diligence would have located *Suhm*, a publication relating to speech recognition and language modeling.

c. *Vo*

Vo is prior art because it was publicly available before the '405 priority date, and persons interested and ordinarily skilled in the art exercising reasonable diligence would have located it.

First, *Vo* includes a 1996 Copyright stamp and price information. Ex1021, pp.26. *Vo* was included in an IEEE publication from 1996, which also includes a 1996 Copyright stamp, as well as an IEEE catalog number, ISBN numbers, a Library of Congress identification number, and instructions for ordering additional copies. Ex1021, pp.16-18. These indicia establish *Vo* was publicly available as of its 1996 publication date. *See Ericsson*, IPR2014-00527, Paper 41, at 10-12.

Second, interested persons would have been able to locate it exercising reasonable diligence. Based on standard IEEE practices, *Vo* would have been made available to the public no later than the last day of the conference, i.e., May 10,

1996. Ex1021, pp.17; *see Intel*, Paper 53 at 29, 34-36, Exhibit 1027. In fact, *Vo* is cited by *Waibel*, showing that skilled artisans actually located *Vo* before July 7, 1997. Moreover, *Vo* was received and catalogued at the Library of Congress in 1997 and made accessible to the public within weeks of receipt. Ex1021 ¶¶5-25. *Vo*'s cataloguing is confirmed by the Library of Congress public catalog record and the MARC record for the 1997 IEEE publication. *Id.* ¶¶16-19. Those skilled in the art exercising reasonable diligence would have been able to locate *Vo*, a publication relating to speech recognition and language modeling.

d. *Manke*

Manke is prior art because it was publicly available before the '405 priority date, and persons interested and ordinarily skilled in the art exercising reasonable diligence would have located it.

First, *Manke* was included in an MIT publication from 1995. Ex1022, ¶¶11-14, pp.11-14. The MIT publication includes a 1995 Copyright stamp, an ISSN number, an ISBN number, instructions for reproducing copies, and a date stamp from the University of Wisconsin library that it received the publication on August 17, 1995. *Id.* These indicia establish *Manke* was publicly available as of its 1995 publication date. *See Ericsson*, IPR2014-00527, Paper 41, at 10-12.

Second, interested persons would have been able to locate it exercising reasonable diligence. *Manke* is cited by *Waibel*, showing skilled artisans actually

located *Manke* before July 7, 1997. Moreover, *Manke* was received and catalogued at the University of Wisconsin library on August 17, 1995, and made accessible to the public within weeks of that date. Ex1022 ¶¶5-19. *Manke*'s cataloguing is confirmed by the University of Wisconsin library public catalog record and the MARC record for the 1995 MIT publication. *Id.* ¶¶15-17. Those skilled in the art exercising reasonable diligence would have located *Manke*, a publication relating to speech recognition and language modeling.

B. *Roberts*

Roberts discloses a speech recognition method training models for use in speech recognition. *Roberts*, 1:6-11. It enables a user to “build acoustic models from tokens that are collected during actual use of the system,” “easily correct errors,” and “develop a language model for use in a speech recognition system” “without requiring much additional effort” to customize the language model for each user’s “particular diction.” *Id.*, 2:17-30.

Roberts's method includes having the user “speak[] a word which may or may not already be in the active vocabulary.” *Id.*, Abstract. The system “displays a list of the words in the active vocabulary which best match the spoken word.” *Id.* “By keyboard or voice command, the user may choose the correct word from the list[,] or may choose to edit a similar word if the correct word is not on the list.” *Id.*, 4:19-20, 16:22-24, 23:32-35. Alternatively, *Roberts* explains that a user may

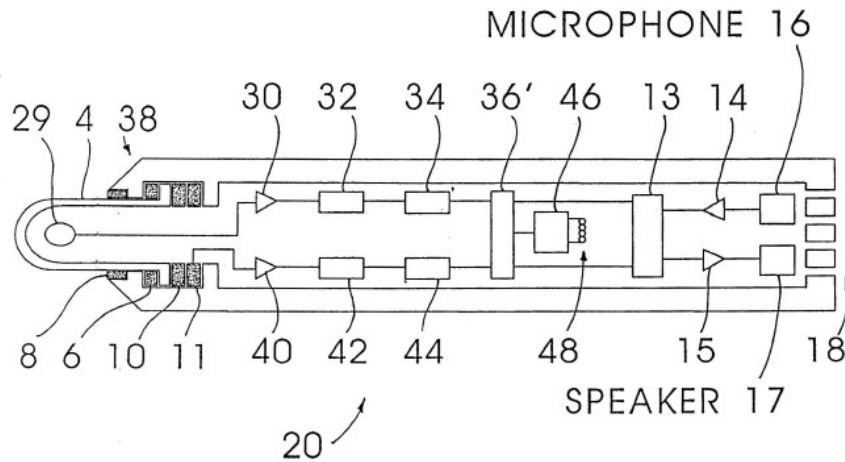
type or speak the first letters of the word. *Id.*, Abstract. Its recognition algorithm will then call words satisfying those letters, and the “choices displayed again” to the user. *Id.* This list may be “displayed from a large backup vocabulary,” with the “best words” selected by “a statistical language model and optionally word models derived from a phonemic dictionary.”² *Id.* Once the correct word is identified, a “speech sample is used to create or update an acoustic model for the word.” *Id.* *Roberts* explains that its system will “use[] the confirmed word . . . to update the language model used by the recognition system.” *Id.*, 13:44-47, Fig. 8, 18:14-19:31, Abstract. In this way, *Roberts* explains that “[t]he system gets more and more word models and keeps improving its performance the more it is used.” *Id.*, Abstract. Decl. ¶¶66-71.

C. *Verrier*

Verrier teaches a stylus with a microphone and speaker. *Verrier*, Abstract. This is shown in Figure 2:

² *Roberts* explains “such language models are well-known” and its invention would work with any of them. *Roberts*, 13:44-60.

FIG. 2



Id., Fig. 2. 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.*, Abstract. *Verrier*’s teaching was well-known in the art. ’405 patent, 2:54-64. Decl. ¶68.

D. Roth

Roth teaches an “Audio-Augmented Handwriting Recognition” system, where handwriting input (element 1) and speech input (element 6) are used to generate text (elements 9, 28). This is shown below in Figure 1:

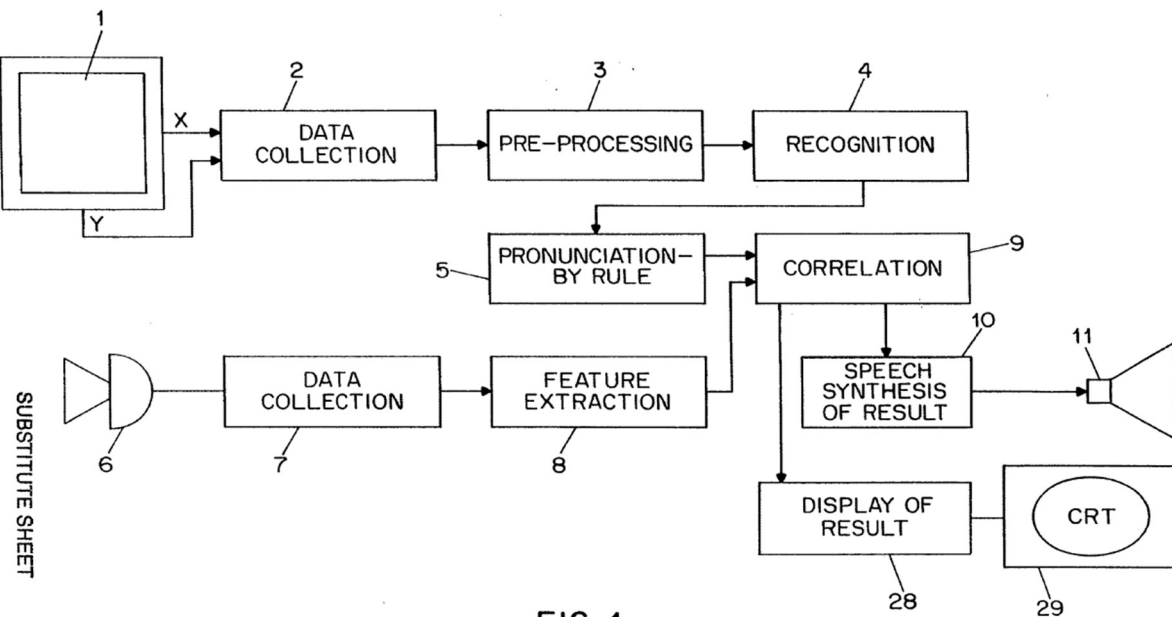


FIG. 1

Id., Fig. 1. In one embodiment, *Roth* discloses analyzing handwritten input to derive a number of candidate words and receiving speech input to correlate with that handwritten input to narrow candidate words to one result before displaying the match word on the display. *Id.*, 4:12-28, 6:23-7:4, 8:11-19, 11:3-15. *Roth* also teaches the “opposite”—analyzing speech input to derive candidate words and then receiving handwritten input to correlate with the speech input to narrow candidates to one result before displaying the result. *Id.*, 11:16-30. *Roth* explains its system could “apply both approaches . . . and choose the result with the higher degree of confidence.” *Id.*, 11:30-32. Decl. ¶¶69-71.

IV. Level of Ordinary Skill

A person of ordinary skill in the art (“POSITA”) at the time of invention 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. Decl. ¶¶33-37.

Work experience can substitute for formal education and additional formal education can substitute for work experience. *Id.*

V. Claim Construction

The Board construes claims in accordance with 37 C.F.R. § 42.100(b) and *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc). Claims need to be construed only to the extent necessary to resolve a controversy. *Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017). Here, no terms require construction because the claims encompass the prior-art mappings under any reasonable construction consistent with the *Phillips* standard. Decl. ¶¶38-42.

VI. Claims 1-12 Are Unpatentable

A. Ground 1: Claims 1-3, 6-9, and 12 Would Have Been Obvious over *Waibel* (incorporating *Suhm*, *Vo*, and *Manke*) and *Roberts*; or Ground 2: *Waibel* in view of *Suhm*, *Vo*, *Manke*, and *Roberts*

Waibel's multimodal recognition method (incorporating *Suhm*, *Vo*, *Manke*), in view of *Roberts*, renders obvious claims 1-3, 6-9, and 12. Decl. ¶¶72-73. If the Board finds *Suhm*, *Vo*, and *Manke* were not incorporated in *Waibel*, it would nevertheless have been obvious to a POSITA to combine these references, for

Waibel's citations to *Suhm*, *Vo*, and *Manke* "provide a justification for combining the references for obviousness purposes." *Commonwealth Sci. & Indus. Rsch. Org. v. Buffalo Tech. (USA), Inc.*, 542 F.3d 1363, 1372 (Fed. Cir. 2008). As such, and in the alternative, *Waibel* with *Suhm*, *Vo*, *Manke*, and *Roberts* renders obvious claims 1-3, 6-9, and 12. Decl. ¶¶72-73.

1. Independent Claim 1

- a. [1P] ***"A system for generating text responsive to message input of a first type and a second type, the system comprising:"***

If the Board finds the preamble limiting, *Waibel* (with *Suhm*, *Vo*, and *Manke*) discloses it. Decl. ¶74. *Waibel* discloses a system of generating text. *Waibel*, 169-70. It explains its system comprises several multimodal processors, including the speech recognition system from *Suhm* and handwriting recognition system from *Manke*. *Id.*, 167. These recognizers receive different inputs, including message input of a first type ("speech") at the speech recognizer, and message input of a second type ("handwriting") at the handwriting recognizer. *Id.* These recognizers output information to the joint interpreter, which in turn "produc[es] an interpretation of user intent." *Id.*, 167-168. This interpretation includes producing an "HTML report" or QuickDoc, which a POSITA would understand includes text. *Id.*, 169-70 (noting "audio file is also run through the speech recognizer to produce an automatic transcription"); Decl. ¶74. If the graphical user interface is available,

the user can highlight erroneous words in the “recognition hypothesis displayed” and “choos[e] among different correction methods.” *Waibel*, 168.

- b. [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;*”

Waibel (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious this limitation. Decl. ¶¶75-76. Indeed, *Waibel* discloses or at least renders obvious having a “*a unified message model*”³ that includes (1) a shared language model in the speech recognizer, handwriting recognizer, and/or joint recognizer (*Waibel*, 167-68); (2) a model in the recognition front-end of the JANUS speech translation system specific to speech input (a “first type of message recognition”) (*id.*, 167); and also (3) a model in the MS-TDNN-based handwriting recognizer specific to a handwriting input (the “second type of message recognition”) (*id.*). Decl. ¶¶75-76. According to *Waibel*, such technique “[r]ecogni[z]es output from the modality processors . . . merged together to produce the combined interpretation” “leads to

³ The ’405 specification explains an exemplary “unified message model” is in Figure 3B. ’405 patent, 22:52-64.

uniform handling of high-level information from all input sources.” *Waibel*, 168; Decl. ¶¶75-76.

i. “(1) a *shared language model*”

The ’405 patent explains a “shared language model” is “any language model that may be shared between two types of subordinate message recognizers included within a message recognizer.” ’405 patent, 22:38-51, Fig. 3B. *Waibel* (with *Suhm*, *Vo*, *Manke*) discloses or at least renders obvious such “*shared language model*.” Decl. ¶¶77-87.

First, *Waibel* discloses a language model used in its speech and/or handwriting recognizers. *Waibel*, 167 (speech recognizer “can be adapted to any task domain by retraining *the language models* and possibly tuning the acoustic models” (emphasis added)); *Suhm*, 222 (“standard word bigram language model”); *Manke*, 1098, 1099 (dictionaries). A POSITA 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. Decl. ¶¶77-87. Not only was the sharing of a language model between different recognizers well-known at the time (Decl. ¶77, citing, e.g., *Roberts*, Fig. 1 (103, 111, 104, 180),

'000 Patent, 6:27-46, *Moran* at 296), *Waibel*, in particular, was designed to favor the use of shared language models, as explained below. Decl. ¶¶77-87.

For example, *Manke* teaches using handwriting language models “originally defined for speech recognition” and chosen to “take advantage of the *synergy effects* between handwriting recognition and speech recognition.” *Manke*, 1099 (emphasis added). *Manke* teaches using the “synergy effects” between handwriting and speech recognition “to integrate our speech recognizer . . . and the proposed NPen⁺⁺ system into a multi-modal system.” *Id.* A POSITA would have understood that such a language model in the handwriting recognizer would have been shared, or would have been obvious to be shared, such that the handwriting recognizer and speech recognizer used the same language model, and that using such a shared language model would have been *Waibel*’s preferred design to integrate the recognizers into a single multi-modal system with synergistic effects. Decl. ¶¶77-87. A POSITA would have also understood that using a shared language model would have increased the performance and accuracy of each recognizer while simplifying computation and minimizing memory requirements in *Waibel*. *Id.*

Similarly, Table 1 in *Manke* shows a handwriting recognition algorithm based on a dictionary that can be changed without rewriting the recognizer. *Id.*; *Manke*, 1098 (Table 1). A POSITA would have understood that the dictionary used in *Waibel*’s recognizer is a language model because of its role in limiting

recognition to words a user can speak, independent of *Waibel's* handwriting algorithm. Decl. ¶¶77-87. A POSITA would also have understood that the dictionary is primed for sharing between speech and handwriting recognizers due to its independence from the handwriting algorithm. *Id.* Thus, a POSITA would have at least found it obvious to share the dictionary language model used in the handwriting recognizer with the speech recognizer in *Waibel*, especially because the speech recognizer also needs to use a dictionary language model (*Suhm*, 223-24), which, for simplicity and minimized memory requirements, could be the same as the dictionary model in the handwriting recognizer. Decl. ¶¶77-87. In fact, *Waibel's* speech and handwriting recognizers were designed to accommodate the same dictionary language model, as reflected by the same way they process respective inputs, for example, by conversion into intermediate forms (phonemes and stroke units, respectively) and then into letters, as modelled by a dictionary. *Id.* In *Waibel's* system, where multiple modalities can use the same language model, it would have been unnecessary to build separate language models for speech and handwriting input. *Id.*

This is supported by other disclosures in *Waibel*. For example, even though the bigram language model in *Waibel's* JANUS speech recognizer is not expressly described as being used with the handwriting recognizer, a POSITA would have understood this to be the case because the bigram language model in the speech

recognizer and the Viterbi search of the Hidden Markov Models in the handwriting recognizer are compatible. *Manke*, 1096-98. A POSITA would have at least found it beneficial to share the bigram language model between the handwriting and speech recognizer to align with *Waibel's* stated desire to achieve synergy for vocabulary management, and to ease computation and memory requirements. *See, e.g., Suhm*, 222; *Manke*, 1099; Decl. ¶¶77-87.

Indeed, *Waibel's* speech recognizer design favors a shared language model. *Id.* Specifically, the speech recognizer is language-independent (*Suhm*, 221-22) and, therefore, model-independent such that its language model can be changed without the need to rebuild the recognizer. *Id.*; *Suhm*, 221-22. Because each model is built from training data, data processing times required to build a model can become large. *See, e.g., Waibel* Table 1 (showing close to 2000 dialogs and 500,000 words for the English Language scheduling model data). Decl. ¶¶77-87. This is true even if the same training data can produce acoustic and language models. *Id.* Performing substantially the same data processing to build multiple models for different recognizers in a system unnecessarily consumes processing times and resources. *Id.* Therefore, a POSITA would have been motivated to save resources by sharing a model between speech and handwriting recognizers when possible. *Id.* Further, because the speech recognizer is designed to use language models that are dynamically changed (*see, e.g., Suhm*, 222, language model

changed based on discourse state), it may require access to more than one language model. As such, a POSITA would have understood that a system with a need for multiple language models would have larger memory needs. And a POSITA would have been motivated to share language models in such a system to reduce these requirements. *Id.*

An aspect of a shared language model is also taught in *Vo. Id.* For example, *Vo* describes a multimodal interpreter including a dialog processor that maintains a single frame per dialog to hold information obtained from speech *and* handwriting inputs, so that new information from each additional input is merged into the frame. *Vo*, 3547 (the “dialog between a user and the system can span several related multimodal events, where each event may contain one or more spoken utterances and/or pen gestures”; “new information from each additional input event is merged into this frame and the scores are weighted appropriately”); Figs. 1 and 2. A POSITA would have understood, or at least found obvious, that the dialog frames described in *Vo*, as used in *Waibel*’s system, is further indicative of another aspect of a shared language model in the context of *Waibel*’s system. Decl. ¶¶77-87. Indeed, frames were a well-known natural language processing construct reflecting syntax and semantics. Decl. ¶¶77-87. As *Vo* describes, frames are filled in a dialog where input is requested, provided, recognized, and incorporated until a frame is complete and the system performs the intended action. *Id.*; *Vo*, 3547. The

dialog in *Vo* interacts with lower-level recognizers, i.e., the speech and handwriting recognizers, by temporarily changing their operative language model. See *Waibel*, 168; *Suhm*, 222. A POSITA would have understood the dialog frames in *Vo* to be a type of shared language model, for they embody a dynamic language model that changes the operation of the recognizers depending on the dialog state, and use shared inputs from the speech and handwriting recognizers. Decl. ¶¶77-87; *Suhm*, Fig. 2, Section 3.2 at 223.

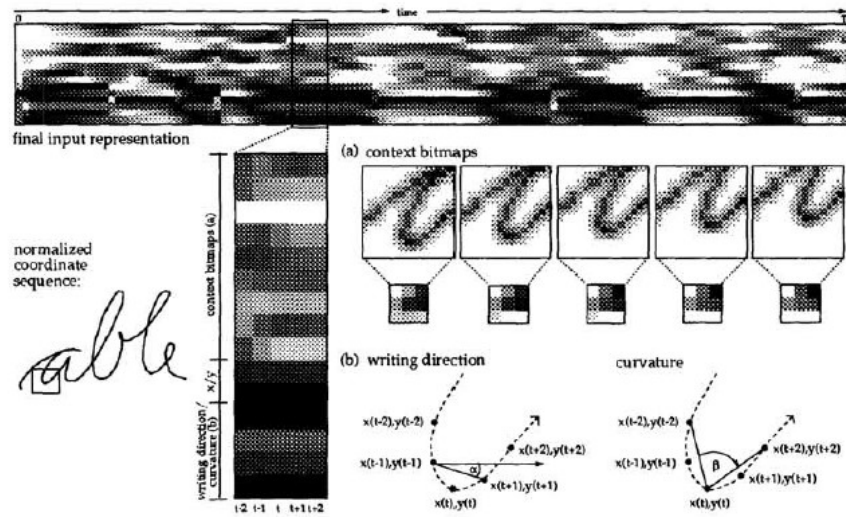
Even if modifications were necessary to use a shared language model with both the speech and handwriting recognizers, a POSITA would have found making such modifications straightforward with no technical challenges due to similarities in the techniques used by the recognizers. Decl. ¶¶77-87. For example, a POSITA would have understood that *Waibel*'s speech recognizer is a common type that decomposes words into phonemes—the smallest unit of speech—before building a three-state Hidden Markov Model representing each phoneme. Decl. ¶¶77-87; *Suhm*, 222. So too, the handwriting recognizer of *Waibel* uses characters, the smallest unit of handwriting, before building a three-state Hidden Markov Model to represent each character. *Manke*, 1096. Accordingly, a POSITA would have recognized the recognizers process words using the same type of algorithm, in turn making the use of a shared language model nearly seamless. Decl. ¶¶77-87.

ii. “(2) a first model specific to a first type of message recognition”

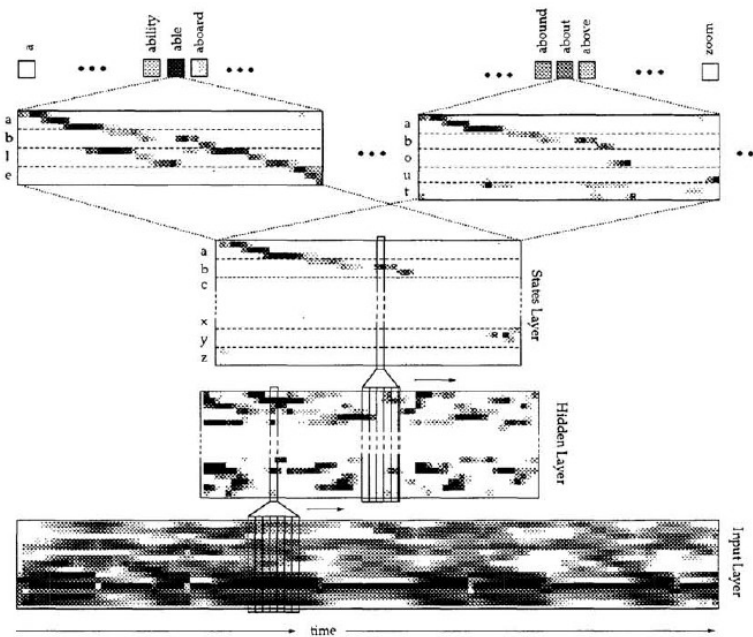
Waibel (with *Suhm*, *Vo*, and *Manke*) discloses this element. *Waibel* discloses an acoustic model for speech recognition. *Waibel*, 167. Its speech recognizer can be adapted “by retraining the language models and possibly tuning the *acoustic models*.” *Id.* (emphasis added). A POSITA would have also understood that *Suhm* describes an acoustic model in the speech recognizer, using “TDNN, MS-TDNN [3], MLP, LVQ [4], and HMM's [5, 12] for acoustic modeling.” *Suhm*, 222. Decl. ¶ 88.

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

Waibel (with *Suhm*, *Vo*, and *Manke*) discloses this element. Decl. ¶ 89. A POSITA would have understood *Manke* discloses several handwriting-specific models, including a “15-dimensional feature vector,” the three-state hidden markov model, a “Viterbi approximation,” and scores that are “derived from the word models.” *Id.*; *Manke*, 1095-97, Fig. 1, Fig. 2. Figure 1 shows a “15-dimensional feature vector”:



Manke, 1095 (Fig. 1). Figure 2 shows a “Viterbi approximation” and “scores [that] are derived from the word models”:



Id., 1097 (Fig. 2). Decl. ¶89.

- c. [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”

Waibel (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious this limitation. Decl. ¶¶90-91. *Waibel* discloses a first message recognizer (the front-end of JANUS speech translation system and joint multimodal interpreter (*Waibel*, 167)) that, responsive to a first type of message input (speech), provides text data (e.g., text output in QuickDoc (*id.*, 169-70)) in accordance with a shared language model (e.g., language model in multimodal interpreter, limitation [1a]) and the first model (e.g., acoustic model in the recognition front-end of the JANUS speech translation system, limitation [1a]). Decl. ¶¶90-91; Section VI.A.1.b; *Vo*, 3456, Fig. 2 (showing output of speech recognizer used by the parser is a text string and the “text string output from the speech recognizer is processed by the Phoenix semantic parser”).

d. [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”

Waibel (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious this limitation. Decl. ¶¶92-93. *Waibel* discloses a second message recognizer (MS-TDNN-based handwriting recognizer and joint multimodal interpreter (*Waibel*, 167)) that, responsive to a second type of message input (handwriting), provides text data (e.g., text in QuickDoc (*id.*, 169-70)) in accordance with the shared

language model (e.g., language model used in the multimodal interpreter, limitation [1a] above) and the second model (e.g., handwriting model in the MS-TDNN-based handwriting recognizer, limitation [1a]). Decl. ¶¶92-93; Section VI.A.1.b; *Vo*, 3455, Fig. 2 (teaching output of the handwriting recognizer could be used by parser as text string for processing by “Phoenix semantic parser” because “person using our Jeanie multimodal calendar can employ” “gesturing with a pen on a touch-sensitive screen, or handwritten words to interact with the system”).

- e. [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*”

Waibel (with *Suhm*, *Vo*, and *Manke*) and *Roberts* render this limitation obvious. Decl. ¶¶94-95. As discussed for limitation [1a], *Waibel* (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious a shared language model. To the extent *Waibel* (with *Suhm*, *Vo*, and *Manke*) does not expressly disclose that this shared model is trainable, it would have been obvious in view of *Roberts. Id.*

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. Decl. ¶¶96-97. For example, after a user dictates for speech recognition, a displayed word list is generated from a large backup vocabulary “using a statistical language model and optionally word models

derived from a phonemic dictionary.” *Roberts*, Abstract. When the user chooses the correct word from the list by either “keyboard or voice command,” the language model is updated. *Id.* (“[w]hen the correct word is chosen by the user, the speech sample is used to create or update an acoustic model for the word”). “As the system is used,” it “constantly updates its statistical language model,” and “[t]he system gets more and more word models[,] and keeps improving its performance.” *Id.*; *id.*, 12:13-14:54 (describing error correction), 16:11-17 (“saves tokens for utterances when the user specifically confirmed the word to be associated with that utterance by means of a pick-choice command” and noting this “lets the user greatly reduce the chance that an incorrectly recognized utterance will be used to train”), 16:22-28, 15:58-16:28 (updating acoustic model), 23:59-25:31 (updating language model), 27:30-33. Figure 1 shows *Roberts*’s training system:

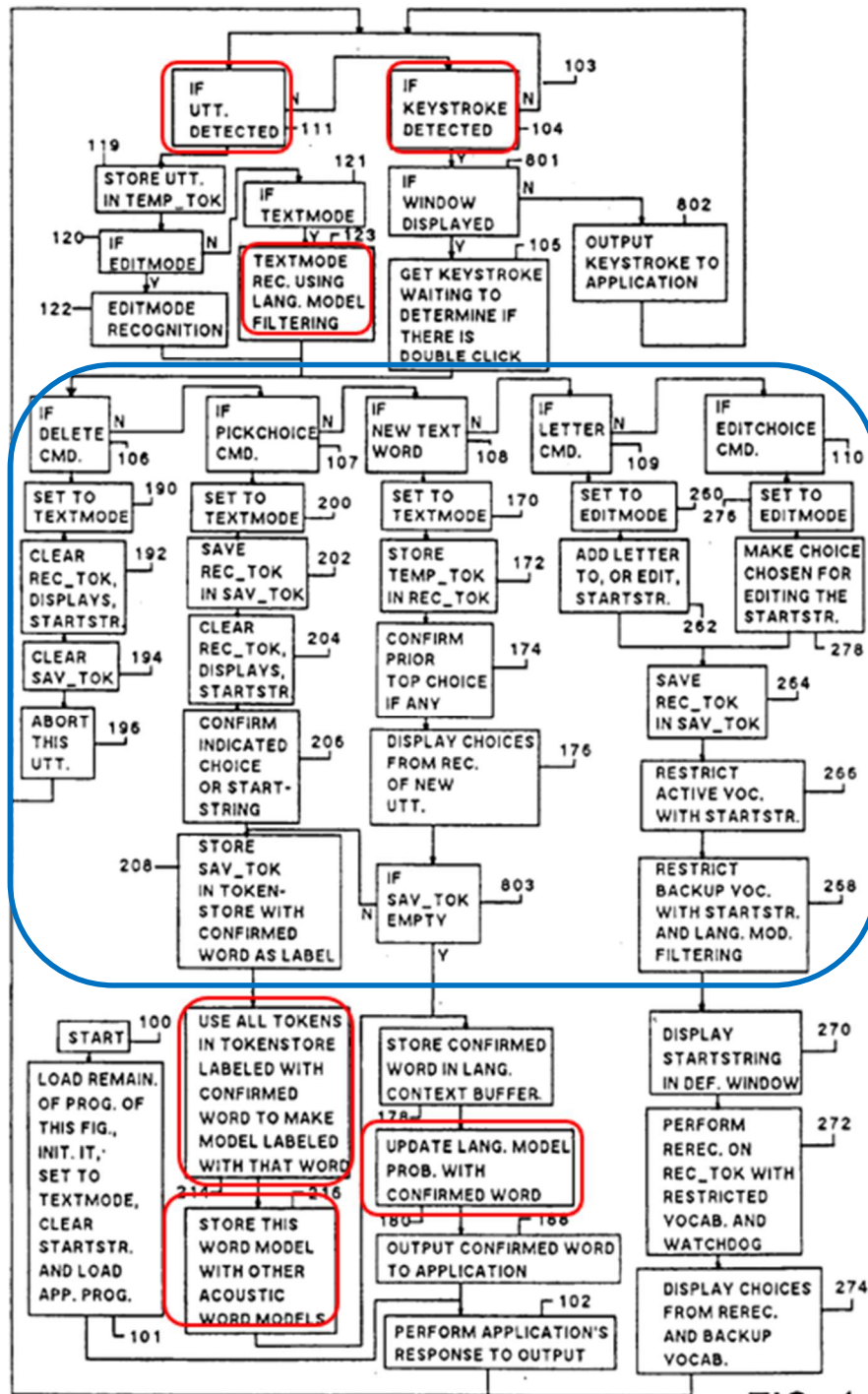


FIG. 1

Id., Fig. 1 (annotated).

As shown, *Roberts* discloses receiving speech and keystroke input (boxes 111 and 104) and using a 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.

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 a 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.

Roberts also discloses that training language models improves the accuracy of the speech recognizer. *Roberts*, 2:5-7, Abstract (“When the correct word is chosen by the user, the speech sample is used to create or update an acoustic model for the word, without further intervention by the user. . . . The system gets more and more word models and keeps improving its performance the more it is used.”), 2:17-22. As such, a POSITA would have found it obvious that the combination of *Roberts* and *Waibel* would also improve the accuracy of the speech recognizer in *Waibel*. Decl. ¶¶98-101. Moreover, because the handwriting recognizer in *Waibel* also includes the language model to also be trained and improved in view of *Roberts*, having such an improved language model leads to a more accurate handwriting recognizer. Decl. ¶¶98-101.

Compatibilities of *Waibel* and *Roberts* would have made any combination well within the grasp of a POSITA, yielding predictable results and requiring no

undue experimentation. Decl. ¶¶98-101. For example, *Waibel* teaches a standard word bigram model (*Suhm*, 222), a compatible format with *Roberts* (see *Roberts*, 13:44-60). Decl. ¶¶98-101. Further, *Waibel* discloses a similar method of error correction to that discussed in *Roberts*, including using a GUI to highlight and correct errors in speech recognition. *Id.*; *Waibel*, 168 (“[If a graphical user interface (GUI) is available, the user can simply highlight erroneous words For error correction, the user provides additional input, choosing among different correction methods” such as “repair by pen gestures (e.g., [.] to delete or insert),” or “typing or selecting among N-best alternatives”); *Roberts*, 4:15-20 (“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”). *Roberts*’s method of improving Hidden Markov Models and language models with training data also applies to *Waibel*’s recognizers because its speech and handwriting recognizers use algorithms based on Hidden Markov Models (JANUS speech translation system, NPen⁺⁺) that can be improved and adapted through additional training data. Decl. ¶¶98-101; see, e.g., *Waibel*, 167-68; *Roberts*, 2:23-25. A POSITA would have had no concern combining *Waibel* and *Roberts* and would have done so with a reasonable expectation of success. Decl. ¶¶98-101.

2. Claim 2

“The system of claim 1 wherein:

(a) the first type of message recognition is speech recognition, the first type of message input is voice input, and the first model is an acoustic model; and

(b) the second type of message recognition is handwriting recognition, the second type of message input is freehand input, and the second model is a handwriting model.”

Waibel (with *Suhm*, *Vo*, and *Manke*) discloses these additional limitations.

Decl. ¶¶102-03. As discussed for limitations 1[a] and 1[b], *Waibel* discloses its first type of message recognition is speech recognition, with the first type of message input being voice input, and the first model being an acoustic model. *E.g.*, *Waibel*, 167 (“Speech: Our speech recognition subsystem is based on the recognition front-end of the JANUS speech translation system . . . capable of processing speaker-independent, spontaneous speech.”); Decl. ¶¶102-03; *supra* Sections VI.A.1.b, VI.A.1.c. As discussed for limitations 1[a] and 1[b], *Waibel* likewise discloses the second type of message recognition is handwriting recognition, with the second type of message input being handwriting (or freehand) input. *E.g.*, *Waibel*, 167 (“MS-TDNN-based handwriting recognizer . . . capable of processing writer-independent, continuous (cursive) handwriting”); Decl. ¶¶102-03; *supra* Sections VI.A.1.b, VI.A.1.c. A POSITA would have also understood the

model used in the MS-TDNN-based handwriting recognizer is a handwriting model. Decl. ¶¶102-03; *supra* Sections VI.A.1.b, VI.A.1.c.

3. Claim 3

“The system of claim 2 further comprising a stylus, wherein the shared language model is trainable responsive to user correction of misrecognition by the first message recognizer by (a) movement of the stylus across visual indicia of dictated text that has been erroneously generated to specify a selection from the text, and (b) correction of the selection.”

Waibel (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious this additional limitation. Decl. ¶¶104-07. Further, *Waibel* and *Roberts* render obvious this limitation. *Id.*

Waibel discloses a stylus (pen) to correct misrecognitions by the recognizer, and a POSITA would have understood that includes misrecognition by the speech recognizer. *Waibel*, 168 (“[R]equiring the user to identify errors by highlighting them, . . . For error correction, we offer . . . pen gestures to indicate where to insert words, or which word to delete.”; “If a graphical user interface (GUI) is available, the user can simply highlight erroneous words . . . [and] choos[e] among different correction methods [including] . . . repair by pen gestures”); Decl. ¶¶104-07. A POSITA would have understood that *Waibel*’s correction of misrecognitions by highlighting erroneous words and “repair[ing] by pen gestures” to delete or insert words is correction of misrecognition by the speech recognizer by some (a)

movement of a stylus across visual indicia of dictated text erroneously generated to specify a selection from the text, and then (b) correction of the selection. *Id.*

Further, *Roberts* discloses a system with a shared language model trainable responsive to user correction of misrecognition by the first message recognizer (e.g., speech recognizer) by a user's speech or keyboard input correction. *Id.*; Section VI.A.1.e. Thus, a POSITA would have found it obvious to train the shared model in *Waibel* (and/or modified in view of *Roberts* (Section VI.A.1.e)) responsive to user correction of misrecognition in the manner described in *Waibel*, e.g., using a stylus. *Roberts*, 4:15-20 (“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”), Abstract (“speech sample is used to create or update an acoustic model for the word”). This would have been obvious because *Roberts* also teaches correction can be performed using different modalities, like speech or a keyboard. *Id.*, 2:48-55, Abstract (with “keyboard or voice command, the user may choose the correct word from the list or may choose to edit a similar word if the correct word is not on the list”). Decl. ¶¶104-07.

4. Claim 6

“The system of claim 1 wherein the shared language model comprises a syntactic model and a semantic model.”

Waibel (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious this additional limitation. As discussed for limitation [1a], *Waibel* (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious a shared language model in its system. *Waibel*, 168; (Section VI.A.1.b). A POSITA would have understood that such a shared language model (as used in *Waibel* and described in *Vo*) would comprise a syntactic model and a semantic model. Decl. ¶108.

First, *Waibel* discloses a bigram language model, which a POSITA would have understood is a syntactic model because it describes probabilities of words appearing in sequence, i.e., syntactic information. *Id.*; Section VI.A.1.b. Second, a semantic model is disclosed in the dialog frames described in *Vo*. *See, e.g., Vo*, 3546 (“text string 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”). As explained in connection with limitation [1a], this shared language model receives inputs from speech and handwriting recognizers and, in response to the calculated dialog state, can change the active language models in the recognizers to improve recognition. *See* Section VI.A.1.b.

Thus, a POSITA would have understood the frames are a semantic model as they relate to the meaning or understanding of the user's intent with the original input.

Decl. ¶108.

5. Independent Claim 7

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

Independent claim 7 is substantially identical to claim 1, except it is directed to a method. For the reasons discussed above for limitations [1P] and [1a], *Waibel* (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious this limitation.

Decl. ¶109; Sections VI.A.1.a, VI.A.1.b.

- b. [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;*”**

This limitation is substantially identical to limitations [1a] and [1b]. For the reasons discussed for limitations [1a] and [1b], *Waibel* (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious this limitation. Decl. ¶110; Sections VI.A.1.b, VI.A.1.c.

- c. [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”***

This limitation is substantially identical to limitations [1a] and [1c]. For the reasons discussed for limitations [1a] and [1c], *Waibel* (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious this limitation. Decl. ¶111; Sections VI.A.1.b, VI.A.1.d.

- d. [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.”***

This limitation is substantially identical to limitations [1a] and [1d]. For the reasons discussed for limitations [1a] and [1d], *Waibel* (with *Suhm*, *Vo*, and *Manke*) and *Roberts* render obvious this limitation. Decl. ¶112; Sections VI.A.1.b, VI.A.1.e.

6. Claim 8

“The method of claim 7 wherein:

(a) performing the first type of message recognition comprises performing speech recognition responsive to voice input in accordance with the shared language model and a first model that is an acoustic model; and

(b) performing the second type of message recognition comprises performing handwriting recognition responsive to freehand input in

accordance with the shared language model and a second model that is a handwriting model.”

This claim is substantially identical to claim 2. For the reasons discussed above for claim 2, *Waibel* (with *Suhm*, *Vo*, and *Manke*) discloses these additional limitations. Decl. ¶113; Section VI.A.2.

7. Claim 9

“The method of claim 8 wherein training the shared language model includes (a) moving a stylus across visual indicia of dictated text that has been erroneously generated to specify a selection from the text, and (b) correcting the selection.”

This claim is substantially identical to claim 3. For the reasons discussed above for claim 3, *Waibel* (with *Suhm*, *Vo*, and *Manke*) alone or with *Roberts* renders obvious this additional limitation. Decl. ¶114; Section VI.A.3.

8. Claim 12

“The method of claim 7 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.”

This claim is substantially identical to claim 6. For the reasons discussed above for claim 6, *Waibel* (with *Suhm*, *Vo*, and *Manke*) discloses or at least renders obvious this additional limitation. Decl. ¶115; Section VI.A.4.

B. Ground 3: Claims 4, 5, 10, and 11 Would Have Been Obvious over *Waibel* (incorporating *Suhm*, *Vo*, *Manke*), *Roberts*, and *Verrier*; or Ground 4: *Waibel*, *Suhm*, *Vo*, *Manke*, *Roberts*, and *Verrier*

1. Claim 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.”

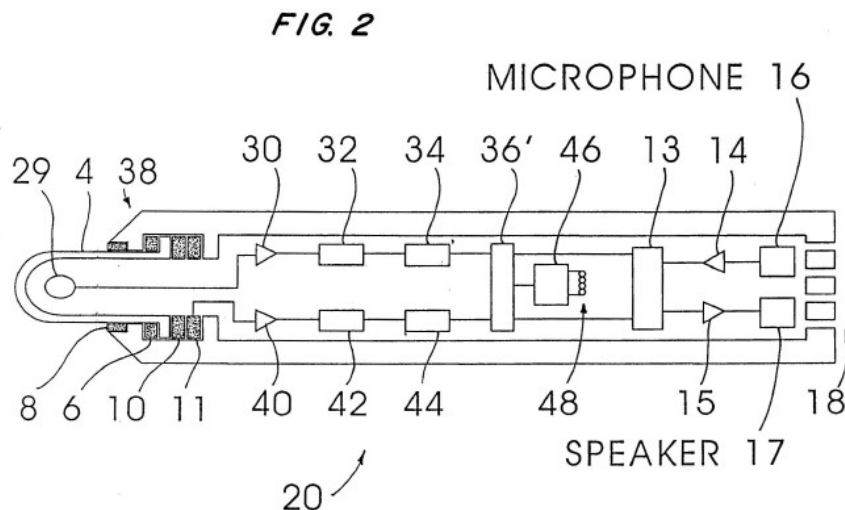
Waibel (with *Suhm*, *Vo*, *Manke*), *Roberts*, and *Verrier* render these additional limitations obvious. Decl. ¶¶116-21.

First, *Waibel* discloses or at least renders obvious a “*stylus includes a pointing interface.*” *Waibel* discloses a stylus, e.g., a pen, for handwriting input into a computer in human-computer interactions. *Id.* A POSITA would have understood that such a stylus would include a pointing interface to interact with the computer receiving the input. *Id.*; *Waibel*, 167 (“computer agents that use speech, gesture, handwriting, pointing, spelling jointly for more robust, natural and flexible human-computer interaction”; “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 [and w]e employ simple

heuristics to decide when to invoke handwriting recognition on pen input ...”),
168 (“user provides additional input, [including] repair by pen gestures (e.g. to
delete or insert), in addition to the standard repair by typing or selecting among N-
best alternatives”).

As to “*the stylus includes . . . a microphone*” and “*voice input is from the
microphone,*” the limitations are not expressly in *Waibel*. Decl. ¶¶116-21. They
were, however, well-known stylus features at the time. *Id.* They would have been
obvious to a POSITA over *Waibel* (with *Suhm, Vo, Manke*) in view of *Roberts* and
Verrier. 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.

This is shown in Figure 2:



Id., Fig. 2. It would have been obvious to a POSITA 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.*

This claim also recites "**a tablet responsive to the pointing interface.**" The '405 patent explains that a tablet can be "any suitable hardware and/or software for performing functions of, inter alia, a speech recognizer and an editor." '405 patent, 6:11-14. *Waibel* discloses such a tablet, e.g., a computer agent with hardware and/or software for performing functions of a speech recognizer and editor. Decl.

¶¶116-21. Further, *Waibel* teaches its tablet is responsive to the pointing interface of the stylus. *Waibel*, 167. *Waibel* discloses users 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.*

Finally, as to last element, “***freehand input is from the pointing interface being manually manipulated on the tablet,***” *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. Decl. ¶¶116-21.

2. Claim 5

“The system of claim 4 wherein the visual indicia of dictated text appears on the tablet.”

Waibel (with *Suhm, Vo, Manke, Roberts, and Verrier* render this additional limitation obvious. Decl. ¶¶122-24. As explained with claims 3 and 4, *Waibel* discloses the use of a tablet as a human-computer interface and display of visual indicia of dictated text on that tablet. *Waibel*, 167 (discussing speech recognition), 168 (“a graphical user interface (GUI) is available [and] the user can simply highlight erroneous words . . . displayed”; “user provides additional input, choosing among different correction methods [including] . . . repair by pen gestures (e.g., [] to delete or insert)”); Decl. ¶¶122-24. A POSITA would have

understood *Waibel* discloses using a tablet that responds to the stylus on the tablet on which visual indicia of dictated text appears. For example, it discloses a navigation system where information is displayed and modified, disclosing a display that both responds to touch input and shows visual indicia of dictated text:

[F]or navigation within visually presented image data, the user typically modifies the view by operations such as zooming and panning. *The integration of speech and gesture increases flexibility by allowing different attributes of an operation to be specified in whatever modality is more appropriate for each attribute; for instance, the effect of a spoken ‘zoom in to this <circle> area’ accompanied by a circle drawn around the desired area is much more difficult to achieve using a single input modality.*

Waibel, 169 (emphases added). A POSITA would have understood the display for the navigation system responds to the pen (e.g., circle drawing) and displays the dictated text (e.g., “integration of speech and gesture” and “spoken ‘zoom in to this <circle> area’”). Decl. ¶¶122-24. At the very least, a POSITA would have found doing so 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 cost. *Id.* And a POSITA would have a reasonable expectation of successfully doing so. *Id.*

3. Claim 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.”

For the reasons discussed above for claims 1, 4, and 7, *Waibel* (with *Suhm, Vo, Manke*), *Roberts*, and *Verrier* render obvious the additional limitations of claim 10. Decl. ¶125; Sections VI.A.1, VI.B.1, VI.A.5.

4. Claim 11

“The method of claim 10 wherein the visual indicia of dictated text appears on the tablet.”

This claim is substantially identical to claim 5. For the reasons discussed for claim 5, *Waibel* (with *Suhm, Vo, Manke*), *Roberts*, and *Verrier* render obvious this additional limitation. Decl. ¶126; Section VI.B.2.

C. Ground 5: *Roth* and *Roberts* Render Obvious Claims 1-3 and 7-9

1. Independent Claim 1

Roth in view of *Roberts* renders obvious claim 1. Decl. ¶127.

a. [1P]

If the Board finds the preamble limiting, *Roth* discloses it. Decl. ¶128. For example, *Roth* discloses a system for generating text in response to message inputs. *E.g.*, *Roth*, Fig. 1, element 28 (display of result), 8:12-13 (“may be a visual display of the match character as indicated by block 28”). *Roth* explains speech can be collected “by microphone 6 and audio storage means 7” (“*message input of a first type*”). *Id.*, Fig. 1, element 6, 6:26-29. And handwritten input can be provided by the user at “touchpad 1” (“*second type*” of message input). *Id.*, Fig. 1, element 1, 4:12-28. *Roth* explains, in response to these inputs, its system generates a “visual display of the match character” (text). *Id.*, Fig. 1, element 28, 8:3-19. These steps are shown in Figure 1:

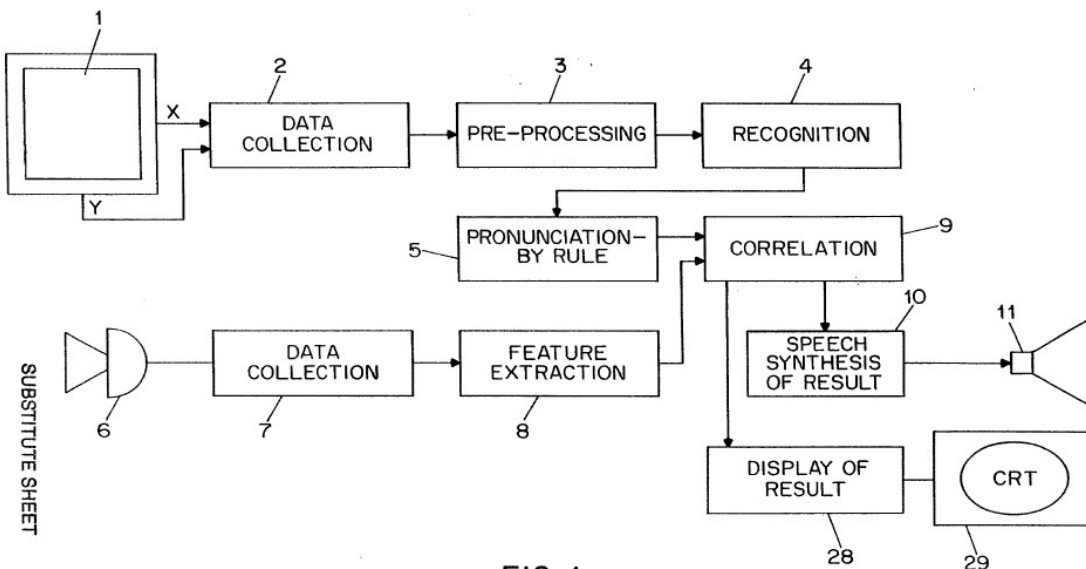


FIG. 1

Id., Fig. 1.

b. [1a]

Roth discloses this limitation. *Roth* discloses a “*unified message model*” in element 2 (data collection), element 3 (pre-processing), element 4 (recognition), element 5 (pronunciation by rule), element 7 (speech data collection), element 8 (speech feature extraction), and element 9 (correlation):

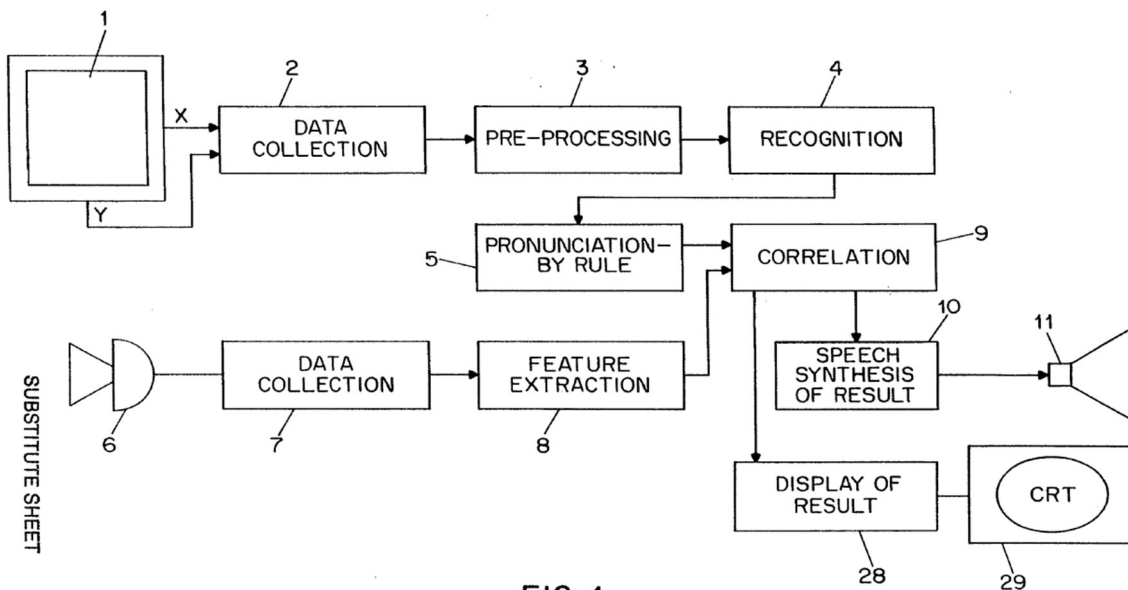


FIG. 1

Id., Fig. 1; Decl. ¶¶129-130.

i. “(1) a shared language model”

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.*

As another example, *Roth* discloses using a template store containing speech templates. *Roth*, 9:14-24, 10:13-25. This store and its templates are used for speech and handwriting recognition because, in addition to correlating speech inputs, the templates correlate word candidates obtained through the handwriting recognizer. *Id.*; Decl. ¶¶131-32. This template store and its correlation is a shared language model because it is used to characterize what may be said or written independently from the speech or the handwriting recognizer, and used to correlate the outputs of the two recognizers to achieve higher accuracy. *Id.*

To the extent Patent Owner argues *Roth* does not teach this limitation in view of limitation 1(d), it would have been obvious in view of *Roberts* and its

shared bigram language model, as explained below. Decl. ¶133; *infra* Section VI.C.1.e.

ii. “(2) a first model specific to a first type of message recognition”

Roth discloses this limitation. Decl. ¶134. *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.

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

Roth discloses or at least renders obvious this limitation. Decl. ¶135. *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.

c. [1b]

Roth discloses or at least renders obvious this limitation. Decl. ¶¶136-38.

Roth discloses a first message recognizer (elements 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.

To the extent Patent Owner argues *Roth* does not teach this limitation in view of limitation 1(d), it would have been obvious to do so in view of *Roberts*, as explained below. Decl. ¶138; *infra* Section VI.C.1.e.

d. [1c]

Roth discloses or at least renders obvious this limitation. Decl. ¶¶139-41.

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 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.

To the extent Patent Owner argues *Roth* does not teach this limitation in view of limitation 1(d), it would have been obvious to do so in view of *Roberts*, as explained in detail below. Decl. ¶141; *infra* Section VI.C.1.e.

e. [1d]

Roth discloses or at least renders obvious this limitation. To the extent Patent Owner argues that *Roth* does not teach this limitation or *Roth*'s pronunciation-by-rule/speech templates correlations are not trainable shared language models responsive to user correction of misrecognition by either the speech or writing recognizers, it would have been obvious in view of *Roberts*. Decl. ¶142.

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 to 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.

A POSITA 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 corrections. Decl. ¶¶152-53. For example, 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.*

A POSITA would also have found it obvious to use *Roberts*'s trainable bigram language model (*Roberts*, 13:51) in *Roth*'s language model to achieve the same benefits (*id.*, 2:10-30). Decl. ¶¶144-51. For example, bigram models were known to effectively reduce errors in recognizers by predicting the likelihood that a second word follows a 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, 8:17-19, 10:13-25. This further reduction of errors would reduce the time it takes to provide a correct result. Decl. ¶¶144-51. Also, although *Roth* focuses on processing single word inputs, a POSITA would have recognized that *Roberts*'

more sophisticated bigram language model would have provided additional error correction benefits for sentences which inherently provide more lexical context that bigram models use to reduce errors. *Id.* Furthermore, *Roth*'s existing ranking methodology makes it easy for an added bigram model to affect and improve the recognition accuracy. *Roth*, 7:26-28, 14:26-17:2. For example, *Roth*'s correlator can use the bigram model to increase or decrease a "confidence level weighting value" (*see, e.g., id.*, 15:4-24) assigned to each match word. Decl. ¶¶144-51. Moreover, a POSITA would have known that because *Roth*'s method of user correction (*Roth*, 8:16-19) is compatible with that in *Roberts* (*Roberts*, 4:16-20), the same method could be used to train the additional bigram model in *Roth*. *Id.*

2. Claim 2

Roth discloses these additional limitations. Decl. ¶¶154-55. *Roth* discloses the first type of message recognition is speech recognition, with the first type of message input being voice input, and the first model being an acoustic model. *Id.*; *see Roth*, 1:1-4, 13:8-14, 11:17-21; *see also supra* Sections VI.C.1.b, VI.C.1.c (discussions for elements [1a] and [1b]).

Roth also discloses the second type of message recognition is handwriting recognition, with the second type of message input being handwriting (or freehand) input. Decl. ¶¶154-55; *Roth*, 5:14-6:2, 11:3-8; *see supra* Sections VI.C.1.b, VI.C.1.d (discussions for elements [1a] and [1c]). It would have been obvious, as

noted above, for the second model to be a handwriting model. *See supra* Sections VI.C.1.b, VI.C.1.d.

3. Claim 3

Roth, alone or with *Roberts*, renders obvious this additional limitation. Decl. ¶¶156-58. *Roth* discloses a stylus to correct misrecognitions, including misrecognitions by a speech recognizer, by (a) movement of the stylus across visual indicia of dictated text that has been erroneously generated to specify a selection from text, and (b) correction of the selection. *Id.* *Roth* explains “[t]ypically an input device such as a touchpad is employed, which receives user input by a stylus (or finger).” *Roth*, 1:7-10. In the “(hopefully rare) case that the match was incorrect, the user will so indicate, either by touchpad input or by speech input.” *Id.*, 8:8-10. *Roth* also teaches “the recognition result is provided to the user on a visual display, and an opportunity is provided so [that] the user may confirm the correctness of the recognition.” *Id.*, 3:12-15. Accordingly, a POSITA would have understood, since *Roth* receives user input via stylus, *Roth* also receives user correction input through a stylus. Decl. ¶¶156-58. A POSITA would have understood the stylus, in correcting the misrecognition, moves across visual indicia of dictated text erroneously generated on a tablet (e.g., “recognition result”). *Id.*

Further, *Roberts* discloses a system with a shared language model trainable responsive to user correction of misrecognition by the first message recognizer (e.g., speech recognizer) by a user's speech or keyboard input correction. *Id.*; see *supra* Section VI.A.1.e. Thus, a POSITA would have found it obvious to train the shared model in *Roth* (and/or modified in view of *Roberts* (element [1d] above) responsive to user correction of misrecognition in the manner described in *Roth*, e.g., using a stylus. See, e.g., *Roberts*, 4:15-20; Abstract. Indeed, this would have been particularly obvious because *Roberts* also teaches that the correction in *Roberts* can be performed using different modalities, like speech or a keyboard. *Id.*, 2:48-55 (“Preferably the menu of character sequences is a list of best matching words produced by speech recognition of one of the training utterances, and the user can select and edit the desired menu choice by voice. It is also preferred that the system is used as part of a speech recognition system which produces as its output for each spoken word either the word recognized or the word edited and selected for that spoken word”), Abstract (“keyboard or voice command, the user may choose the correct word from the list or may choose to edit a similar word”). Decl. ¶¶156-58.

4. Independent Claim 7

a. [7P]

Independent claim 7 is substantially identical to claim 1, except it is directed to a method. For the reasons discussed above for limitations [1P] and [1a], *Roth* renders obvious this limitation. Decl. ¶159; *supra* Sections VI.C.1.a, VI.C.1.b.

b. [7a]

This limitation is substantially identical to limitations [1a] and [1b]. For the reasons discussed for limitations [1a] and [1b], *Roth* and *Roberts* render obvious this limitation. Decl. ¶160; *supra* Sections VI.C.1.b, VI.C.1.c.

c. [7b]

This limitation is substantially identical to limitations [1a] and [1c]. For the reasons discussed for limitations [1a] and [1c], *Roth* alone or with *Roth* and *Roberts* render obvious this limitation. Decl. ¶161; *supra* Sections VI.C.1.b, VI.C.1.d.

d. [7c]

This limitation is substantially identical to limitations [1a] and [1d]. For the reasons discussed for limitations [1a] and [1d], *Roth* alone or with *Roth* and *Roberts* render obvious this limitation. Decl. ¶162; *supra* Sections VI.C.1.b, VI.C.1.e.

5. Claim 8

This claim is substantially identical to claim 2. For the reasons discussed above for claim 2, *Roth* discloses these additional limitations. Decl. ¶163; *supra* Section VI.C.2.

6. Claim 9

This claim is substantially identical to claim 3. For the reasons discussed above for claim 3, *Roth* and *Roberts* render obvious this additional limitation. Decl. ¶164; *supra* Section VI.C.3.

D. Ground 6: *Roth*, *Roberts*, and *Verrier* Render Obvious Claims 4, 5, 10, and 11

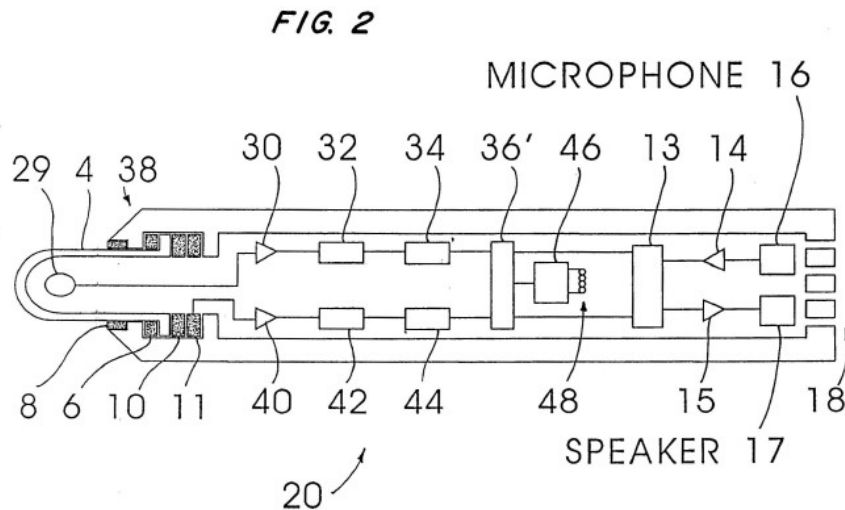
1. Claim 4

Roth, *Roberts*, and *Verrier* render these limitations obvious. Decl. ¶¶165-69.

First, *Roth* discloses “*stylus includes a pointing interface*” and “*freehand input is from the pointing interface being manually manipulated on the tablet*,” for it discloses using a stylus for freehand input on a touchpad. *Roth*, 1:7-10 (“an input device such as a touchpad is employed, which receives user input by a stylus (or finger) and provides information about the position of the stylus”); *id.*, claim 3. A POSITA would have understood this is an input from the pointing interface of a stylus being manually manipulated on the tablet. Decl. ¶¶165-68.

Though *Roth* does not expressly disclose its stylus includes “*a microphone*” or “*the voice input is from the microphone*,” incorporating a microphone in the

Roth's pen/stylus and providing voice input from the microphone would have been obvious to a POSITA. *Id.* As explained above, *Verrier* teaches a stylus with a microphone for receiving voice input. *See, e.g., Verrier*, Abstract ("microphone and speaker are included in an untethered stylus" for "communication between the user and the pen-based computer system"), 18:35-40, Fig. 2. This stylus is shown in Figure 2:



Id., Fig. 2. 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 a 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.*

Finally, *Roth* discloses “*a tablet responsive to the pointing interface.*” *Roth*, Fig. 1, element 1; 2:35-3:2 (“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”), 8:30-33 (“touchpad tablet 1 with interface 22 plugged into an expansion slot of bus 27, microphone 6 with speech recognition (feature extraction) interface 24, and an optional speaker with speech synthesis card 26”). Decl. ¶169.

2. Claim 5

Roth, *Roberts*, and *Verrier* render this additional limitation obvious. As explained for claims 3 and 4, *Roth* teaches the use of a tablet as a human-computer interface and the display of visual indicia of dictated text on that tablet. Decl. ¶¶170-71; *see Roth*, 3:12-15 (“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”), 8:11-19 (“[T]here may be a visual display of the match character as indicated by block 28. . . [T]he process continues with the next handwriting to be recognized. As with the audio synthesis technique just described, . . . the user will so indicate, either by touchpad input or by speech input to the system”). The fast and commonplace development of touch sensitive screens in portable devices (e.g., tablets) was well-known at the 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.

3. Claim 10

For the reasons discussed above for claims 1, 4, and 7, *Roth, Roberts*, and *Verrier* render obvious the additional limitations of claim 10. Decl. ¶172; *supra* Section VI.C.1, VI.D.1, VI.C.4.

4. Claim 11

This claim is substantially identical to claim 5. For the reasons discussed for claim 5, *Roth*, *Roberts*, and *Verrier* render obvious this additional limitation. Decl. ¶173; *supra* Section VI.D.2.

E. Ground 7: *Roth*, *Roberts*, and *Vo* Render Obvious Claims 6 and 12

1. Claim 6

Roth, *Roberts*, and *Vo* render obvious the shared language model comprises “*a syntactic model*” because a POSITA would have understood the disclosed bigram models describes the probabilities of words appearing in sequence, i.e., syntactic information. *Roberts*, 13:50-53 (describing word bigrams); Decl. ¶174. Although *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. *Id.*; *see also* Ex1026, 172-175 (describing syntactic and semantic models). *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”); Section VI.A.4. Even the ’405 patent admits the use of semantic and syntactic models in language models in recognizers was well known. *See, e.g.*, ’405 patent, 26:11-21.

A POSITA would have found it obvious to incorporate a semantic model in *Roth* to improve upon its recognition accuracy and “deal gracefully with ungrammatical sentences.” *Vo*, 3546; *see also supra* Section VI.A.4. And a POSITA would have had a reasonable expectation of success in doing so. Decl. ¶174.

2. Claim 12

This claim is substantially identical to claim 6. For the reasons discussed for claim 6, *Roth*, *Roberts*, and *Vo* render obvious this additional limitation. Decl. ¶175; *supra* Section VI.E.1.

VII. Noninstitution Would Be Improper Under 35 U.S.C. § 314

Non-institution would be improper under *Fintiv*. The ’405 patent has been asserted in four litigations, three of which have settled as of the filing of this Petition. *Infra* Section X.B. While Google is not a named party in any, products incorporating its Gboard application were accused in the fourth litigation against Motorola in the Northern District of Illinois (“NDIL”). As of the filing of this Petition, NDIL had not entered a scheduling order, discovery had not started, and the case had been stayed since December 2022. Ex1024. Yesterday, the Court ruled on Motorola’s motion to dismiss. In NDIL, the median time to trial is under 50 months, and stays pending IPRs are granted over 70% of the time. Ex1025.

These factors, and the strength of the arguments in this Petition, show a denial under Section 314 would be improper.

VIII. Noninstitution Would Be Improper Under 35 U.S.C. § 325(d)

The Board considers the following factors in deciding whether to exercise its discretion under § 325(d)—(1) whether the same or substantially the same prior art and/or arguments were previously presented to the PTO; and (2) if so, whether the Petitioner demonstrated the PTO erred in a manner material to patentability. *See Advanced Bionics, LLC v. MED-EL Elektromedizinische Geräte GmbH*, IPR2019-01469, Paper 6 (P.T.A.B. Feb. 13, 2020). Neither factor favors denying institution

None of *Waibel*, *Suhm*, *Vo*, *Manke*, *Roberts*, or *Roth* were evaluated by the Examiner. '405 patent, References Cited; e.g., Ex1002. Although another prior-art reference attributed to *Waibel*, U.S. Patent No. 5,855,000, was considered by the Examiner, it describes a generalized system, which is not the same as the specific system in *Waibel*. Decl. ¶45. But even if it described the same system as *Waibel*, it was not considered with *Roberts*, the combination proposed herein. As such, the Examiner did not consider the same or substantially the same arguments and art. *Versa Prods. v. Varidesk, LLC*, IPR2020-00387, Paper 13, at 17-18 (PTAB, July 10, 2020).

Further, Applicant's comments distinguishing the '000 *Waibel* patent over its claims were misleading at best. For example, Applicant claimed it "does not

believe that *Waibel* teaches or suggests any type of language model . . . shared between the recognizers for the primary and secondary input signals” Ex1002 at 55-56. But as relevant to the claims, a POSITA would have understood the experiments in the ’000 *Waibel* patent included different language models shared between recognizers for the primary and secondary input signals. Decl. ¶45; *see, e.g.*, ’000 patent, 15:43-51, 15:61-64, 6:28-65, 7:19-26. For example, a POSITA would have understood that the purpose of the language models in the ’000 patent is to facilitate recognition, and in a system where inputs, including speech and handwriting, are considered primary or secondary inputs, ’000 patent, 5:39-41, it would be a poor design to use the language model for one input modality and not the other. Decl. ¶45. In fact, when recognition is constrained to vocabulary and grammar of a particular task, a POSITA would have found it counterproductive to apply one model to one recognizer, but a different model to another. *Id.*

IX. Mandatory Notices

A. Real Party-in-Interest

Google LLC is the real party-in-interest for this petition.⁴

Out of an abundance of caution and without conceding that it is a real party in interest, Google discloses that Motorola Mobility LLC was named as a defendant in a litigation based, in part, on alleged activities relating to products that incorporate Google's Gboard application. Whether such a third party is a real party in interest need not be resolved here because it has no impact on this proceeding, including because even a determination that it is a real party in interest would not result in the application of any estoppel or time-bar under 35 U.S.C. § 315. *Unified Patents, LLC v. MemoryWeb, LLC*, Decision Granting Director Review Vacating-in-part the Final Written Decision and Vacating Board Order, IPR2021-01413, Paper 76 at 5 (PTAB May 22, 2023) (citing *SharkNinja Operating LLC v. iRobot Corp.*, IPR2020-00734, Paper 11 at 18 (PTAB Oct. 6, 2020) (precedential)); *see also* Ex1028 (Motorola was served on September 12, 2022).

⁴ Google LLC is a subsidiary of XXVI Holdings Inc., which is a subsidiary of Alphabet Inc. XXVI Holdings Inc. and Alphabet Inc. are not real parties in interest to this proceeding.

B. Related Matters

To Petitioner's best knowledge, the '405 patent is or has been involved in the following district court litigations:

Name	Number	Court	Filed
<i>Buffalo Patents, LLC v. ZTE Corp.</i>	6-22-cv-00423	W.D. Tex.	April 27, 2022
<i>Buffalo Patents, LLC v. TCL Tech. Grp. Corp.</i>	6-22-cv-00422	W.D. Tex.	April 27, 2022
<i>Buffalo Patents, LLC v. OnePlus Tech. (Shenzhen) Co.</i>	6-22-cv-00421	W.D. Tex.	April 27, 2022
<i>Buffalo Patents, LLC v. Motorola Mobility LLC</i>	1-22-cv-04540	N.D. Ill.	August 25, 2022

C. Lead and Back-Up Counsel, and Service Information

Lead Counsel	Back-Up Counsel
Erika H. Arner, Reg. No. 57,540 erika.arners@finnegan.com Finnegan, Henderson, Farabow, Garrett & Dunner, LLP 1875 Explorer Street Reston, VA 20190-6023 Tel: 571-203-2700 Fax: 202-408-4400	Sydney R. Kestle, Reg. No. 78,725 sydney.kestle@finnegan.com Jia Lu, Reg. No. 61,543 jia.lu@finnegan.com Finnegan, Henderson, Farabow, Garrett & Dunner, LLP 901 New York Avenue, NW Washington, DC 20001-4413 Tel: 202-408-4000 Fax: 202-408-4400

Please address all correspondence to lead and back-up counsel at the addresses shown above. Petitioner also consents to electronic service by email.

X. Grounds for Standing

Petitioner certifies the '405 patent is available for IPR and Petitioner is not barred or estopped from requesting an IPR challenging the claims on the grounds identified herein.

XI. Conclusion

The Board should institute and cancel each challenged claim. The Office may charge any required fees to Deposit Account No. 06-0916.

Date: September 12, 2023

Respectfully submitted,

/Erika H. Arner/

Erika H. Arner, Lead Counsel

Reg. No. 57,540

CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. § 42.24(a)(1)(i), the undersigned hereby certifies that the foregoing PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 6,904,405 contains 13,479 words, excluding parts of this Petition exempted under § 42.24(a), as measured by the word-processing system used to prepare this paper.

/Erika H. Arner/
Erika H. Arner, Lead Counsel
Reg. No. 57,540

CERTIFICATE OF SERVICE

The undersigned hereby certifies that a copy of the above PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 6,904,405 and accompanying EXHIBITS 1001-1029 and power of attorney were served September 12, 2023, via FedEx Priority Overnight on the following address of record for the subject patent:

Dean Munyon
KOWERT, HOOD, MUNYON, RANKIN & GOETZEL P.C.
1120 S. Capital of Texas Hwy
Building 2, Suite 300
Austin, TX 78746

A courtesy copy has been mailed to litigation counsel at the following:

Matthew James Antonelli
Antonelli, Harrington & Thompson LLP
4306 Yoakum Blvd, Suite 450
Houston, TX 77006

Date: September 12, 2023

By: /Lisa C. Hines/
Lisa C. Hines
Senior Litigation Legal Assistant
FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER, LLP