

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

BUFFALO PATENTS LLC,
Patent Owner.

Case No. IPR2023-01387
U.S. Patent No. 8,204,737

**DECLARATION OF CHRISTOPHER SCHMANDT IN SUPPORT OF
PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 8,204,737**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. QUALIFICATIONS	2
III. MATERIALS CONSIDERED	6
IV. LEGAL STANDARDS	8
V. LEVEL OF ORDINARY SKILL IN THE ART	11
VI. CLAIM CONSTRUCTION	12
VII. BACKGROUND OF THE '737 PATENT	14
A. Overview of the '737 Patent.....	14
B. Earliest Available Priority Date	16
VIII. GROUNDS OF UNPATENTABILITY.....	16
IX. CLAIMS 1-16 and 19 WOULD HAVE BEEN OBVIOUS TO A POSITA BASED ON THE PRIOR ART IN GROUNDS 1-8.....	18
A. Ground 1: Claims 1, 4, 12, 13, 15, 16, and 19 Would Have Been Obvious Over <i>Ditzik</i> , <i>Tchorzewski</i> , and <i>Roberts</i>	18
1. Overview of <i>Ditzik</i>	18
2. Overview of <i>Tchorzewski</i>	21
3. Overview of <i>Roberts</i>	22
4. Independent Claim 1	23
5. Claim 4.....	46
6. Claim 12	47
7. Independent Claim 13	49
8. Claim 15	51

9.	Claim 16.....	53
10.	Independent Claim 19.....	53
B.	Ground 2: Claims 2, 3, 5, 6, 11, and 14 Would Have Been Rendered Obvious Over <i>Ditzik, Tchorzewski, Roberts, and Verrier</i>	57
1.	Overview of <i>Verrier</i>	57
2.	Motivation to Combine	59
3.	Claim 2	60
4.	Claim 3	65
5.	Claim 5	67
6.	Claim 6	69
7.	Claim 11	70
8.	Claim 14	72
C.	Ground 3: Claims 9 and 10 Would Have Been Obvious Over <i>Ditzik, Tchorzewski, Roberts and Bellegarda</i>	74
1.	Overview of <i>Bellegarda</i>	74
2.	Motivation to Combine	74
3.	Claim 9.....	76
4.	Claim 10	78
D.	Ground 4: Claims 7 and 8 Would Have Been Obvious Over <i>Ditzik, Tchorzewski, Roberts, and Campbell</i>	81
1.	Overview of <i>Campbell</i>	81
2.	Claim 7: “The system of claim 1, wherein the system is configured to select the particular user message model based at least in part on an identifying signal received from the microphone.”	81

3.	Claim 8: “The system of claim 7, wherein the system is configured to access the particular user message model via a wide area network.”	84
E.	Ground 5: Claims 1, 2, 4, 12, 13, 16, 19 Would Have Been Obvious Over <i>Root</i> , <i>Roberts</i> , and <i>Lewis</i>	84
1.	Overview of <i>Root</i>	84
2.	Overview of <i>Lewis</i>	85
3.	Claim 1	85
4.	Claim 2	97
5.	Claim 4: “The system of claim 1, wherein the system is further configured to train the particular user message model by updating the acoustic model.”	99
6.	Claim 12: “The system of claim 1, wherein the particular user message model includes user-specific acoustic model information and user-specific handwriting model information.”	100
7.	Claim 13	101
8.	Claim 16: “The method according to claim 13, further comprising training the selected user message model.”	103
9.	Claim 19	104
F.	Ground 6: Claims 3, 5, 6, 11, and 14 Would Have Been Rendered Obvious Over <i>Root</i> , <i>Roberts</i> , <i>Lewis</i> , and <i>Verrier</i>	107
1.	Claim 3: “The system of claim 2 wherein the microphone is housed in the stylus.”	107
2.	Claim 5: “The system of claim 1, further comprising a proximity sensor configured to sense a proximity of a user to the microphone.”	108
3.	Claim 6: “The system of claim 1, further comprising an identification circuit configured to identify an owner of	

	the stylus, wherein the system is configured to select the particular user message model at least in part in response to the identification circuit.”	108
4.	Claim 11: “The system of claim 1, further comprising an activation circuit configured to sense movement of the stylus toward the tablet surface, wherein the system is further configured to receive the freehand input in response to the activation circuit sensing the movement.”	109
5.	Claim 14: “The method according to claim 13, further comprising switching from receiving audio input to receiving freehand input in response to detecting movement of the stylus towards a tablet.”	109
G.	Ground 7: Claims 9 and 10 Would Have Been Obvious Over <i>Root, Roberts, Lewis, and Bellegarda</i>	110
1.	Claim 9: “The system of claim 1, wherein the system adjusting the language model includes adjusting a syntactic model and a semantic model included in the language model.”	110
2.	Claim 10	110
H.	Ground 8: Claims 7, 8, and 15 Would Have Been Obvious Over <i>Root, Roberts, Lewis, and Campbell</i>	111
1.	Claim 7: “The system of claim 1, wherein the system is configured to select the particular user message model based at least in part on an identifying signal received from the microphone.”	111
2.	Claim 8: “The system of claim 7, wherein the system is configured to access the particular user message model via a wide area network.”	111
3.	Claim 15: “The method according to claim 13, further comprising accessing the selected user message model via a network.”	113
X.	Conclusion	114

'737 PATENT CLAIMS

1. A system for generating text responsive to voice and handwriting input, comprising:
 - a tablet surface, wherein the system is configured to receive freehand input based at least in part on manipulation of a stylus relative to the tablet surface; and
 - a microphone for receiving voice input;
wherein the system is configured to:
 - select a particular user message model from a plurality of user message models;
 - adjust a language model of a local message model based at least in part on the particular user message model, wherein the local message model includes the language model, a handwriting model, and an acoustic model;
 - generate text data based at least in part on the freehand input, the adjusted language model, and the handwriting model; and
 - generate text data based at least in part on the voice input, the adjusted language model, and the acoustic model.
2. The system of claim 1, wherein the system is further configured to:
 - identify information relating to a particular user;
 - wherein the particular user message model is selected based at least in part on the information relating to the particular user; and,
 - wherein the plurality of user message models includes two or more user message models corresponding to users that are different from the particular user.

3. The system of claim 2 wherein the microphone is housed in the stylus.
4. The system of claim 1, wherein the system is further configured to train the particular user message model by updating the acoustic model.
5. The system of claim 1, further comprising a proximity sensor configured to sense a proximity of a user to the microphone.
6. The system of claim 1, further comprising an identification circuit configured to identify an owner of the stylus, wherein the system is configured to select the particular user message model at least in part in response to the identification circuit.
7. The system of claim 1, wherein the system is configured to select the particular user message model based at least in part on an identifying signal received from the microphone.
8. The system of claim 7, wherein the system is configured to access the particular user message model via a wide area network.
9. The system of claim 1, wherein the system adjusting the language model includes adjusting a syntactic model and a semantic model included in the-language model.
10. The system of claim 9, wherein the semantic model comprises at least one semantic cluster selected from the group consisting of:
 - a semantic cluster of a first type that is a mathematically derived group of vectors representing text segments;
 - a semantic cluster of a second type that is a discrete grouping of text having a particular context; and
 - a semantic cluster of a third type that is defined with respect to an individual reference word that is indicative of a particular context.
11. The system of claim 1, further comprising an activation circuit configured to sense movement of the stylus toward the tablet surface,

wherein the system is further configured to receive the freehand input in response to the activation circuit sensing the movement.

12. The system of claim 1, wherein the particular user message model includes user-specific acoustic model information and user-specific handwriting model information.

13. A method, comprising:

receiving, at a device, freehand input based at least in part on manipulation of a stylus;

receiving, at the device, audio input based at least in part on information received at a microphone;

selecting, via the device, a selected user message model from a plurality of user message models;

adjusting, via the device, a language model of a local message model, wherein the local message model includes the language model, a handwriting model, and an acoustic model, wherein the adjusting the language model is based at least in part on the selected user message model;

generating, by the device, converted handwriting text data based at least in part on the freehand input, the adjusted language model, and the handwriting model; and

generating, by the device, converted speech text data based at least in part on the audio input, the adjusted language model, and the acoustic model.

14. The method according to claim 13, further comprising switching from receiving audio input to receiving freehand input in response to detecting movement of the stylus towards a tablet.

15. The method according to claim 13, further comprising accessing the selected user message model via a network.

16. The method according to claim 13, further comprising training the selected user message model.

19. An article of manufacture including a non-transitory computer-readable medium having stored thereon instructions executable by a computing device to cause the computing device to perform operations comprising:

receiving freehand input based at least in part on manipulation of a stylus;

receiving audio input based at least in part on information received at a microphone;

selecting a particular user message model from a plurality of user message models;

adjusting a language model of a local message model based at least in part on the particular user message model, the local message model including the language model, a handwriting model and an acoustic model;

generating converted handwriting text data based at least in part on the freehand input, on the adjusted language model, and on the handwriting model; and

generating converted speech text data based at least in part on the audio input, on the adjusted language model, and on the acoustic model.

I. INTRODUCTION

1. I, Christopher Schmandt, submit this declaration to state my opinions on the matters described below.

2. I have been retained on behalf of Google LLC to serve as an expert in this proceeding before the United States Patent & Trademark Office.

3. I am over 21 years of age and, if I am called upon to do so, I would be competent to testify as to the matters set forth herein.

4. My compensation is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of any proceeding.

5. I understand that this proceeding involves U.S. Patent No. 8,204,737 which is entitled “Message Recognition Using Shared Language Model” (“the ’737 patent”), and I have been asked to provide my opinions as to the patentability or unpatentability of the claims of the ’737 patent.

6. I understand the ’737 patent has been provided for as Ex. 1001. I have reviewed and am familiar with the ’737 patent. I understand that the file history has been provided as Ex. 1002. I have reviewed and am familiar with the file history of the ’737 patent.

7. This declaration sets forth my opinions, which I formed based on my study of the evidence; my understanding as an expert in the field; and my training, education, research, knowledge, and personal and professional experience.

8. Throughout this declaration, I refer to specific pages, figures, and/or line numbers of various exhibits. These citations are illustrative and are not intended to suggest that they are the only support for the propositions for which they are cited.

II. QUALIFICATIONS

9. I retired four years ago after a 40-year career at the Massachusetts Institute of Technology (MIT); for most of that time I was employed as a Principal Research Scientist at the Media Laboratory. In that role I also served as faculty for the MIT Media Arts and Sciences academic program. I was a founder of the Media Laboratory, a research lab which now spans two buildings.

10. I received my B.S. degree in Electrical Engineering and Computer Science from MIT in 1978, and my M.S. in Visual Studies (Computer Graphics) also from MIT the following year. I was employed at MIT since 1980, initially at the Architecture Machine Group which was an early computer graphics and interactive systems research Lab. In 1985, I helped found the Media Laboratory and continued to work there until retirement. I was director of a research group titled "Living Mobile." My research spanned distributed communication and

collaborative systems, with an emphasis on multi-media and user interfaces, with a strong focus on speech-based systems and portable devices. I have over 70 published conference and journal papers and one book in the field of speech technology and user interaction.

11. For the first fifteen years of my career, my research emphasized speech recognition and speech user interfaces. I built the first system utilizing multimodal user input – speech and gesture – starting in 1980 (“Put That There”); this remains known as seminal work in the field. Put That There was also the first conversational computer system utilizing speech recognition and synthesis to engage in a voice user interaction. I continued to innovate speech user interfaces using recognition, text-to-speech synthesis, and recorded audio in a wide variety of projects. I built one of the first graphical user interfaces for audio editing, employing text keyword recognition on voice memos in 1982 (Intelligent Ear). I built the first research-grade unified messaging system, which combined text and voice messages into a single inbox, with speech recognition over the phone for remote access, and a graphical user interface for desktop access in 1983 (Phone Slave). I built a speech recognizer to detect acoustic correlates of mouth configuration to drive a real time animated character (Lip Sync, 1983). Along with my students we built the first system for real time spoken driving directions, including speech-accessible maps of Cambridge, Massachusetts in 1987 (Back

Seat Driver). We built some of the earliest speech-based personal assistants for managing messages, calendar, contacts, etc. (Conversational Desktop 1985, Chatter 1993, MailCall 1996). We applied speech recognition to large bodies of everyday conversations captured with a wearable device and utilized as a memory aid (Memory Prosthesis 2004). We used speech recognition on radio newscasts to build a personalized version of audio newscasts (Synthetic News Radio, 1999) and also investigated adding speech recognition to a mouse-based window system a few years earlier.

12. With the appearance of handheld devices, some of which were capable of recognizing hand writing, I used and built applications for these. The first was the Apple Newton, for which we built a voice based reminder system with multiple “to do” categories and time compressed speech playback (VoiceNotes, 1992). We built quite a few systems employing speech recognition in handheld mobile devices (ComMotion 1999, Nomadic Radio 2000, Impromptu 2001, and Symphony 2004, for example). We applied speech recognition to large bodies of everyday conversations captured with a wearable device and utilized as a memory aid (Memory Prosthesis 2004). In this context we worked with devices such as the Palm Pilot, which used Graffiti for handwriting recognition, the HP Ipaq handheld computer, pagers, flip phones, and finally smart phones.

13. I was later awarded the prestigious Association for Computing Machinery (ACM) Computer Human Interface (CHI) Academy membership specifically for those years of work pioneering speech user interfaces.

14. In the course of my research, I built a number of speech recognition client/server distributed systems, with the first being in 1985. Much of the initial motivation for a server architecture was that speech recognition required expensive digital signal processing hardware that we could not afford to put on each computer, so a central server with the required hardware was used. Later versions of the speech recognition server architecture allowed certain computers to perform specialized tasks serving a number of client computers providing voice user interfaces, either on screens or over telephone connections.

15. Because of my early work with distributed speech systems, I served for several years in the mid-1990s with a working group on the impact of multimedia systems on the Internet reporting to the Internet Engineering Task Force (IETF) and later the Internet Activities Board (IAB). This work impacted emerging standards such as Session Initiation Protocol (SIP).

16. In my faculty position I taught graduate level courses in speech technology and user interaction design, and directly supervised student research and theses at the Bachelors, Masters, and PhD level. I oversaw the Masters and PhD thesis programs for the entire Media Arts and Sciences academic program

during my more senior years. I also served on the Media Laboratory intellectual property committee for many years, including a period during which I was the chair.

III. MATERIALS CONSIDERED

17. In forming my opinions, I have reviewed the following documents, and any other document cited in this declaration:

Exhibit	Description
Ex. 1001	U.S. Patent No. 8,204,737 to Suominen (“the ’737 patent”)
Ex. 1002	Prosecution History of U.S. Patent No. 8,204,737
Ex. 1005	U.S. Patent No. 6,167,376 to Ditzik (“ <i>Ditzik</i> ”)
Ex. 1006	U.S. Patent No. 5,027,406 to Roberts (“ <i>Roberts</i> ”)
Ex. 1007	Canadian Patent No. 1,335,002 (“ <i>Tchorzewski</i> ”)
Ex. 1008	European Patent Application Publication No. 0 622 724 A2 (“ <i>Verrier</i> ”)
Ex. 1009	Bellegarda, et al., “A Novel Word Clustering Algorithm Based on Latent Semantic Analysis,” IEEE, 1996 (“ <i>Bellegarda</i> ”)
Ex. 1010	Campbell, “Speaker Recognition: A Tutorial,” IEEE, 1997 (“ <i>Campbell</i> ”)
Ex. 1011	U.S. Patent No. 5,600,781 to Root (“ <i>Root</i> ”)
Ex. 1012	U.S. Patent Publication No. 2001/0053978 (“ <i>Lewis</i> ”)
Ex. 1013	A. Waibel et al., “Multimodal Interfaces for Multimedia Information Agents,” IEEE (1997) (“ <i>Waibel</i> ”)

Exhibit	Description
Ex. 1014	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> ”)
Ex. 1015	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> ”)
Ex. 1016	B. Suhm et al., “JANUS: Towards Multilingual Spoken Language Translation,” Proc. ARPA SLT Workshop 95 (Austin, Texas) (“ <i>Suhm</i> ”)
Ex. 1017	U.S. Patent No. 5,857,099 (“ <i>Mitchell</i> ”)
Ex. 1018	U.S. Patent No. 5,839,106 (“’106 patent”)
Ex. 1021	U.S. Patent No. 5,459,798 (“ <i>Bailey</i> ”)
Ex. 1026	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> ”)

18. I have also relied on my education, experience, research, training, and knowledge in the relevant art, and my understanding of any applicable legal principles described in this declaration.

19. All of the opinions contained in this declaration are based on the documents I reviewed and my knowledge and professional judgment. My opinions have also been guided by my understanding of how a person of ordinary skill in the art would have understood the claims of the ’737 patent at the time of the alleged

invention. For purposes of this declaration, I have been asked to assume that the date of the alleged invention is the earliest claimed priority date: July 17, 1999.

20. I reserve the right to supplement and amend any of my opinions in this declaration based on documents, testimony, and other information that becomes available to me after the date of this declaration.

IV. LEGAL STANDARDS

21. For purposes of this declaration, I have been asked to opine only on issues regarding obviousness under 35 U.S.C. § 103. I have been informed of the following legal standards, which I have applied in forming my opinions.

22. I have been asked to provide my opinions as to whether the cited prior art teaches or renders obvious claims 1-16 and 19 of the '737 patent from the perspective of a person of ordinary skill in the art at the '737 patent's priority date in 1999, as described in more detail below.

23. For purposes of this declaration, I have been informed and understand certain aspects of the law as it relates to my opinions.

24. I have been advised and understand that there are two ways in which prior art may render a patent claim unpatentable. First, I have been advised that the prior art can "anticipate" a claim. Second, I have been advised that the prior art can make a claim "obvious" to a person of ordinary skill in the art. I understand that for

an invention claimed in a patent to be patentable, it must not be anticipated and must not be obvious based on what was known before the invention was made.

25. I have been advised and understand the information used to evaluate whether an invention was new and not obvious when made is generally referred to as “prior art.” I understand that prior art includes patents and printed publications that existed before the earliest filing date of the patent (which I have been informed is also called the “effective filing date”). I have been informed and understand that a patent or published patent application is prior art if it was filed before the earliest filing date of the claimed invention and that a printed publication is prior art if it was publicly available before the earliest filing date.

26. I have been advised and understand that a dependent claim is a patent claim that refers back to another patent claim. I have been informed and understand that a dependent claim includes all of the limitations of the claim to which it refers.

27. I have been advised and understand that a patent claim may be invalid as obvious under 35 U.S.C. § 103 if the differences between the subject matter claimed and the prior art are such that the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time the invention was made. I have also been advised that several factual inquiries underlie a determination of obviousness. These inquiries include (1) the scope and content of

the prior art, (2) the level of ordinary skill in the field of the invention, (3) the differences between the claimed invention and the prior art, and (4) any objective evidence of non-obviousness (which I have been informed may also be called “secondary considerations”).

28. I have also been advised and understand that, when obviousness is based on a combination of references, that party must identify a reason why a person skilled in the art would have been motivated to combine or modify the asserted references in the manner recited in the claims and to explain why one skilled in the art would have had a reasonable expectation of success in making such combinations or modifications.

29. I have been advised and understand that the law permits the application of “common sense” in examining whether a claimed invention would have been obvious to a person skilled in the art. For example, I have been advised that combining familiar elements according to known methods and in a predictable way may suggest obviousness when such a combination would yield nothing more than predictable results. I understand, however, that a claim is not obvious merely because every claim element is disclosed in the prior art and that a party asserting obviousness must still provide a specific motivation to combine or modify the references as recited in the claims and explain why one skilled in the art would have reasonably expected to succeed in doing so.

30. I have been advised and understand that two references are considered to be in the same field of art when the references are either (1) in the same field of endeavor, regardless of the problems they address, or (2) reasonably pertinent to the particular problem being solved by the inventor in his patent.

31. I am not aware of any evidence of secondary considerations that would support a determination of non-obviousness of the claimed subject matter in the '737 patent.

32. I have been informed that in *inter partes* review proceedings, such as this one, the party challenging the patent bears the burden of proving unpatentability by a preponderance of the evidence. I understand that a preponderance of the evidence means “more likely than not.”

V. LEVEL OF ORDINARY SKILL IN THE ART

33. In rendering the opinions set forth in this declaration, I have been asked to consider the '737 patent's claims and the prior art through the eyes of a person of ordinary skill in the art (which I may also refer to as “one skilled in the art,” “skilled artisan,” “POSITA,” or similar variation). I have considered factors such as the educational level and years of experience of those working in the pertinent art, the types of problems encountered in the art, the teachings of the prior art, patents and publications of other persons or companies, and the sophistication of the technology.

34. I have been instructed to assume a person of ordinary skill in the art is not a specific real individual, but rather a hypothetical individual having the qualities reflected by the factors discussed above.

35. Taking these factors into consideration, it is my opinion that a person of ordinary skill in the art of the '737 patent as of its filing date, would have at least a minimum of 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. Work experience can substitute for formal education and additional formal education can substitute for work experience.

36. By July 17, 1999, I was at least a person of ordinary skill in the art, as shown by my qualifications and work experience above.

37. In this declaration, and for all of my opinions herein, I have applied the knowledge of a person of ordinary skill in the art as of July 17, 1999.

VI. CLAIM CONSTRUCTION

38. I have been instructed that the words of a claim are typically given their ordinary and customary meaning, as they would have been understood by a person of ordinary skill in the art at the time of the invention, in view of the intrinsic record (discussed below). In this case, I have been instructed to assume

that the “time of the invention” for purposes of claim construction is July 17, 1999, which is the earliest claimed priority filing date for the ’737 patent. The opinions herein pertain to that time frame, except where expressly stated otherwise.

39. I have been instructed that the “intrinsic record” includes the patent itself, including the claims, description, and figures (Ex. 1001), and the patent’s prosecution history—i.e., the record of proceedings at the U.S. Patent and Trademark Office (“Patent Office”) concerning the patent (Ex. 1002). I understand that, like the claims and written description, the prosecution history provides evidence to a person of ordinary skill in the art of how the inventor intended his patent to be understood, and how the Patent Office understood the patent. I understand that the inventor is permitted to apply a special definition to the terms or to limit the scope of claim terms in his patent claims, which may differ from the term’s plain and ordinary meaning. That special definition or limitation on scope may be provided in the patent’s written description, the patent’s prosecution history, or both.

40. I understand that claim interpretation may also be informed by “extrinsic evidence” (that is, evidence outside of the patent record itself). I have been informed that extrinsic evidence may include dictionaries, technical treatises, and other materials evidencing the meaning of a claim term and the understanding held by a POSITA in the relevant time period.

41. I have been asked for purposes of this declaration to apply the plain and ordinary meaning of the claim terms as they would have been understood by a person of ordinary skill in the art at the earliest claimed priority date of the '737 patent (July 17, 1999).

42. I understand that no claim term's construction is in dispute at this time. Based on my review of these materials and my personal knowledge and experience, I have considered each term of the '737 patent as it would have been understood by one skilled in the art as of July 17, 1999.

VII. BACKGROUND OF THE '737 PATENT

A. Overview of the '737 Patent

43. The '737 patent relates to a system for generating text in response to voice and handwriting inputs. ('737 patent, 5:61-63.) The system includes a microphone that receives voice inputs, and a tablet surface that receives freehand inputs from a stylus. (*Id.*, 6:2-6.) These voice and handwriting inputs are respectively recognized by a speech and a handwriting recognizer in the system. (*See, e.g., id.*, 22:10-13.)

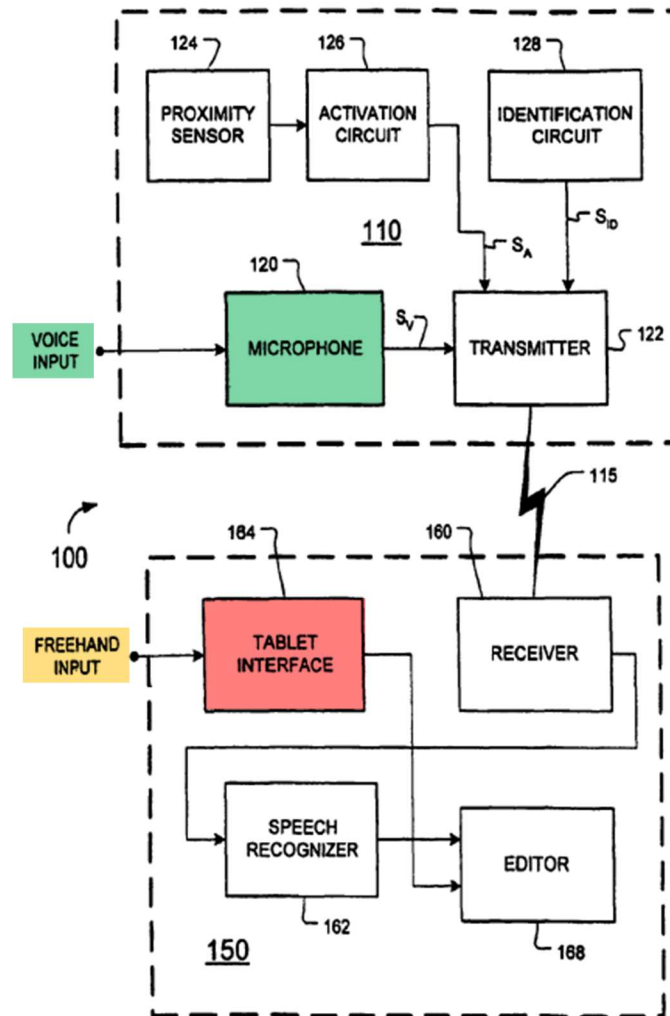


FIG. 1A

(*Id.*, FIG. 1A (annotated).)

44. The '737 patent also explains its system contains message models, which are “model[s] used in recognition of a message,” namely the speech or freehand inputs noted above. (*Id.*, 13:1-4.) A message model can be any “algorithm mathematical expression, set of rules, etc.” for “defining operation of a message recognizer to provide text data responsive to message input” (*id.*, 22:14-25). A

message model “may include a language model as well as an acoustic and/or handwriting model.” (*Id.*, 13:3-4.)

45. The patent further explains that its system specifically comprises a plurality of user message models and a local message model. A local message model is “any message model that is local to a message recognizer operating in accordance with the message model,” and it does not require any “communications outside its local processing environment.” (*Id.*, 23:17-26; *see id.* (“For example, a message model implemented in the same memory space, virtual machine, or CPU as a message recognizer is said to be local to the message recognizer.”)) A “user message model” is “a message model that is specific to a particular user.” (*Id.*, 12:66-13:4.) According to the patent, a “message model is said to be specific to a particular user when th[at] message model may be adapted by training input from the user,” such as through a user’s “corrections of misrecognized message input.” (*Id.*, 23:31-36; *see also id.*, 23:36-25:29.)

B. Earliest Available Priority Date

46. I understand that the application that matured into the '737 patent was filed March 22, 2005. It claims priority to U.S. Provisional Application No. 60/144,481, filed on July 17, 1999.

VIII. GROUNDS OF UNPATENTABILITY

47. Based on my review of the materials set forth above, including my application of the knowledge of a person of ordinary skill in the art, it is my opinion that claims 1-16 and 19 of the '737 patent would have been obvious to a person of ordinary skill in the art.

48. In particular, it is my opinion that claims 1-16 and 19 of the '737 patent would have been obvious to a person of ordinary skill in the art based on the combinations shown below:

Grounds of Unpatentability	
Ground 1:	<i>Ditzik, Tchorzewski, and Roberts</i> render obvious claims 1, 4, 12, 13, 15, 16, and 19.
Ground 2:	<i>Ditzik, Tchorzewski, Roberts and Verrier</i> render obvious claims 2, 3, 5, 6, 11, and 14.
Ground 3:	<i>Ditzik, Tchorzewski, Roberts, and Bellegarda</i> render obvious claims 9 and 10.
Ground 4	<i>Ditzik, Tchorzewski, Roberts, and Campbell</i> render obvious claims 7 and 8.
Ground 5	<i>Root, Roberts, and Lewis</i> render obvious claims 1, 2, 4, 12, 13, 16, and 19.

Grounds of Unpatentability	
Ground 6	<i>Root, Roberts, Lewis, and Verrier</i> render obvious claims 3, 5, 6, 11, and 14.
Ground 7	<i>Root, Roberts, Lewis, and Bellegarda</i> render obvious claims 9 and 10.
Ground 8	<i>Root, Roberts, Lewis, and Campbell</i> render obvious claims 7, 8, and 15.

IX. CLAIMS 1-16 and 19 WOULD HAVE BEEN OBVIOUS TO A POSITA BASED ON THE PRIOR ART IN GROUNDS 1-8

A. Ground 1: Claims 1, 4, 12, 13, 15, 16, and 19 Would Have Been Obvious Over *Ditzik, Tchorzewski, and Roberts*

49. In my opinion, claims 1, 4, 12, 13, 15, 16, and 19 would have been obvious to a POSITA based on *Ditzik* in view of *Roberts* and *Tchorzewski*.

1. Overview of *Ditzik*

50. *Ditzik* discloses a “computer system with speech recognition systems and handwriting recognition systems” that receive “speaker-dependent and speaker-independent” speech input through a microphone and handwriting input through a stylus. (*Ditzik*, Abstract.) It converts “this information to computer recognizable text, which may be displayed onto a display device.” (*Id.*) *Ditzik* also discloses correcting speech recognition errors, such as “via a pen input device”

“converted to text and graphics,” and integrating that pen information “into speech recognized text” to be “stored in a data storage system.” (*Id.*) *Ditzik*’s “system may be adapted to detect a person’s spoken speech patterns (audio data) via a **microphone**” and other components. (*Ditzik*, 3:11-13.) This is shown below in

Figure 3:

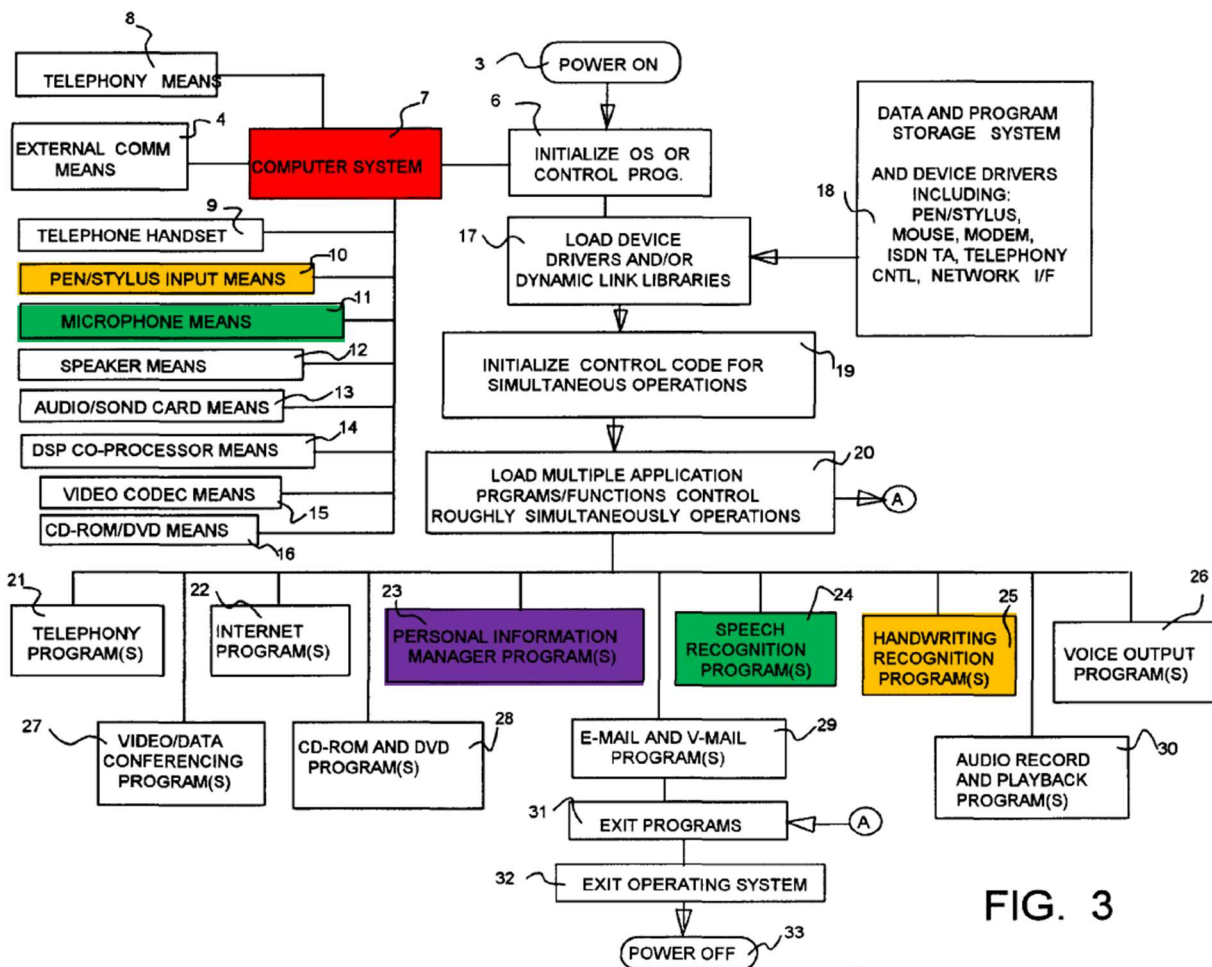


FIG. 3

(*Ditzik*, FIG. 3 (annotated).)

51. *Ditzik* explains that its system also includes **phonological rules database 60** and **phoneme model database 62, which store** phoneme models;

language modeling program 58 and dictionary and grammar databases 64, which store language models for speech recognition; and handwriting stroke/pattern database dictionary 86 and linguistic rules database 90, which store handwriting models and language models for handwriting recognition. (*Ditzik*, FIGS. 1-2, 3:52-64, 5:28-53.) This is shown below in Figures 1 and 2:

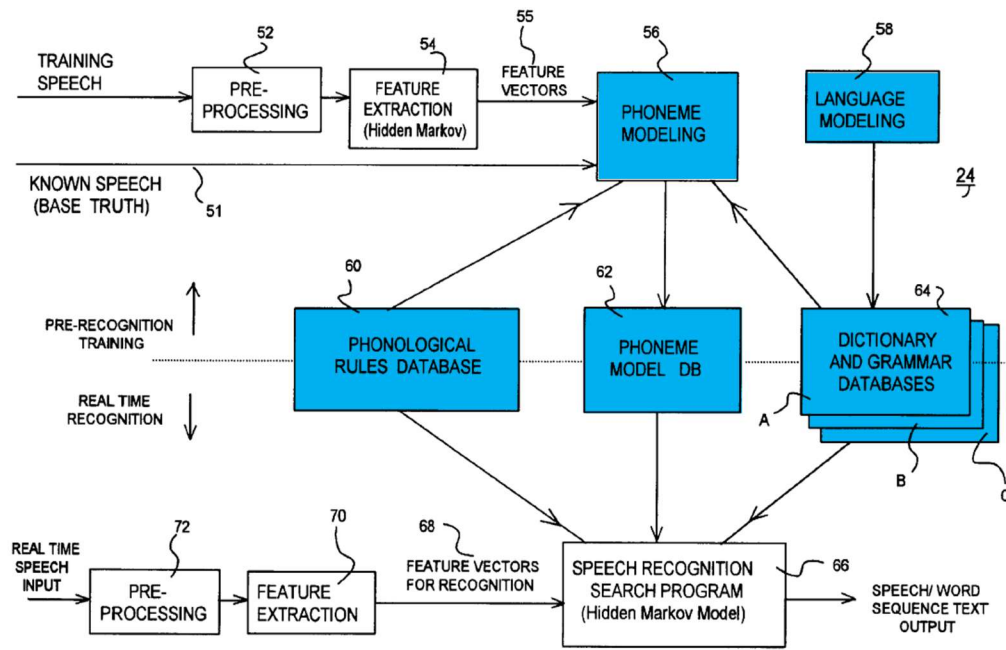


FIG. 1

PRIOR ART

(*Ditzik*, FIG. 1 (annotated).)

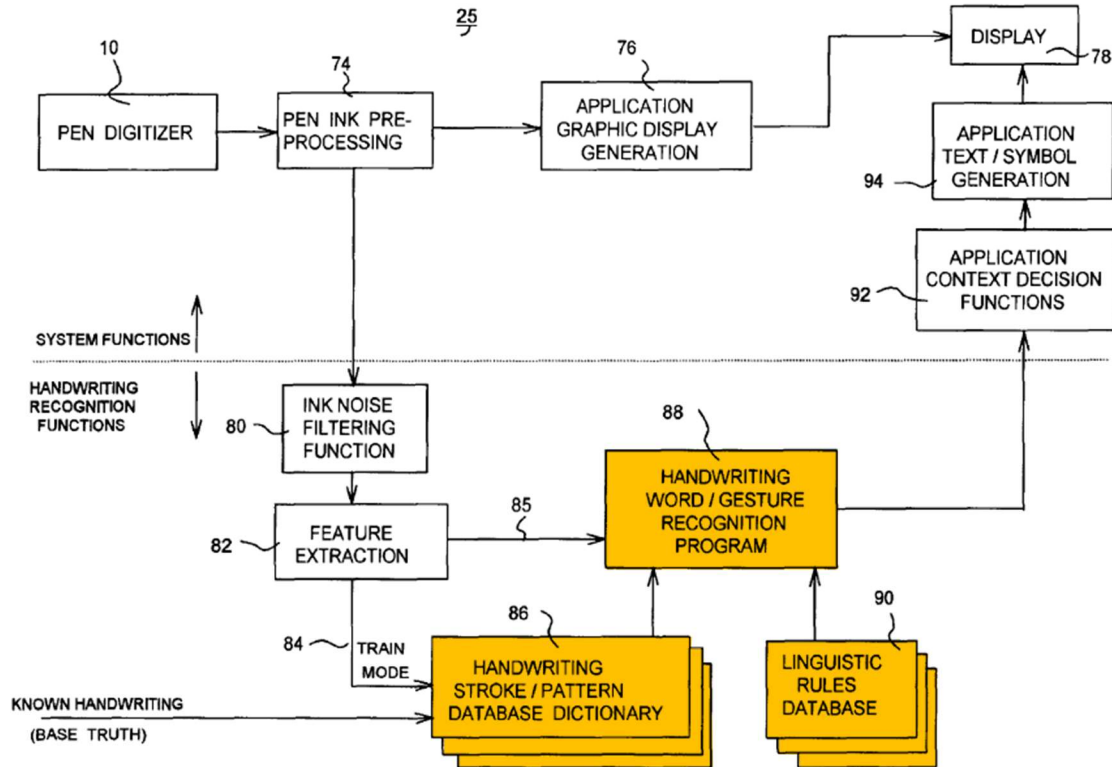


FIG. 2

(Ditzik, FIG. 2 (annotated).)

2. Overview of *Tchorzewski*

52. *Tchorzewski* discloses “a speech recognition system capable of dealing with a plurality of users” by training “the recogni[z]er with his or her own voice.” (See, e.g., *Tchorzewski*, 1:1-8.) *Tchorzewski* generates speaker-dependent “patterns” and stores them as “‘templates’ in Random Access Memory (RAM) within the speech recogni[z]er.” (*Id.*, 1:8-9.) *Tchorzewski* also discloses using “a common database” “containing the templates of all the users,” and “[w]hen any particular user activates his extension, one of the . . . recogni[z]ers is arbitrarily selected and assigned to his use, and his templates are then loaded into that

recogni[z]er.” (*Tchorzewski*, 2:1-7; *see also id.*, 2:23-3:5.) “When a user terminal is activated, one of the recognition means is assigned to deal with voiced instructions input via that user terminal,” and the database receives an instruction, via the recognizer, “to provide a copy of relevant speaker-dependent templates to the assigned recognition means.” (*Tchorzewski*, 2:23-3:5.)

3. Overview of *Roberts*

53. *Roberts* discloses a “method for creating word models for a large vocabulary, natural language dictation system.” (*Roberts*, Abstract.) *Roberts* uses language models for speech recognition and discloses “such language models are well-known to those skilled in the art of natural language speech recognition” and its invention “would work with any of them.” (*Roberts*, 13:47-50.)

54. If a user uses a vocabulary not already in the vocabulary database, *Roberts*’s system “displays a list of the words in the active vocabulary which best match the spoken word,” and a user can select a displayed word using a keyboard or a voice command. (*Roberts*, 13:47-50.) When a user corrects a speech recognition error by selecting an alternative word choice, *Roberts*’s system “uses the confirmed word, if any, to update the language model used by the recognition system.” (*Roberts*, 13:44-47.) It “constantly updates its statistical language model” when it receives “more word models and keeps improving its performance the more it is used.” (*Roberts*, Abstract.) This is shown in Figures 29-31:

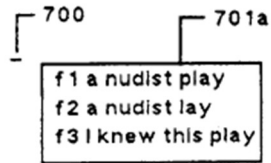


FIG. 29

User says "a new display"

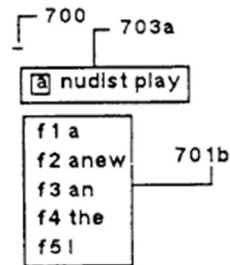


FIG. 30

User double-clicks f1

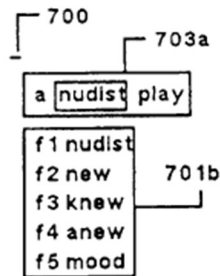


FIG. 31

User presses right arrow

(Roberts, FIGS. 29-31.)

4. Independent Claim 1

- a. 1[pre] "A system for generating text responsive to voice and handwriting input, comprising:"

55. In my opinion, *Ditzik* discloses 1[pre]. For example, *Ditzik* discloses "a computer system with speech recognition systems and handwriting recognition systems . . . that work closely together to improve the total recognition accuracy of each system alone." (*Ditzik*, Abstract.) The speech recognition and handwriting systems are shown below in Figures 1 and 2:

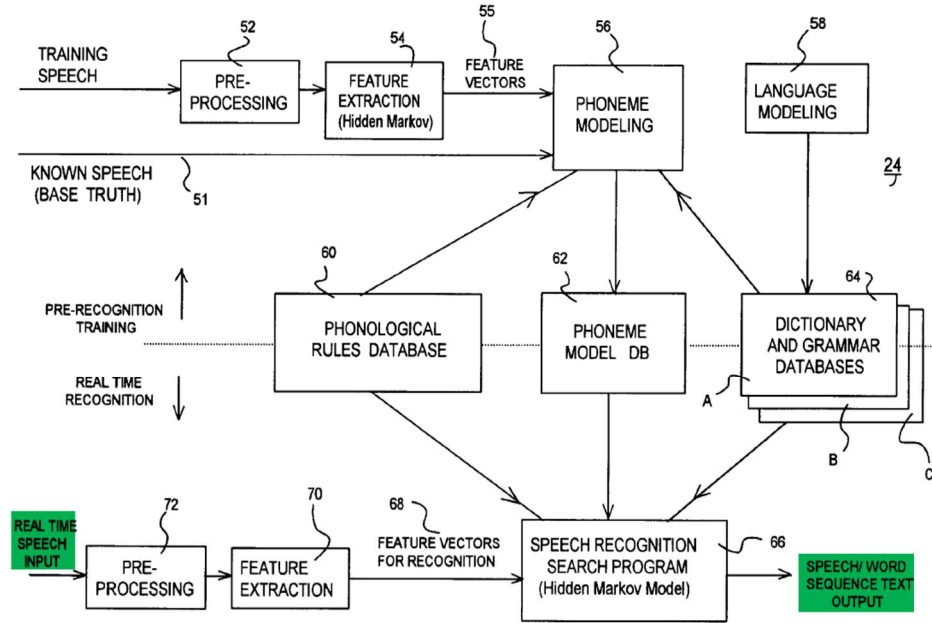


FIG. 1

PRIOR ART

(Ditzik, FIG. 1 (annotated).)

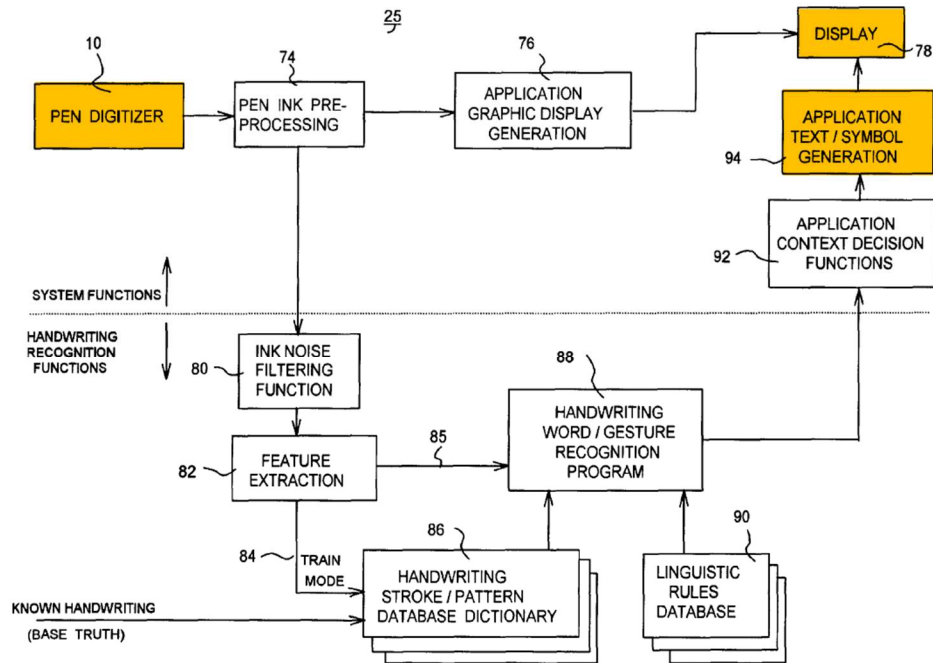


FIG. 2

(Ditzik, FIG. 2 (annotated).)

56. *Ditzik* explains that its speech recognition system “capture[s] and record[s] verbal information” at the speech recognizer and converts that information “into text and/or graphic data formats.” (*Ditzik*, 1:13-14.) Its handwriting recognition system “may transform hand written letters, numbers, punctuation, and symbols into recognizable computer generated text.” (*Ditzik*, 4:32-35.)

- b. 1[a] “a tablet surface, wherein the system is configured to receive freehand input based at least in part on manipulation of a stylus relative to the tablet surface; and”

57. In my opinion, *Ditzik* discloses this limitation. *Ditzik*’s system includes “a pen/stylus input means 10(*Ditzik*, 6:21-24; FIG. 3.) The pen/stylus input means includes a pen/stylus and “an opaque graphic tablet or transparent device, placed over a display screen.” (*Id.*) The tablet “captures a handwriting or graphic symbols from a person” when the pen/stylus is moved across its surface, and it digitizes that “pen stroke data” (“st[r]oke contours, intersections, and shape features, which may differentiate one handwritten word or mark”) on its display screen. (*Id.*, 4:29-41; FIG. 4.) These features are shown in Figures 3 and 4:

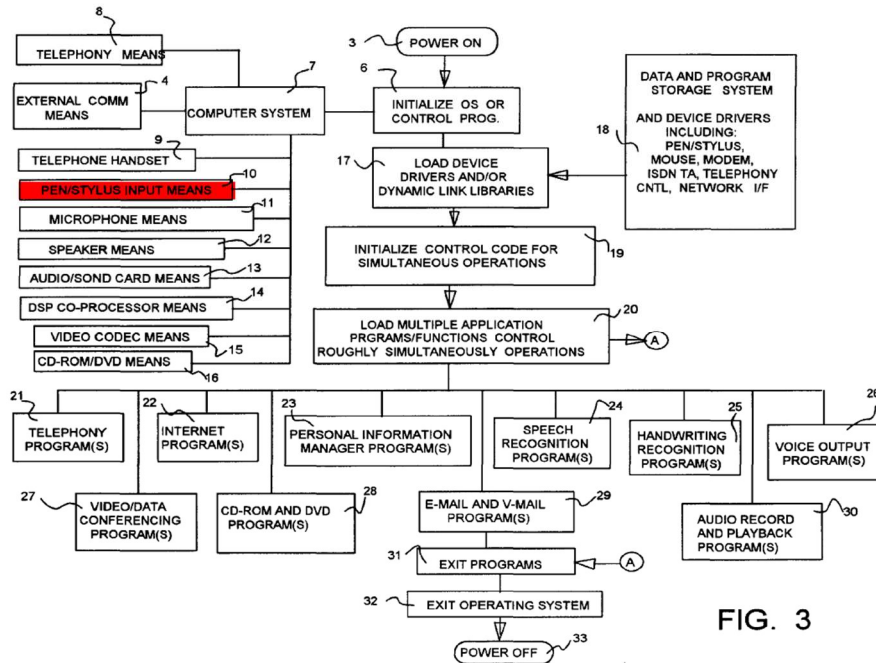


FIG. 3

(Ditzik, FIG. 3 (annotated).)

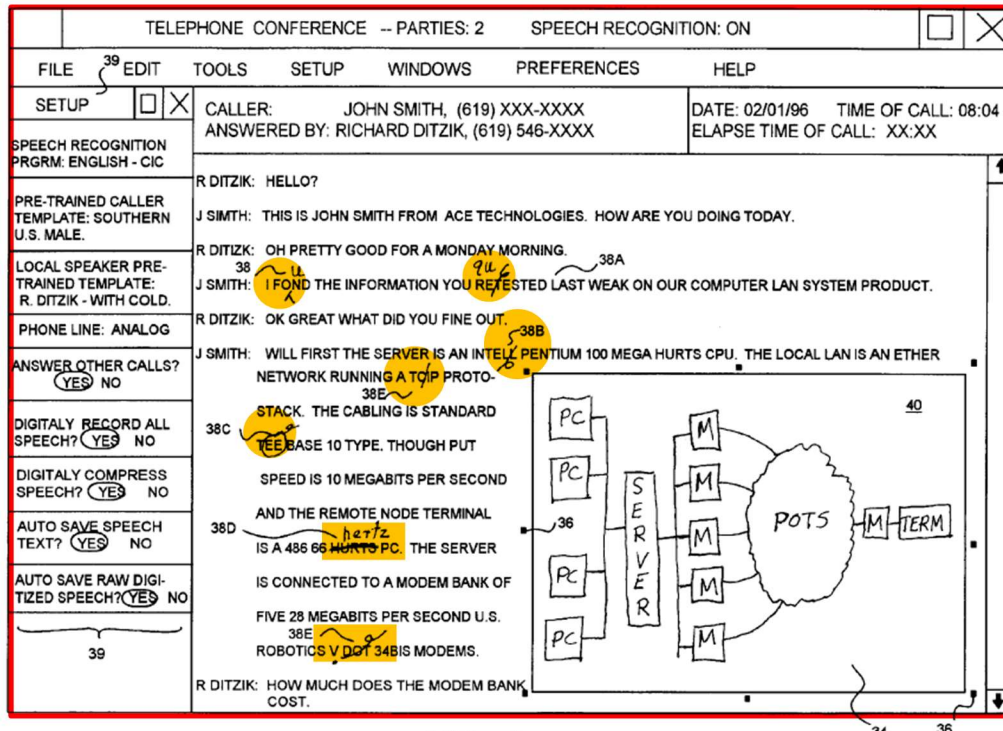
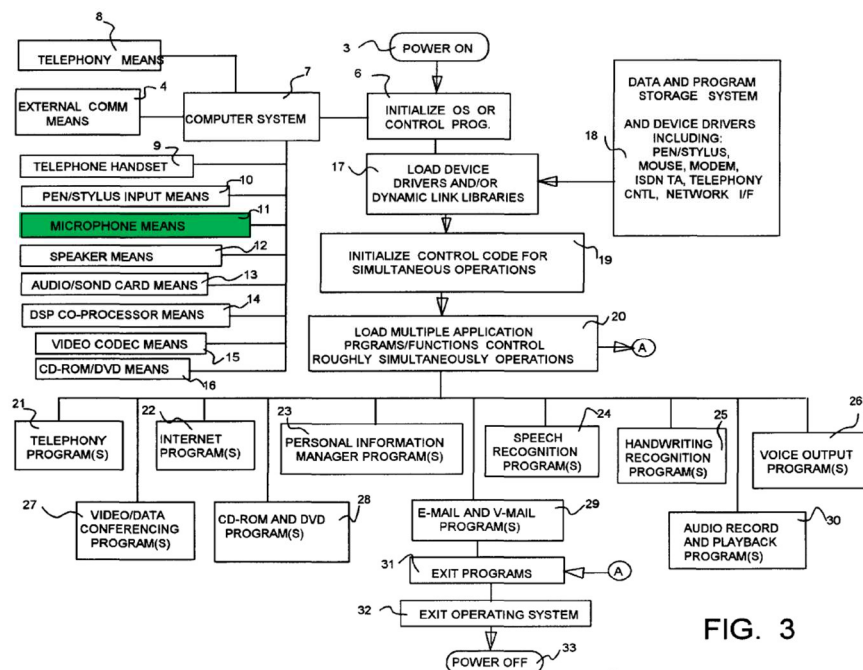


FIG. 4

(Ditzik, FIG. 4 (annotated).)

c. 1[b] “a microphone for receiving voice input;”

58. In my opinion, *Ditzik* discloses this limitation. *Ditzik*’s system comprises “microphone means 11” to receive voice input. (*Ditzik*, 6:24-31.) *Ditzik* discloses “speech input recognition program/functions 24 and associated hardware” that include “several functions[,] including speech capture (via microphone or telephone line), digitize data, process data, store raw speech data, and storage of recognized text.” (*Ditzik*, 6:59-63.) *Ditzik*’s “microphone means 11” receives voice input (*id.*, 6:24-31) and is shown below in Figure 3:



(*Ditzik*, FIG. 3 (annotated).)

d. 1[c] “wherein the system is configured to: select a particular user message model from a plurality of user message models;”

59. I understand that the specification of the '737 patent defines a user message model as “a message model that is specific to a particular user,” ('737 patent, 12:66-67), and a message model as including “a language model as well as an acoustic model and/or handwriting model” (*id.*, 13:3-4.) *Ditzik* alone or in view of *Tchorzewski* renders this limitation obvious.

60. *Ditzik* discloses a user message model. First, *Ditzik* discloses a user-specific language model that is “specific to a particular user” through training. ('737 patent, 12:66-67, 13:3-4.) *Ditzik* discloses a user adaptive recognition system where the computer is “adapted to detect a person’s spoken speech patterns” and handwriting patterns. (*Ditzik*, 3:11-16, 5:28-37.) In addition to stating that the system “adapts,” *Ditzik* discloses that the system can utilize either speaker independent or speaker dependent speech recognition. As was well known, speaker dependent recognition is called as such precisely because it must be trained to work with a particular user. Further, the description of “training mode” for handwriting demonstrates that a particular user can update the models for handwriting recognition as well. This includes a “language modeling program 58” that “input[s] data to the word and grammar dictionary 64” that in turn “provide[s] data to the phoneme modeling program 56.” (*Ditzik*, 3:59-62.) *Ditzik* explains that

the word and grammar dictionary 64 in language modeling program 58 contains language models with “the rules of allowable word sequences” and “several general or specialized vocabularies” where “specialized for several identified speech recognition applications.” (*Ditzik*, 3:52-58.) Such language models were well known at the time in speech recognition. (*Roberts*, 13:47-50; *Starner*, 1-3.) *Ditzik*’s reference to “specialized vocabularies” indicates that its language models are trained based on user-specific data.

61. To the extent the claim requires a language model for handwriting recognition, in my opinion, *Ditzik* discloses “linguistic rules database 90” and “multiple rules databases,” which include “different rules for different written languages and/or regional subsets” that serve as language models used for handwriting recognition. (*Ditzik*, 5:50-53.) The linguistic rules database may be a special rules database including “multiple rules databases, for example, different written languages and/or regional subsets.” (*Ditzik*, 2:51-52, 5:50-53.) Thus, a POSITA would have understood that the handwriting recognition language model can be specialized for each user through training.

62. Although not required by the claims of the ’737 patent, to the extent that Patent Owner argues the language models for speech recognition and handwriting recognition are shared, a POSITA would have found it obvious to modify the language modeling program 58, word and grammar dictionary 64, and

linguistic rules database 90 so that they share the same grammatical and linguistic rules to harmonize *Ditzik*'s speech and handwriting recognition. A POSITA would have done so with a reasonable expectation of success given that a shared language model was a well-known feature in the art (*see, e.g., 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); *Bailey*, 9:30-39, 10:18-25, 10:30-44, 10:59-11:34, 12:2-11, 12:53-55), and applying it to *Ditzik*'s system would have simply involved combining known elements according to known speech and handwriting recognition methods to ensure consistency between handwriting and speech recognition.

63. Second, *Ditzik* also discloses a user-specific acoustic model. *Ditzik*'s speech recognition system includes “a **phonological rules database 60** [that] may also provide data to the **phoneme modeling program**” 56.¹ (*Ditzik*, 3:59-64.) The phoneme modeling function 56 inputs acoustic data to phoneme model database 62, which stores **phoneme models**, which are a type of acoustic model. (*Ditzik*,

¹ *Ditzik* calls element 56 a “phoneme modeling function” and “phoneme modeling program” interchangeably. A POSITA would have likewise understood these terms are interchangeable.

3:37-39.) *Ditzik* discloses that these acoustic models are user specific because they can be trained using “[s]peech patterns from one or more persons.” (*Ditzik*, 3:19-24.) Further, as noted above, *Ditzik* provides for speaker dependent speech recognition, which is always trained by a specific user. *Ditzik*’s user-specific acoustic model is shown below in Figure 1:

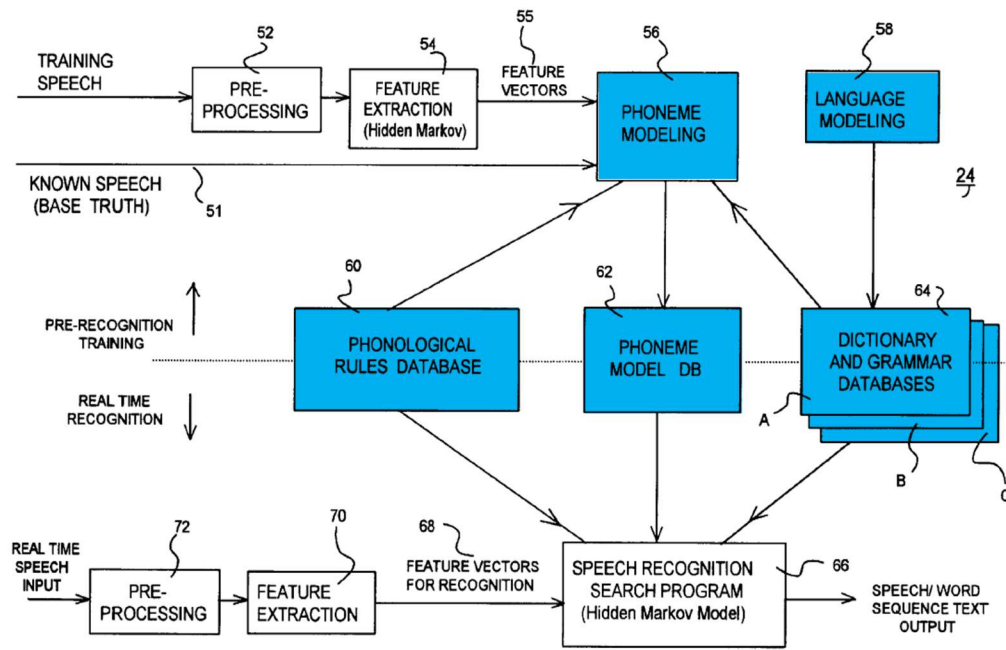


FIG. 1

PRIOR ART

(*Ditzik*, FIG. 1 (annotated).)

64. Last, *Ditzik* discloses a user-specific handwriting model. *Ditzik* discloses that, if its “system is in the training mode, data 84 is input to a handwriting stroke and/or pattern database 86,” which contains handwriting models, for example, “a plurality of hand writing/sketch patterns for one or more

languages and/or special symbol applications.” (*Ditzik*, 5:28-53.) *Ditzik* further discloses that the database 86 “may also accept handwriting data from a known source with known writing samples, which represent a set of base truth data.” (*Ditzik*, 5:33-35.) Using handwriting models, such as those stored in database 86, for handwriting recognition was well known at the time of the ’737 patent. (*Lewis*, [0020]-[0022], FIG. 1; *Root*, 3:66-4:3, 5:44-6:22, FIGS. 1-6.) These trained handwriting models are user specific because they reflect different user’s specific handwriting and sketch patterns. These user-specific handwriting models are shown below in Figure 2:

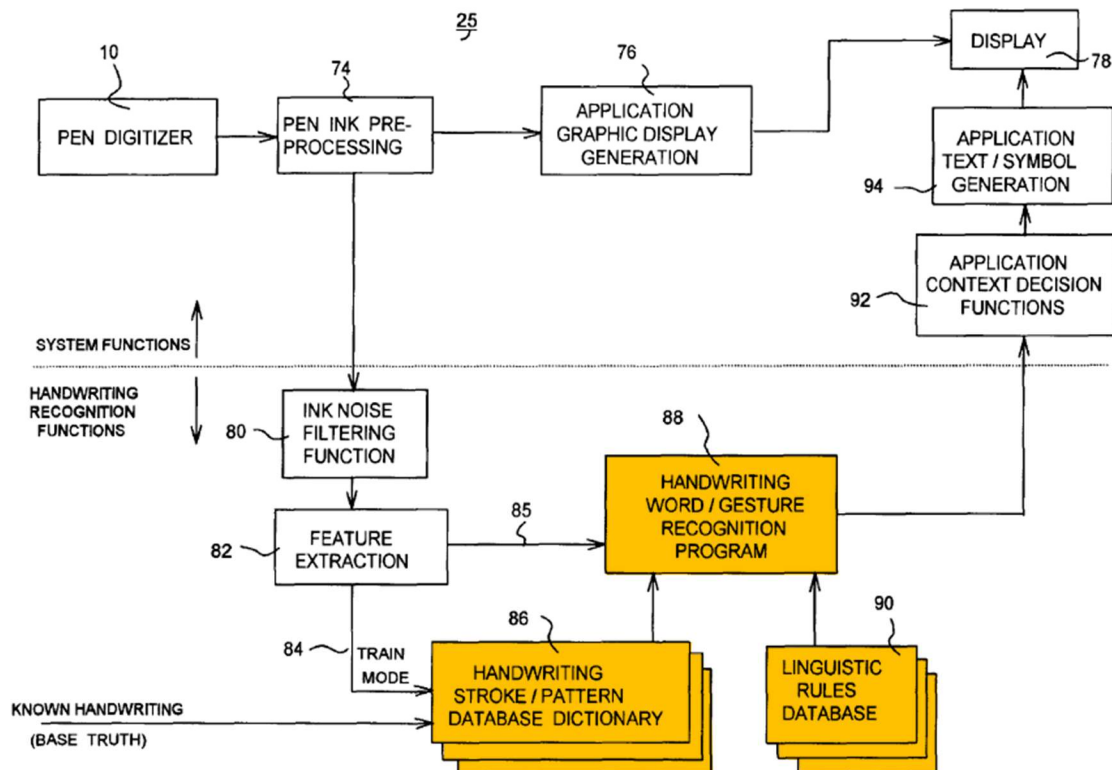


FIG. 2

(*Ditzik*, FIG. 2 (annotated).)

65. Because *Ditzik* discloses a user-specific “language model as well as acoustic model and/or handwriting model,” it discloses a [user message model](#).

66. *Tchorzewski* discloses a [plurality of user message models](#), each tailored to a specific user, and selecting a particular [user message model](#) from that plurality. *Tchorzewski* discloses “a speech-recognition system for the provision of a speech recognition facility to a multiplicity of user terminals.” (*Tchorzewski*, 2:23-30.) The speech recognition facility includes “a database containing respective templates for each of the user terminals.” (*Id.*) Each template includes a particular user’s “voice patterns” (acoustic models) used to “interpret voiced instructions.” (*Tchorzewski*, 1:8-14.) *Tchorzewski* discloses selecting a particular user message model from a plurality of user message models. When a user terminal is activated, a speech recognizer in the speech recognition facility is assigned to that user terminal. (*Tchorzewski*, 2:32-3:5.) The speech recognizer then sends a request to the database, which holds different “speaker-dependent templates,” to send a user template corresponding to the user terminal. (*Tchorzewski*, 2:32-3:5.) The database then “provide[s] a copy of relevant [speaker-dependent templates](#) to the assigned recognition means.” (*Id.*, 2:34-3:5, 3a:11-23.)

67. It would have been obvious to store *Ditzik*’s language model, acoustic model, and handwriting model as a user template disclosed in *Tchorzewski*, so that the models required for speech and handwriting recognition are efficiently stored

and retrieved. It would have been obvious to apply *Tchorzewski*'s selection of one user-specific template from multiple different stored user templates to *Ditzik* to improve speech and handwriting recognition where there are multiple users of the same device. Indeed, it was known to adapt and store a plurality of models including both "an acoustic model and a contextual model" for different users, to have customized, accurate speech and handwriting recognition. (*Mitchell*, 5:46-56.) The more models tailored to specific users, the more likely the system is to accurately recognize user-specific speech and handwriting input.

68. A POSITA would have been motivated to combine *Ditzik* and *Tchorzewski* because they are in the same field of endeavor as each other, as well as the '737 patent, each of which are directed to improving accuracy for speech and handwriting recognition. First, *Ditzik* and *Tchorzewski* are in the same field of endeavor as each other and the '737 patent, all of which are directed to improving the accuracy of speech recognition systems. Second, *Ditzik* discloses training the user message model—training the language model and the phoneme model using speech input, and training the handwriting model using the handwriting input for a given user. (*Ditzik*, 3:9-48.) It was well understood that a tablet device, such as that in *Ditzik*, can be used by multiple users with different speech and handwriting patterns. A POSITA would have found it desirable to store different user message models for different users in a database, so that *Ditzik*'s recognition system could

instead efficiently retrieve a user message model for each user and have improved handwriting and speech recognition. (*Tchorzewski*, 1:5-19, 1:30-2:22.)

69. A POSITA would have done so with a reasonable expectation of success. First, a POSITA would have understood *Tchorzewski*'s telephone-based system is compatible with *Ditzik*'s system because *Ditzik* discusses using its system over telephone connections and identifying a user with a telephone switch and extension, akin to identifying a user with a user identification techniques, such as a username and password. (*Ditzik*, Abstract, 6:13-31, 6:49-53, 6:59-63, 7:19-23, 7:36-38, 7:48-50, 8:19-24, 8:35-39.) Second, *Ditzik* discloses storing user-specific speech and handwriting data (*Ditzik*, 3:59-64, 5:38-42), in separate databases. Modifying the databases to store a plurality of user-specific message models in view of *Tchorzewski* would have simply involved combining known elements according to known speech and handwriting recognition methods to yield predictable results of improving user-specificity of *Ditzik*'s speech and handwriting recognition for more accurate recognition. It would have also been obvious to try, choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success. A POSITA would have understood storing different user message models would offer streamlined costs and efficiency benefits. (*Tchorzewski*, 1:5-19.)

- e. **1[d] “adjust a language model of a local message model based at least in part on the particular user**

message model, wherein the local message model includes the language model, a handwriting model, and an acoustic model;”

70. In my opinion, *Ditzik*, *Tchorzewski*, and *Roberts* renders this limitation obvious.

71. As discussed for 1[c], *Ditzik* discloses a particular user message model including a language model, a handwriting model, and an acoustic model that are trained for a specific user. (§ IX.A.4.d.)

72. *Ditzik* also discloses a speaker-independent local message model including a language model, handwriting model, and acoustic model. *Ditzik*’s system stores these models and explains they can be trained with a user’s training input. (*Ditzik*, 3:43-48, 5:33-37.) For example, as explained above for limitation 1[c], the dictionary and grammar databases 64 in the language modeling program includes “contains language models of “several general . . . vocabularies.” (*Ditzik*, 3:53-56.) The dictionary and grammar database 64 includes the “grammar dictionary” that “contain[s] the rules of allowable word sequences,” and the “word dictionary” that “include[s] several general or specialized vocabularies.” (*Ditzik*, 3:53-56.) A POSITA would have understood that these constitute a language model stored locally in the databases 64 and 90 in the tablet. *Ditzik* also discloses the phoneme model database 62, which stores an acoustic model based on “a base or ground truth of voice data consisting of known words or speech patterns,”

installed on a tablet device “during factory setup” for “user independent recognition systems.” (*Ditzik*, 3:43-48.) *Ditzik* discloses a handwriting model stored in the handwriting stroke/pattern database dictionary 86, which may be based on “handwriting data from a known source with known writing samples, which represent a set of base truth data” for “user independent recognition systems.” (*Ditzik*, 5:33-37.) These models are stored locally in the handwriting stroke and pattern database dictionary. (*Ditzik*, 5:28-33, 5:40-42.)

73. *Ditzik* also discloses training the message model and adapting to a particular user’s speech and handwriting patterns. (*Ditzik*, Abstract, 3:14-18.) *Ditzik* discloses a system which is adaptive to a particular user’s speech and handwriting patterns so it clearly assumes that the databases can be changed by user interaction. *Ditzik* further discloses, as just discussed, a language model consisting of dictionaries of words and grammar rules for how those words can be combined. *Ditzik* also discloses that the language model may be changed or updated. (*Ditzik* 3:57-60.) A POSITA would have found this obvious because *Ditzik* discloses training options for the acoustic model for speaker dependent training, which include training particular words, and a POSITA would have understood that training particular words updates the language model too. The exemplary embodiment in *Ditzik* includes transcribing or taking notes on telephone conversations, so *Ditzik*’s system would benefit from knowing details about a

user's schedule and contact list, but a POSITA would have understood that such names and places would need to be incorporated into a user-specific language model in order to recognize the user.

74. A POSITA would have found “adjusting a language model of a local message model based in part on a particular user message model” obvious in view of the combination of *Ditzik* and *Roberts*. A POSITA would have found it obvious to apply *Roberts*'s language model updating scheme to *Ditzik*'s system as modified by *Tchorzewski*, such that *Ditzik*'s language model in the local message model is updated based on user-specific data in user message models. *Roberts* discloses adjusting a language model based on user-specific information. *Roberts* discloses a system where a user can correct misrecognized words using a user input (e.g., selecting the correct word from a list of alternative words in a pop-up window), which is used to train and “update the language model used by the recognition system.” (*Roberts*, 13:44-47; *see also id.*, Abstract, 4:51-62, 14:10-41, 19:32-45, 23:56-24:2, 25:19-46, 26:47-27:2, 27:48-28:19, 29:33-54, 36:26-37, FIGS. 1, 25.) This is consistent with the '737 patent, which discloses updating the language model using “manual editing of a tabulation,” where a user manually identifies and purges “from the tabulation of data structure 1200” any “[p]rimary text segments known to be poor indicators of context.” ('737 patent, 31:59-32:40, 33:3-25.) For example, *Roberts* discloses using “the most recent confirmed word stored in the

language context buffer” to update the statistical probability of the word being spoken. (*Roberts*, 13:50-60, 16:7-28.)

75. A POSITA would have been motivated to apply *Roberts*’s language model updating scheme to *Ditzik*’s recognition system as modified by *Tchorzewski* (see § IX.A.4.d), such that *Ditzik*’s language models are trained to increase recognition accuracy. First, both *Ditzik* and *Roberts* are in the same field of endeavor of recognition systems, where speech recognition is combined with some type of user input (e.g., handwriting input in *Ditzik* and user selection of correct words in *Roberts*) to improve recognition accuracy. Second, a POSITA would have thus found it desirable to adjust (e.g., train) language models based on *Ditzik*’s user message models, so that they can effectively limit “the recognition vocabulary to such a number of words which the language models considers most likely to occur in the current language context” for a particular speaker. (*Roberts*, 19:28-31.) It was known that “[w]hen recognition is being performed against phonemic as well as against individually trained acoustic models, it is desirable that a language model be used to limit the recognition vocabulary to a number of words which the recognizer can handle in a response time which the user finds satisfactory.” (*Roberts*, 19:23-28.) A POSITA would have similarly been motivated to adjust *Ditzik*’s language models used for handwriting recognition for the same reasons.

76. A POSITA would have done so with a reasonable expectation of success. Combining user message models in *Ditzik*'s system with *Roberts*'s process for adjusting a language model based on user-specific training data would have been combining known elements according to known speech and handwriting recognition methods to yield predictable results, namely, increasing recognition accuracy with personalized speech and handwriting data. It would also have been using a known technique for adjusting language models to improve similar speech and handwriting recognition devices in the same way to increase recognition accuracy. A POSITA would have found it obvious to try because there are only two available options—adjust or do not adjust the language model in *Ditzik*'s system—and a POSITA would have found it obvious that adjusting the language model would improve recognition accuracy. (*Roberts*, 1:65-67.)

77. Furthermore, a POSITA would have found that *Roberts*'s language model updating scheme is compatible with *Ditzik*'s language model. *Roberts*'s language updating scheme consists of changing the probabilities of word appearing in sequences in a bigram language model (*Roberts* 13:44-60), which is the same type of language model used in *Ditzik*. *Ditzik* describes that its language model is the “grammar dictionary may contain the rules of allowable word sequences,” which a POSITA would have understood to be a bigram language model (*Ditzik* 3:43-55). In fact, *Ditzik* explicitly states that its recognizer may be the Dragon

System's Naturally Speaking recognizer (*Ditzik*, 3:13-3-6) and *Roberts*, which is assigned to Dragon Systems, Inc., is describes an earlier version of the same product line.

f. 1[e] “generate text data based at least in part on the freehand input, the adjusted language model, and the handwriting model; and”

78. In my opinion, *Ditzik* and *Roberts* render this limitation obvious.

79. *Ditzik* discloses a handwriting recognition system that “transform[s] hand written letters, numbers, punctuation, and symbols into recognizable computer generated text.” (*Ditzik*, 4:51-53.) *Ditzik*, in view of *Roberts*, uses freehand input, the adjusted language model, and the handwriting model to do so.

80. *Ditzik*'s handwriting recognition system extracts “certain features of the raw data” from user freehand input. (*Ditzik*, 4:41-46, 5:40-45.) That raw data is then compared “to one or more databases, which may be comprised of language dictionaries (lexicons) and writing rules.” (*Id.*) Database 86, for example, includes “one or more dictionaries with a plurality of hand writing/sketch patterns for one or more languages and/or special symbol applications.” (*Id.*) As explained above in limitations 1[c] and 1[d], it would have been obvious in view of *Roberts* for such handwriting language model to be trained with user-specific handwriting data (the “adjusted language model”). (§§ IX.A.4.d-IX.A.4.e.) For example, when in training mode, handwriting “data 85 is input to a handwriting stroke and/or pattern

database 86,” which is “comprised of one or more dictionaries with a plurality of hand writing/sketch patterns for one or more languages and/or special symbol applications.” (*Ditzik*, 5:28-33, 5:49-50.) The language model works with a handwriting model, like handwriting word/gesture recognition program 88, to generate text data. The handwriting word/gesture recognition program 88 is “based on HMM techniques, which determines the best match or matches, by comparing the feature vectors to databases 86 and 90.” (*Ditzik*, 5:42-45.) A POSITA would have understood that the handwriting word/gesture recognition program includes a handwriting model because it uses Hidden Markov Model “HMM” techniques that are based on the data stored in the handwriting stroke/pattern and linguistic rules databases. Once this comparison is done, the information is sent to application text/symbol generation program 94, which then provides text data on display 78. (*Ditzik*, 5:1-14; 5:60-67.) This process of receiving freehand input and using an adjusted language model and handwriting model to generate text data is shown below in Figure 2:

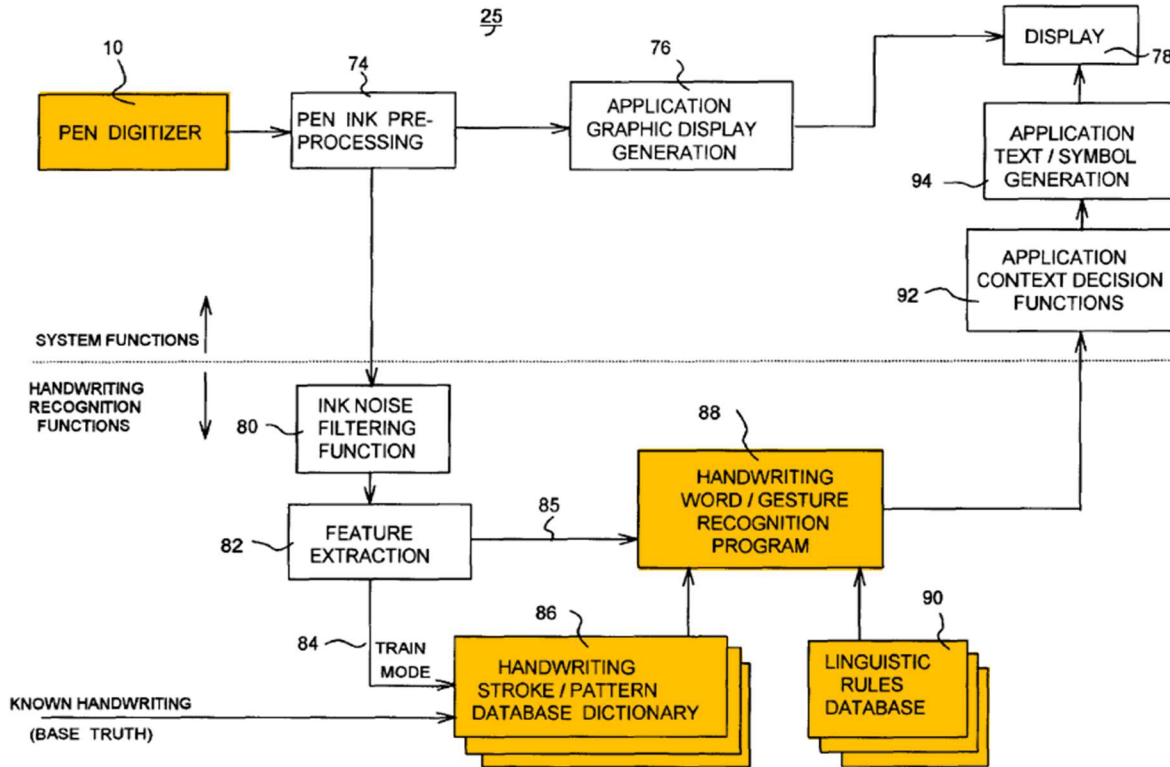


FIG. 2

(Ditzik, FIG. 2.)

81. To the extent *Ditzik* does not disclose this limitation, *Ditzik* combined with *Roberts* would have rendered it obvious. As explained for limitation 1[d], *Ditzik* combined with *Roberts* discloses “the adjusted language model.”

(§ IX.A.4.e) For the same reasons explained for limitations 1[c]-1[d], a POSITA would have found it obvious to use the adjusted language model with the freehand input and the handwriting model to increase recognition accuracy. (§§ IX.A.4.d-IX.A.4.e.) For example, a POSITA would have found it obvious to use the adjusted language model and the handwriting model to generate text so that the generated

text reflects user specific linguistic tendencies and handwriting patterns for more accurate handwriting recognition.

- g. 1[f] “generate text data based at least in part on the voice input, the adjusted language model, and the acoustic model.”**

82. In my opinion, *Ditzik* and *Roberts* render this limitation obvious.

83. *Ditzik* discloses that its system “convert[s] recognized speech data to computer text at near realtime or delayed.” (*Ditzik*, 8:32-34.) *Ditzik*, in view of *Roberts*, uses voice input, the adjusted language model, and the acoustic model to do so. Specifically, *Ditzik*’s “speech input recognition program/functions 24” performs “several functions including speech capture (via microphone or telephone line), digitize data, process data, store raw speech data, and storage of recognized text.” (*Ditzik*, 6:59-63.) The “[s]peech patterns from one or more persons may be captured and sent to a pre-processing function 52,” (voice input). (*Ditzik*, 3:20-27.) The data from the pre-processing function “is input to a feature extraction function 54, which may be based on the Hidden Markov Model (HMM),” a language model “well known to those skilled in the art.” (*Id.*) As explained in limitations 1[c] and [1d], it would have been obvious in view of *Roberts* for such language models to be trained with user-specific speech data (“adjusted language model”). (§§ IX.A.4.d-IX.A.4.e.) The output from the feature extraction function is then

“input to a phoneme modeling function 56,” an acoustic model. (*Ditzik*, 3:27-30.)

These steps are shown below in Figure 1:

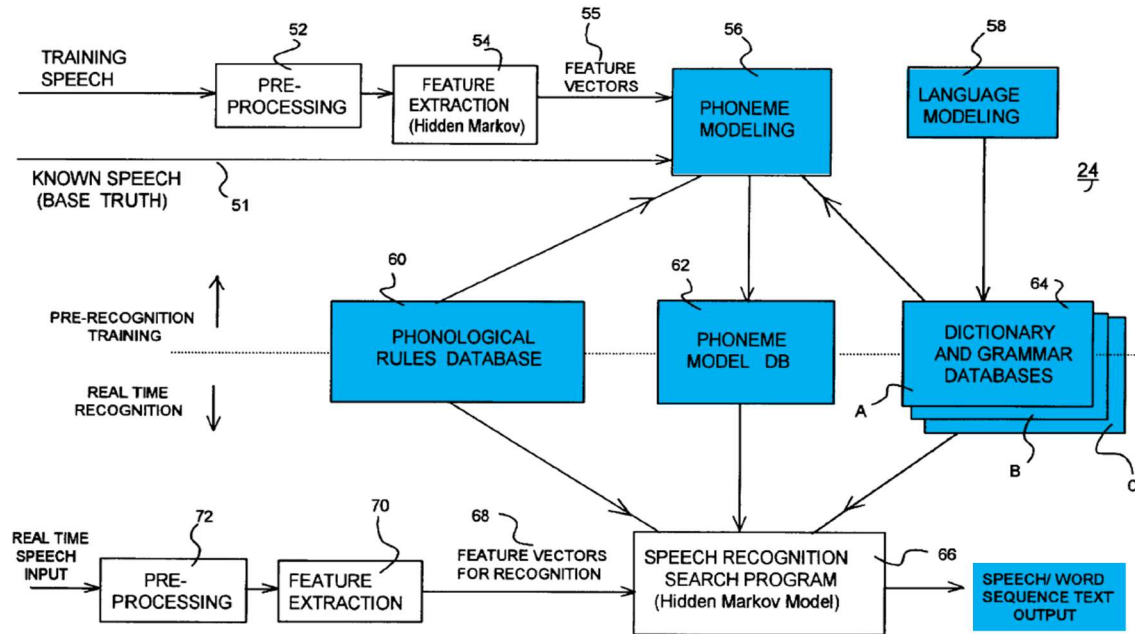


FIG. 1

PRIOR ART

(*Ditzik*, FIG. 1.) The end result is for the received and processed speech data to be converted “to computer text,” including the “speech/word sequence text output” as shown above in Figure 1. (*Ditzik*, 8:32-34.)

84. To the extent *Ditzik* does not disclose this limitation, *Ditzik* combined with *Roberts* would have rendered it obvious. As explained for limitation 1[d] above, *Ditzik* combined with *Roberts* discloses “the adjusted language model.” (§ IX.A.4.e.) For the same reasons explained for limitations 1[c]-1[d], a POSITA would have found it obvious to use the adjusted language model with voice input and the acoustic model to increase recognition accuracy. (§§ IX.A.4.d-IX.A.4.e.)

For example, a POSITA would have found it obvious to use the adjusted language model and the acoustic model to generate text so that the generated text reflects user specific linguistic tendencies and speech patterns for more accurate handwriting recognition.

5. Claim 4

- a. “The system of claim 1, wherein the system is further configured to train the particular user message model by updating the acoustic model.”**

85. In my opinion, *Ditzik*, *Tchorzewski*, and *Roberts* render obvious this additional limitation.

86. *Roberts* discloses training the particular user message model by updating the acoustic model. *Roberts* discloses updating acoustic models. *Roberts* explains, “[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, without further intervention by the user.” (*Roberts*, Abstract.) “If the system has existing models for the word or phonemes for which a new model is being built,” “then the accumulated statistics from previous training are combined with the” updated statistics “to make a combined maximum likelihood estimate of the statistical parameters.” (*Roberts*, 18:3-10.) *Roberts* discloses that “[a]s is well-known to those skilled in the art, this operation will allow adaptive training of the acoustic models, as well as zero-based training if no previous model exists.” (*Roberts*, 18:10-13.) Because the user

message model in the *Ditzik*, *Roberts*, and *Tchorzewski* system includes a user-specific acoustic model (*see supra* limitations [1c], [1d]), updating that acoustic model trains the user model.

87. A POSITA would have been motivated to combine *Roberts*'s process for updating the acoustic model with *Ditzik*'s system as modified by *Tchorzewski*, so that the user model is continuously trained and refined to maximize recognition accuracy. (§ IX.A.4.e.) A POSITA would have understood that the accuracy of the user model increases with more input training data, for example, training data for the acoustic model. (*Roberts*, Abstract.) In fact, updating the acoustic model in the user model was a known concept at the time. (*Mitchell*, 5:48-52.) A POSITA would have had a reasonable expectation of success because combining *Ditzik*'s system as modified by *Tchorzewski* with *Roberts*'s process would have involved combining known elements according to known recognition methods to yield predictable results of increasing recognition accuracy using personalized recognition data. (§ IX.A.4.e.)

6. Claim 12

- a. “The system of claim 1, wherein the particular user message model includes user-specific acoustic model information and user-specific handwriting model information.”**

88. *Ditzik*, *Tchorzewski*, and *Roberts* render obvious this additional limitation.

89. As explained for 1[c]-[d], *Ditzik* and *Tchorzewski* render obvious selecting a particular user message model from a plurality of user message models. (§§ IX.A.4.d-IX.A.4.e.) As discussed, those user message models include a user-specific language model, as well as a user-specific acoustic model and a user-specific handwriting model. (§§ IX.A.4.d-IX.A.4.e.) A POSITA would have understood that the information in such a user message model would include user-specific acoustic model information and user-specific handwriting model information. (§ IX.A.4.e.) For example, *Ditzik* discloses capturing “[s]peech patterns from one or more persons,” which are user-specific acoustic model information stored in phoneme model database 62. (*Ditzik*, 3:19-24.) *Ditzik* also includes speaker dependent speech recognition, which is by definition specific to an individual user (*Ditzik*, 3:13-16). *Ditzik* also discloses that when “system is in the training mode, data 84 is input to a handwriting stroke and/or pattern database 86,” which contains, for example, “a plurality of hand writing/sketch patterns for one or more languages and/or special symbol applications.” (*Ditzik*, 5:28-53.) A POSITA would have understood that the handwriting and sketch patterns stored in the database 86 are user-specific handwriting model information at least because it was a well-known feature in the art. (*Lewis*, [0020]-[0022], FIG. 1; *Root*, 3:66-4:3, 5:44-6:22, FIGS. 1-6.)

7. Independent Claim 13

a. 13[pre]: “A method, comprising:”

90. Independent claim 13 is substantially similar to independent claim 1, except that it is directed to a method. *Ditzik*, *Roberts*, and *Tchorzewski* render obvious the claimed method for the same reasons discussed above for claim 1.

(§ IX.A.4.a.)

b. 13[a]: “receiving, at a device, freehand input based at least in part on manipulation of a stylus;”

91. This limitation is substantially identical to limitation 1[a].

(§ IX.A.4.b.) *Ditzik* also discloses this limitation as explained for 1[a]. (*Id.*)

Moreover, *Ditzik* discloses receiving freehand input at a tablet, which is a device.

(*Id.*)

c. 13[b]: “receiving, at the device, audio input based at least in part on information received at a microphone;”

92. This limitation is substantially identical to limitation 1[b].

(§ IX.A.4.c.) *Ditzik* also discloses this limitation as explained for 1[b].

(§ IX.A.4.c.) Moreover, *Ditzik* discloses receiving audio input through a microphone connected to the tablet. (*Id.*)

- d. **13[c]: “selecting, via the device, a selected user message model from a plurality of user message models;”**

93. This limitation is substantially identical to limitation 1[c].

(§ IX.A.4.d.) *Ditzik* and *Tchorzewski* render this limitation obvious as explained for 1[c]. (§ IX.A.4.d.) A POSITA would have also understood that *Ditzik* as modified by *Tchorzewski* would use user identification information in *Ditzik*’s tablet to select the corresponding user message model from the plurality of user message models for different users.

- e. **13[d]: “adjusting, via the device, a language model of a local message model, wherein the local message model includes the language model, a handwriting model, and an acoustic model, wherein the adjusting the language model is based at least in part on the selected user message model;”**

94. This limitation is substantially identical to limitation 1[d]. *Ditzik*, *Tchorzewski*, and *Roberts* render this limitation obvious as explained for 1[d]. (§ IX.A.4.e.) A POSITA would have also understood *Ditzik* as modified by *Tchorzewski* and *Roberts* would adjust the language model of a local message model using *Ditzik*’s tablet because the tablet receives the acoustic and handwriting inputs used to update (or adjust) the language model.

- f. **13[e]: “generating, by the device, converted handwriting text data based at least in part on the**

**freehand input, the adjusted language model, and the
handwriting model; and”**

95. This limitation is substantially identical to limitation 1[e].

(§ IX.A.4.f.) *Ditzik* and *Roberts* render this limitation obvious as explained for 1[e]. (§ IX.A.4.f.) *Ditzik*’s tablet converts handwriting input into “computer generated text” and displays the text on its display. (*Ditzik*, Abstract, 1:7-16, 2:6-26, 4:32-36, FIG. 2.)

g. 13[f]: “generating, by the device, converted speech text data based at least in part on the audio input, the adjusted language model, and the acoustic model.”

96. This limitation is substantially identical to limitation 1[f].

(§ IX.A.4.g.) *Ditzik* and *Roberts* render this limitation obvious as explained for 1[f]. (§ IX.A.4.g.) *Ditzik*’s tablet generates “computer text of the speech input captured” and displays the text on its display. (*Ditzik*, Abstract, 1:7-16, 2:6-26, 4:12-15, FIG. 1.)

8. Claim 15

a. “The method according to claim 13, further comprising accessing the selected user message model via a network.”

97. *Ditzik* and *Tchorzewski* render this additional limitation obvious.

98. *Ditzik* combined with *Tchorzewski* renders the remaining limitation obvious. *Ditzik* discloses using “a new unique computer telephony application program,” which uses “telephone-like connections to communication networks

such as telephone, Internet, and cable networks.” (*Ditzik*, 7:19-23.) *Tchorzewski* discloses “a plurality of user terminals connected to a plurality of single-speaker dependent speech recognizers” where the connection of the user terminal causes “request[s] for a copy of one of the templates from one of the recogni[z]ers,” which “instruct[s] the database to send a complete copy of the requested template without interruption by another such request from another terminal.” (*Tchorzewski*, 3:27-3a:10.) Networked databases, such as those in *Tchorzewski*, were well known. A POSITA would have understood that *Ditzik* combined with *Tchorzewski* would use a network, such as the Internet, to transfer a user message model from a database to a recognizer when it is assigned to a particular user.

99. A POSITA would have been motivated to implement *Tchorzewski*’s process of accessing user message models (e.g., templates) in the user message model database in *Ditzik*’s system, so that each user message model would be efficiently retrieved from the database via *Ditzik*’s network. A POSITA would have had a reasonable expectation of success of making this combination because it would have simply been applying a known technique (e.g., accessing user message models via a network) to a known speech and handwriting recognition device to yield predictable results (e.g., efficient retrieval of each user message model.)

9. Claim 16

- a. “The method according to claim 13, further comprising training the selected user message model.”**

100. This limitation is substantially identical to claim 4. *Roberts* discloses training a selected user message model as explained for claim 4. (§ IX.A.5.) *Ditzik* and *Roberts* thus render obvious the additional limitation of claim 16. (*Id.*)

10. Independent Claim 19

- a. 19[pre]: “An article of manufacture including a non-transitory computer-readable medium having stored thereon instructions executable by a computing device to cause the computing device to perform operations comprising:”**

101. Independent claim 19 is substantially similar to claim 1, but it is directed to “[a]n article of manufacture including a non-transitory computer-readable medium having stored thereon instructions executable by a computing device to cause the computing device to perform operations.” *Ditzik*’s “computer system,” e.g., a tablet computer (article of manufacture), includes a “microprocessor” (a computing device) and a “main memory” (a non-transitory computer-readable medium). (*Ditzik*, 6:1-7:18, Fig. 3, Claims 6, 10, 12-13.) The main memory stores instructions for the microprocessor to perform operations, as explained in 19[a]-19[f]. (*Id.*; §§ IX.A.10.b-IX.A.10.g.) This is also shown in Figure 3 below:

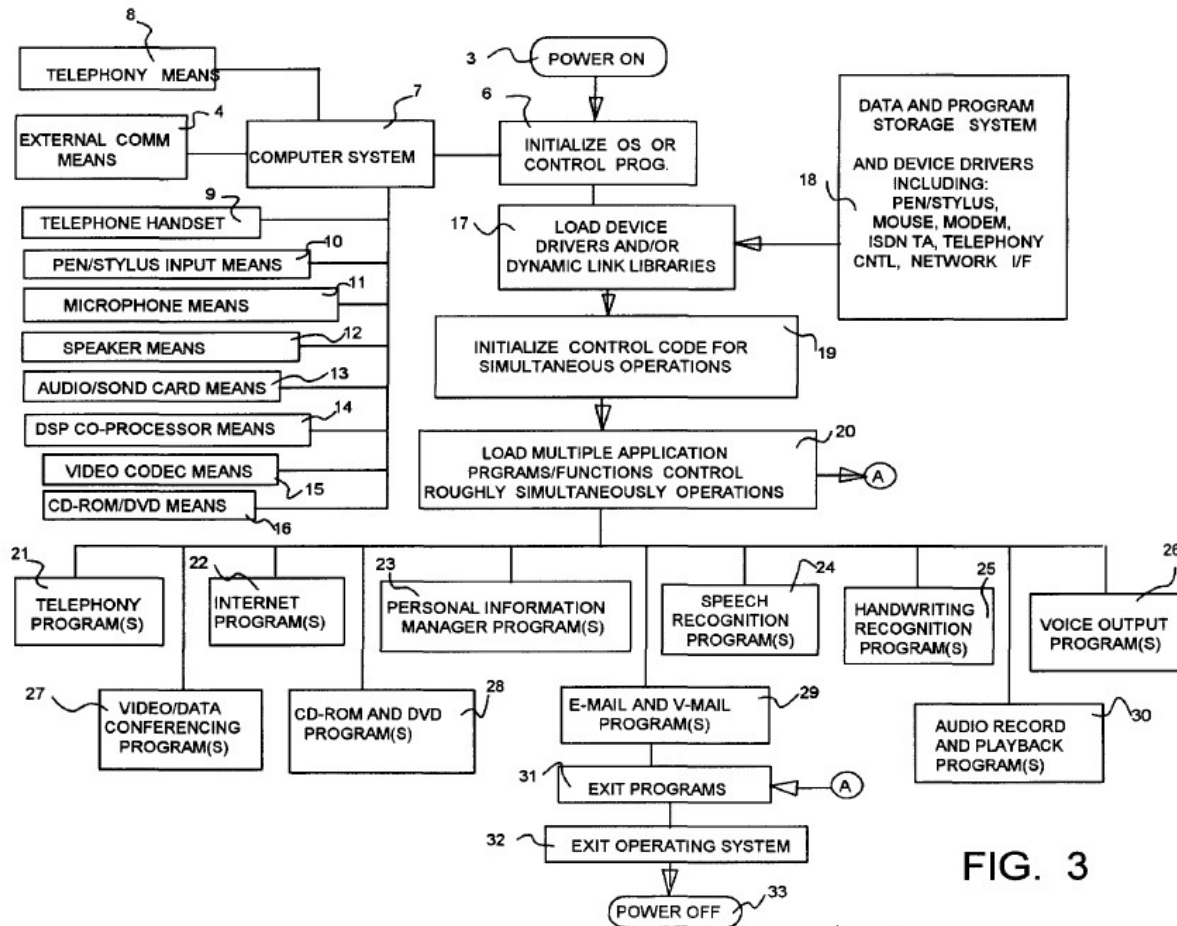


FIG. 3

(Ditzik, Fig. 3.)

b. 19[a]: “receiving freehand input based at least in part on manipulation of a stylus;”

102. As discussed above for limitation [19pre], *Ditzik* discloses a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input.

(*Ditzik*, 6:1-7:18, Fig. 3, Claims 6, 10, 12-13.) As discussed for limitations 1[a] and 13[a], *Ditzik* also discloses its tablet receiving freehand input based on manipulation of a stylus. (§§ IX.A.4.b, IX.A.7.b.)

c. 19[b]: “receiving audio input based at least in part on information received at a microphone;”

103. As discussed above for limitation [19pre], *Ditzik* discloses a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input. (*Ditzik*, 6:1-7:18, Fig. 3, Claims 6, 10, 12-13.) As discussed for limitations 1[b] and 13[b], *Ditzik* also discloses its tablet receiving audio input (voice input) based on information received at a microphone. (§§ IX.A.4.c, IX.A.7.c.) *Ditzik* thus discloses the instructions cause the tablet microprocessor to perform this limitation.

d. 19[c]: “selecting a particular user message model from a plurality of user message models;”

104. As discussed above for limitation [19pre], *Ditzik* discloses a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input. (*Ditzik*, 6:1-7:18, Fig. 3, Claims 6, 10, 12-13.) As discussed for limitations 1[c] and 13[c], *Ditzik* and *Tchorzewski* render obvious selecting a particular user message model from a plurality of user message models. (§§ IX.A.4.d, IX.A.7.d.)

e. 19[d]: “adjusting a language model of a local message model based at least in part on the particular user message model, the local message model including the language model, a handwriting model and an acoustic model;”

105. As discussed above for limitation [19pre], *Ditzik* discloses a tablet with a memory storing instructions to be executed by the microprocessor to

perform operations that generate text data based on speech and handwriting input.

(*Ditzik*, 6:1-7:18, Fig. 3, Claims 6, 10, 12-13.) As discussed for limitations 1[d] and 13[d], *Ditzik*, *Tchorzewski*, and *Roberts* render obvious adjusting a language model of a local message model based at least in part on the particular user message model, the local message model including a language, handwriting, and acoustic model. (§§ IX.A.4.e, IX.A.7.e.)

f. 19[e]: “generating converted handwriting text data based at least in part on the freehand input, on the adjusted language model, and on the handwriting model; and”

106. As discussed above for limitation [19pre], *Ditzik* discloses a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input. (*Ditzik*, 6:1-7:18, Fig. 3, Claims 6, 10, 12-13.) As discussed for limitations 1[e] and 13[e], *Ditzik* and *Roberts* render obvious generating converted handwriting text data based on the freehand input, adjusted language model, and handwriting model. (§§ IX.A.4.f, IX.A.7.f.) *Ditzik* and *Roberts* thus render obvious the instructions causing the tablet microprocessor to perform this limitation.

g. 19[f]: “generating converted speech text data based at least in part on the audio input, on the adjusted language model, and on the acoustic model.”

107. As discussed above for limitation [19pre], *Ditzik* discloses a tablet with a memory storing instructions to be executed by the microprocessor to

perform operations that generate text data based on speech and handwriting input.

(*Ditzik*, 6:1-7:18, Fig. 3, Claims 6, 10, 12-13.) As discussed for limitations 1[f] and

13[f], *Ditzik* and *Roberts* render obvious generating converted speech text data

based on the audio (voice) input, adjusted language model, and acoustic model.

(§§ IX.A.4.g, IX.A.7.g.) *Ditzik* and *Roberts* thus render obvious the instructions

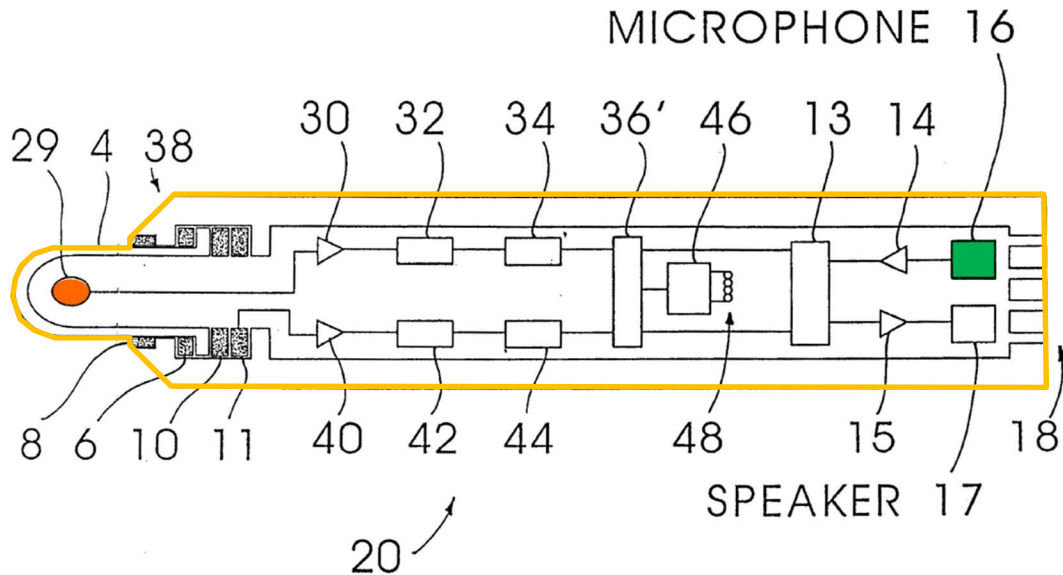
causing the tablet microprocessor to perform this limitation.

**B. Ground 2: Claims 2, 3, 5, 6, 11, and 14 Would Have Been
Rendered Obvious Over *Ditzik*, *Tchorzewski*, *Roberts*, and *Verrier***

1. Overview of *Verrier*

108. *Verrier* discloses a personal **stylus** with “an improved voice communications feature.” (*Verrier*, 3:40-43.) *Verrier*’s **stylus** contains a **microphone**, which receives a speech input from a user, and a **sensor** in the tip of the stylus, which detects the proximity of the stylus to a tablet and the mechanical touchdown of the tip on the tablet. (*Verrier*, 4:22-36.) These features are shown in Figure 2:

FIG. 2



(Verrier, FIG. 2 (annotated).)

109. The **microphone** in the stylus “can be selectively coupled to the transmitter in the stylus to position indicating signals” and “selectively disconnected from the transmitter in the stylus in response to the contact detection mechanism or the radiative pickup mechanism determining that the stylus tip is in contact with the digitizing display.” (Verrier, 4:56-5:7.) The stylus also includes a “personal identification number storage 104,” “which stores a personal identification number or a password or other security information, which can be transmitted from the stylus 20 to the digitizing display 22, for processing in the pen-based computer 74, to validate the stylus or the user of the stylus.” (Verrier, 18:13-20.)

2. Motivation to Combine

110. A POSITA would have been motivated to combine *Ditzik*, *Tchorzewski*, and *Roberts* with *Verrier* to improve upon the combined *Ditzik*, *Tchorzewski*, *Roberts* system. As explained above for limitation 1[a] in Ground 1, *Ditzik* discloses a stylus for inputting information to the tablet. (§ IX.A.4.b.) A POSITA would have been motivated to improve *Ditzik*'s stylus in view of *Verrier* in several ways. First, a POSITA would have been motivated to move the microphone from *Ditzik*'s table to its stylus to minimize receiving inputs from multiple sources and efficiently transition between receiving audio input and handwriting input. (*Verrier*, 3:27-36.) This would lead to improved speech and handwriting recognition, as well as improved text generation in response to speech and handwriting inputs. (*Id.*) Second, a POSITA would have been motivated to include a proximity sensor in *Ditzik*'s stylus to improve the efficiency of transitioning between receiving speech input and handwriting input depending on the location of the stylus relative to the tablet. This too would lead to improved speech and handwriting recognition, and improved text generation in response to speech and handwriting inputs. (*Id.*) Third, a POSITA would have been motivated to modify *Ditzik*'s stylus to allow for user identification upon activation of the stylus, and in turn retrieval of a particular user message model (as described in *Tchorzewski*) based on the user's identification. This modification would have

provided personalized user experiences and improved the accuracy of any associated speech recognition.

111. A POSITA would have made these changes with a reasonable expectation of success. Indeed, improving *Ditzik*'s stylus in view of *Verrier* would have simply been substituting one known element for another to obtain predictable results (personalized speech and handwriting recognition for improved accuracy). It would have involved combining a known tablet with a known stylus according to known methods in the same way to yield predictable results (including improved speech and handwriting recognition and user identification).

3. Claim 2

a. 2[pre]: "The system of claim 1, wherein the system is further configured to:"

112. In my opinion, as discussed, *Ditzik*, *Tchorzewski*, and *Roberts* render claim 1 obvious. (§ IX.A.4.) The system is configured to perform the steps as explained below.

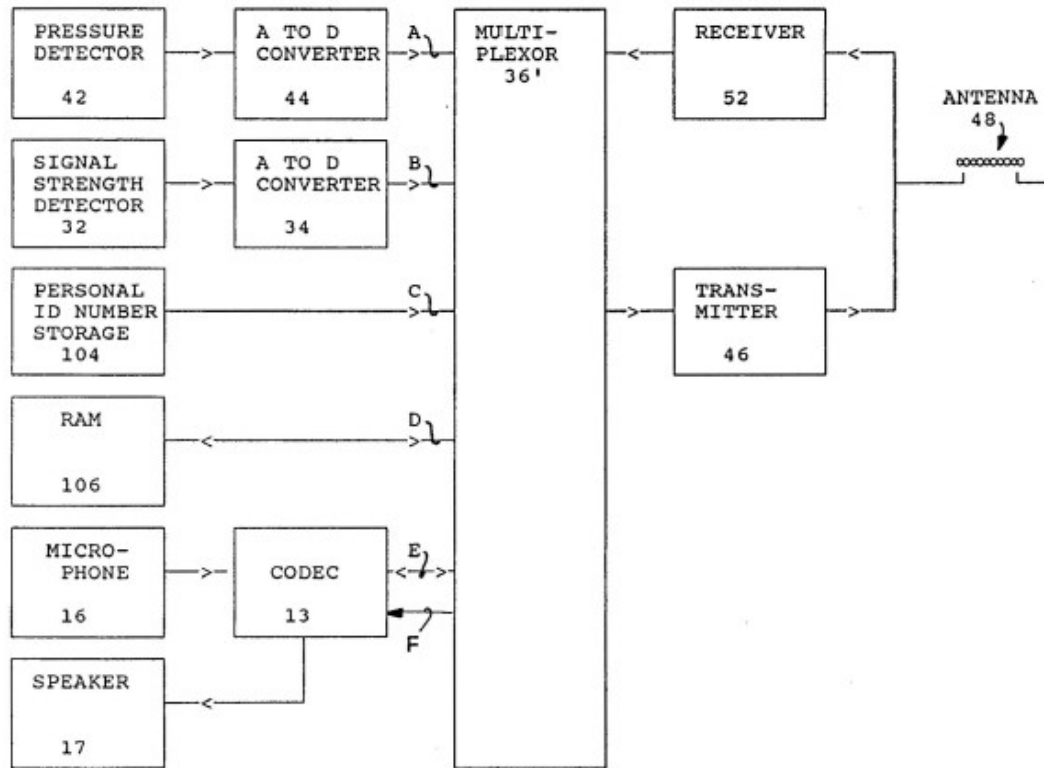
b. 2[a]: "identify information relating to a particular user;"

113. *Ditzik*, *Roberts*, *Tchorzewski*, and *Verrier* render obvious this limitation. *Verrier* discloses a "personal stylus" with a "personal identification number storage" that "stores personal identification number of a password or other security information, which can be transmitted from the stylus 20 to the digitizing display 22, for processing in the pen-based computer 74, to validate the stylus or

the user of the stylus" with the tablet. (*Verrier*, Abstract, 8:53-58, 9:11-17, 18:13-

20.) This is shown in Figure 3A:

FIG. 3A



(*Verrier*, FIG. 3A.)

114. A POSITA would have been motivated to combine *Ditzik*, *Roberts*, and *Tchorzewski* with *Verrier*. Each of *Ditzik*, *Roberts*, and *Tchorzewski* discloses receiving user input, and it was well known at the time that a stylus can be used to input user identification information to a computer, such as a tablet. *Tchorzewski*, for example, discloses identifying the user when the system receives an indication that a user terminal has been activated. (*Tchorzewski*, 4:1-10.) In fact, as explained

for 1[a], *Ditzik* discloses using a stylus for inputting information to the tablet.

(§ IX.A.4.b.) A POSITA would have understood that *Ditzik*'s stylus can be modified such that the *Ditzik*'s system identifies the user upon the activation of *Verrier*'s stylus and retrieves the user's message model as described in *Tchorzewski*. A POSITA would have understood that such modification would have provided personalized user experience and more accurate speech recognition.

115. A POSITA would have done so with a reasonable expectation of success. *Ditzik* as modified by *Roberts* and *Tchorzewski* already includes a stylus (§ IX.A.4.b); thus, replacing that stylus with *Verrier*'s personalized stylus would have involved a simple substitution of one known element for another to obtain predictable results (e.g., personalized speech and handwriting recognition for improved accuracy). It would have involved combining prior art tablet and personalized stylus according to known methods to yield predictable results, and using a known personalized stylus in *Verrier* to a known device in *Ditzik* as modified by *Roberts* and *Tchorzewski* in the same way.

c. 2[b]: “wherein the particular user message model is selected based at least in part on the information relating to the particular user; and,”

116. In my opinion, *Ditzik*, *Roberts*, *Tchorzewski* combined with *Verrier* render this limitation obvious. As explained for 1[c], *Ditzik* combined with *Tchorzewski* renders obvious selecting a particular user message model from a

plurality of user message models based on which user terminal is activated.

(§ IX.A.4.d.) For example, *Tchorzewski* discloses that “speaker-dependent templates” (e.g., user message model) are transferred from the database to an activated user terminal. (*Tchorzewski*, 2:32-3:5.) It would have been further obvious in view of *Verrier* to have the identification number in the stylus trigger the selection of the particular user message model, instead of the telephone extension associated with the user terminal, as in *Tchorzewski*. (*Verrier*, 8:56-57, 9:11-15, 18:13-20.) Doing so would have been the simple substitution of known elements in a known way to achieve predictable results.

117. A POSITA would have found it obvious to replace the stylus in *Ditzik*’s system as modified by *Roberts* and *Tchorzewski* with *Verrier*’s microphone-embedded stylus. First, *Ditzik*, *Roberts*, *Tchorzewski*, and *Verrier* are in the same field of endeavor as each other and the ’737 patent for improving the accuracy of speech and/or handwriting recognition. As explained in 1[a], *Ditzik* as modified by *Roberts* and *Tchorzewski* discloses using a stylus to provide handwriting input to a tablet. (§ IX.A.4.b.) A POSITA would have been motivated to replace *Ditzik*’s stylus with *Verrier*’s to improve the detection of the touchdown of the stylus on the tablet and to accommodate voice communications at the user’s location. (*Verrier*, 3:4-36.)

118. A POSITA would have done so with a reasonable expectation of success. Modifying *Ditzik*'s stylus with *Verrier*'s microphone stylus would have involved combining prior art elements (e.g., replacing a regular stylus with a microphone-embedded stylus) according to known methods to yield predictable results of improving the accuracy and convenience of known recognition systems. (*Verrier*, 3:4-36.) It would have been a simple substitution of a known stylus for *Verrier*'s microphone-embedded stylus to obtain predictable results—a system that can accurately receive both speech and handwriting input via the stylus.

- d. **2[c]: “wherein the plurality of user message models includes two or more user message models corresponding to users that are different from the particular user.”**

119. *Ditzik* and *Tchorzewski* render obvious this limitation. As explained for limitation 1[d], *Ditzik* and *Tchorzewski* render obvious a plurality of user-specific message models. (§ IX.A.4.e.) A POSITA would have understood the plurality would include models for a number of users, such as at least three users—a “particular user,” and “two or more” users that are “different from the particular user.” *Tchorzewski* discloses that user-trained patterns “are stored as ‘templates’ in Random Access Memory (RAM),” which is “a common database . . . containing the templates of all the users.” (*Tchorzewski*, 1:8-9, 2:1-2.)

120. A POSITA would have been motivated to apply *Tchorzewski*'s user-specific templates to *Ditzik*'s language models, acoustic models, and handwriting

models, so that a combination of these models for a specific user is retrieved efficiently upon request. A POSITA would have understood that storing different user message models for different users maximizes speech and handwriting recognition because one message model may not accurately reflect the speech and handwriting patterns of another user. Any system which supports models for multiple users will be capable of storing models which are different from those of any particular user, as long as there are two or more users of the system.

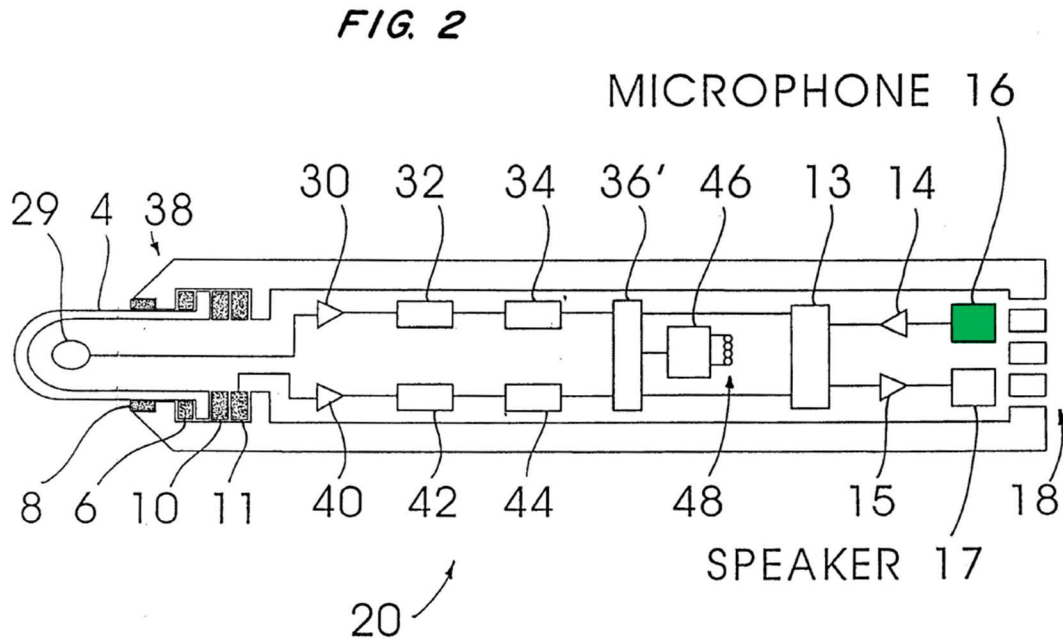
121. A POSITA would have done so with a reasonable expectation of success. It would have involved combining prior art elements (e.g., applying *Tchorzewski*'s storage of user templates) according to known methods to yield predictable results of improving the accuracy of speech and handwriting recognition by using a personalized user message model for a given user. It would also have been obvious to try given the finite number of identified, predictable solutions (e.g., use a generic message model for all users v. use a personalized message model for each user) with a reasonable expectation of success.

4. Claim 3

a. “The system of claim 2 wherein the microphone is housed in the stylus.”

122. *Ditzik, Tchorzewski, Roberts, and Verrier* render obvious this additional limitation.

123. In my opinion, *Verrier* discloses this limitation. *Verrier* discloses a personal stylus housing “a microphone and a speaker.” (*Verrier*, Abstract, 4:24-27, 8:57-58, 9:18-20.)



(*Verrier*, FIG. 2 (annotated).)

124. A POSITA would have been motivated to modify *Ditzik*'s stylus in a similar way to provide a personalized user experience and improve the accuracy of speech and handwriting recognition, as discussed above. A POSITA would have had a reasonable expectation of success in doing so. It would have been obvious to add a microphone to *Ditzik*'s stylus to smoothly transition between speech and handwriting input and avoid clashes between the two types of inputs. (*Verrier*, 3:4-36.)

5. Claim 5

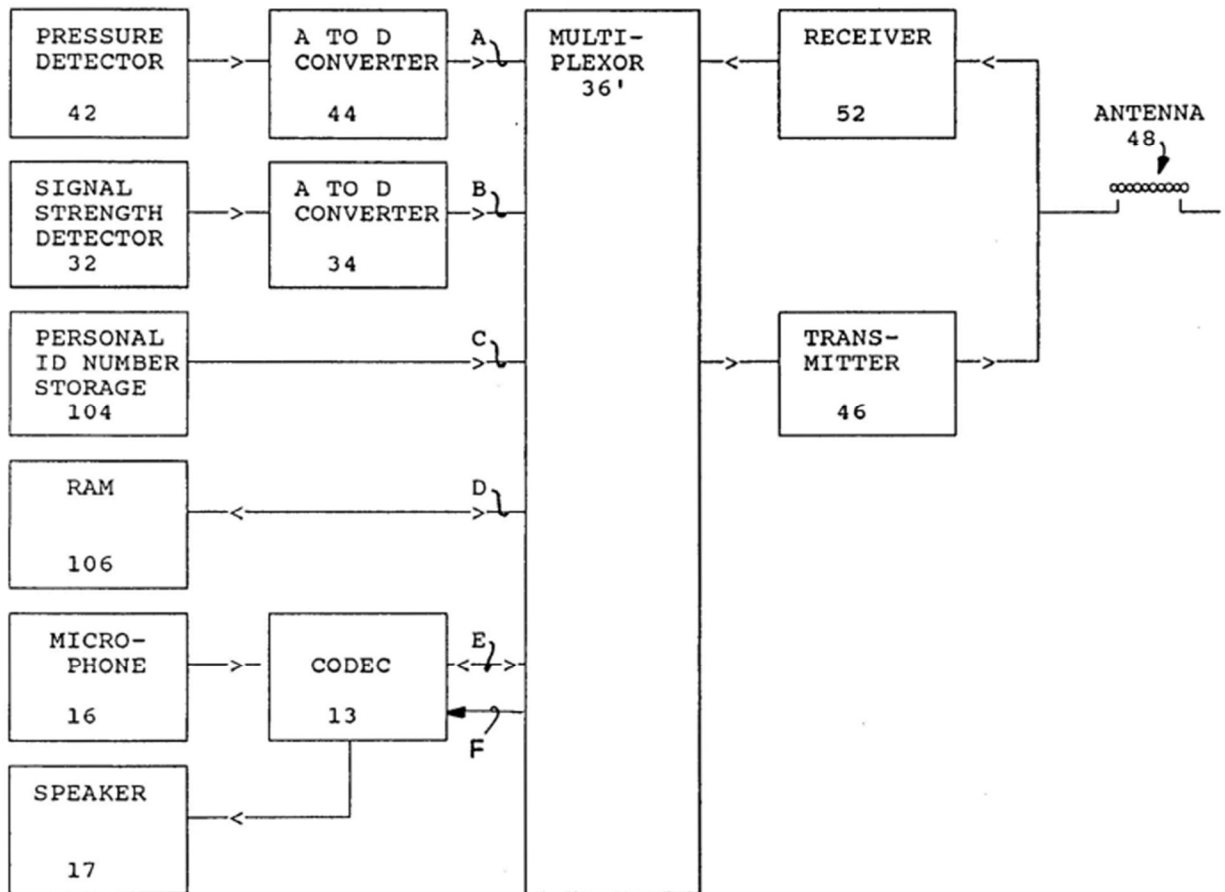
- a. “The system of claim 1, further comprising a proximity sensor configured to sense a proximity of a user to the microphone.”**

125. *Ditzik, Tchorzewski, Roberts, and Verrier* render obvious this additional limitation.

126. In my opinion, *Verrier* discloses a proximity sensor that senses the proximity of the user to the microphone in the stylus. *Verrier*’s stylus has a sensor “in the tip of the stylus [that] picks up the electromagnetic signals radiated from the display overlay, and converts those signals into a suitable form for retransmission back to the pen-based computer system.” (*Verrier*, 2:46-51.) “The relative signal strength of the signals picked up by the stylus is used to infer the relative separation distance *Z* of the tip of the stylus with respect to the overlay.” (*Verrier*, 2:51-54.) This is a proximity sensor because it detects the strength of a radiative signal to determine whether the “user [has] brought the stylus 20 into a relatively close position with respect to the electromagnetic tablet 54.” (*Verrier*, 10:34-42.) When the sensor does not detect an electromagnetic signal indicating that the stylus is close to the tablet, but receives a voice signal, *Verrier* explains that the stylus sensor determines that the stylus is close to the speaker and enables a microphone. (*Verrier*, 10:15-30.) Thus, a POSITA would have understood that

the sensor in the stylus tip is configured to sense a proximity of a user to the microphone. This signal strength detector is shown below in Figure 3A:

FIG. 3A



(Verrier, FIG. 3A.)

127. A POSITA would have been motivated to modify *Ditzik*'s stylus in a similar way to improve the accuracy of speech recognition, as discussed above. A POSITA would have had a reasonable expectation of success in doing so.

6. Claim 6

- a. “The system of claim 1, further comprising an identification circuit configured to identify an owner of the stylus, wherein the system is configured to select the particular user message model at least in part in response to the identification circuit.”**

128. *Ditzik, Tchorzewski, Roberts, and Verrier* render obvious this additional limitation.

129. In my opinion, *Verrier* discloses this limitation. As explained for claim 2, *Verrier* discloses that its personal stylus comprises an identification circuit. (§ IX.B.3.) *Verrier* discloses a “personal identification number storage 104,” which stores user identification information, such as “a password or other security information” “to validate the stylus or the user of the stylus.” (*Verrier*, Abstract, 8:53-58, 9:11-17, 18:13-20.) A POSITA would have understood that *Verrier* contains a user identification circuit to use the user identification circuit to use the user identification information and identify and validate that user.

130. Moreover, as explained for 2[b], *Ditzik, Roberts, Tchorzewski* and *Verrier* render obvious selecting a particular user message model based in part on information relating to the particular user, and that would have been in response to the identification circuit identifying and validating the specific user. (§ IX.B.3.c.) A POSITA would have understood that the electronic components in *Verrier* used for user identification are in an electronic circuit. (*Verrier*, 18:9-12.)

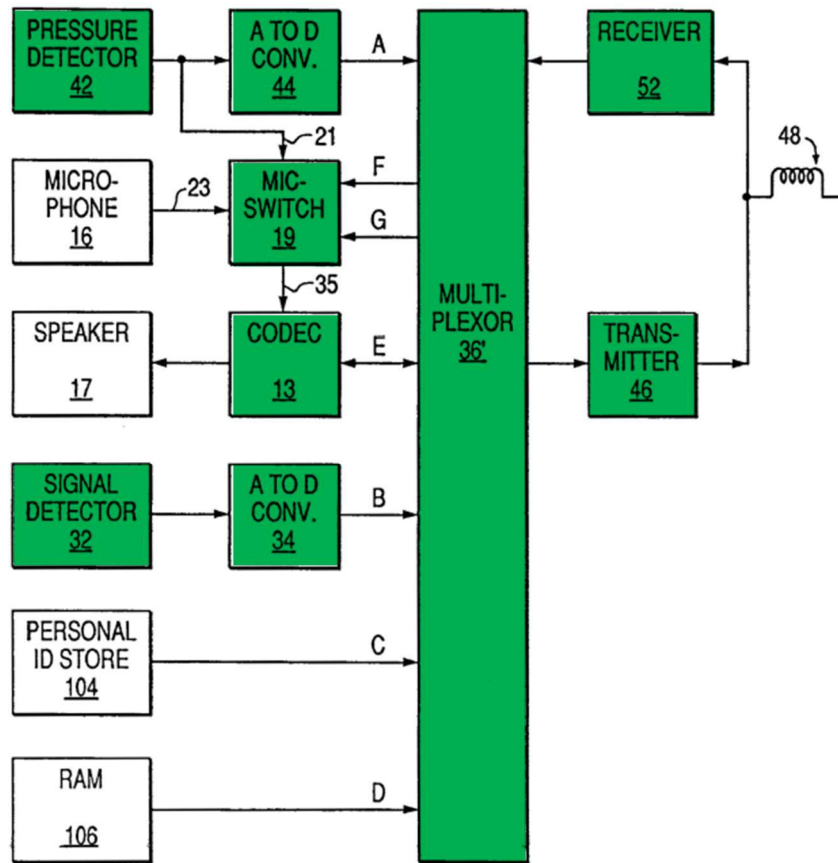
7. **Claim 11**

- a. **“The system of claim 1, further comprising an activation circuit configured to sense movement of the stylus toward the tablet surface, wherein the system is further configured to receive the freehand input in response to the activation circuit sensing the movement.”**

131. *Ditzik, Tchorzewski, Roberts, and Verrier* render obvious this additional limitation.

132. In my opinion, *Verrier* discloses this limitation. *Verrier* discloses the proximity sensor in the tip of the stylus measures the “relative signal strength of the signals picked up by the stylus . . . to infer the relative separation distance Z of the tip of the stylus with respect to the overlay” in the tablet. (*Verrier*, 2:46-54.) The sensor in *Verrier*’s stylus is connected to a pressure detector 42 and signal strength detector 32, which feed signal strength data to the multiplexor 36. (*Verrier*, 7:43-9:17.) Together, the components act as an **activation to circuit** that senses the movement of the stylus toward the tablet surface. This is shown in Figure 3B:

FIG. 3B



(*Verrier*, FIG. 3B (annotated).) A POSITA would have understood that these components are in an electronic circuit. (*Verrier*, 18:9-12.)

133. In response to detecting a proximity signal and “a mechanical touchdown of the tip 4 of the stylus 20 on the surface of the digitizing display 54,” *Verrier*’s system receives freehand input from the stylus on the tablet. (*Verrier*, 11:5-20, 3:4-26.) A POSITA would have been motivated to modify *Ditzik*’s stylus in a like way to improve the accuracy of handwriting recognition, as discussed. A POSITA would have had a reasonable expectation of success in doing so.

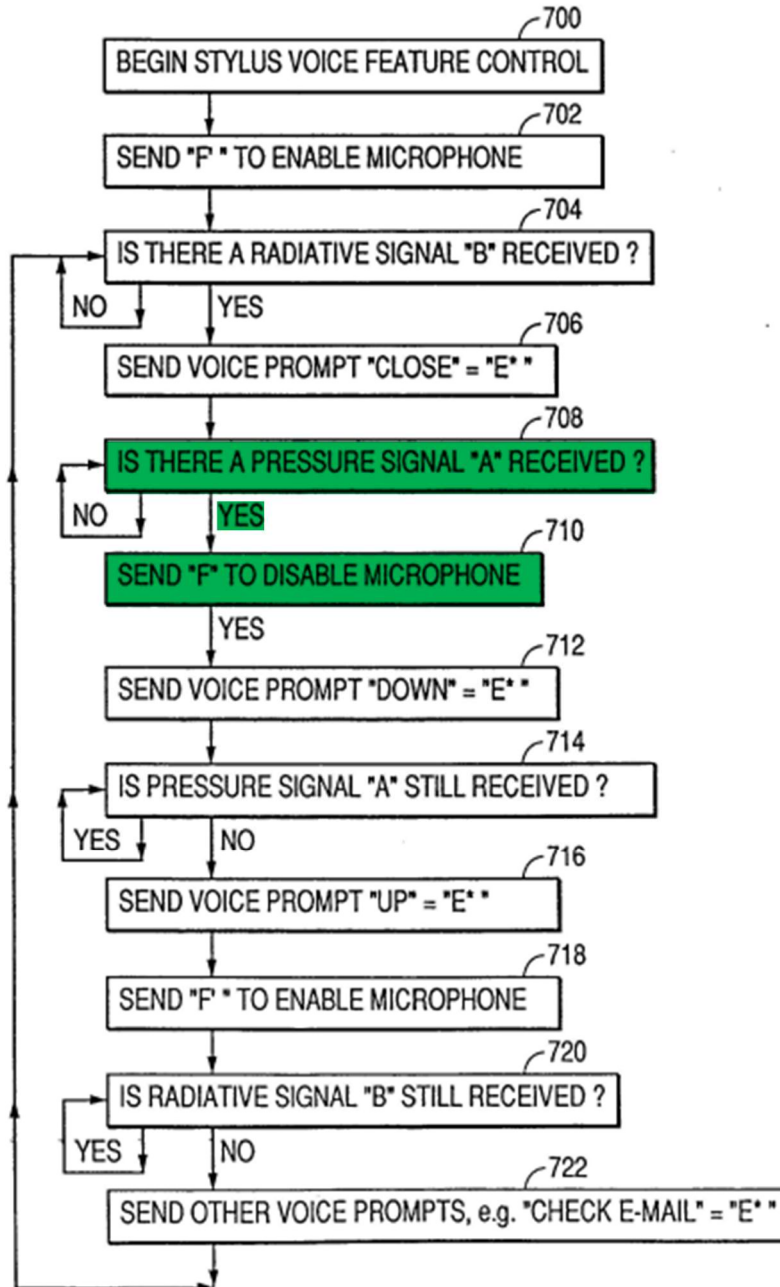
8. Claim 14

- a. “The method according to claim 13, further comprising switching from receiving audio input to receiving freehand input in response to detecting movement of the stylus towards a tablet.”**

134. *Ditzik, Tchorzewski, Roberts, and Verrier* render obvious this additional limitation.

135. In my opinion, *Verrier* discloses this limitation. When *Verrier*’s system detects a mechanical touchdown of the tip of the stylus on the tablet display, “it sends a F signal to the stylus 20 to disable the microphone” and “decouple the microphone from the transmitter 46 in the stylus, in order to avoid picking up the spurious acoustic noise from the stylus tip 4 as it is rubbed against the surface of the tablet.” (*Verrier*, 11:5-42.) This is shown below in Figure 7:

FIG. 7



(Verrier, FIG. 7 (annotated.))

136. A POSITA would have been motivated to modify *Ditzik*'s stylus in a like way to improve the accuracy of speech and handwriting recognition, as

discussed above. A POSITA would have had a reasonable expectation of success in doing so.

C. Ground 3: Claims 9 and 10 Would Have Been Obvious Over *Ditzik, Tchorzewski, Roberts and Bellegarda*

1. Overview of *Bellegarda*

137. *Bellegarda* discloses “a latent semantic analysis” that uses clustering of words based on their contexts, which is used to determine which words are most likely to co-occur. (*Bellegarda*, 172.) *Bellegarda*’s latent semantic analysis involves three steps: 1) “a singular value deposition (SVD) is performed” on “a matrix of co-occurrences between words and documents (articles),” “producing singular vector representations for words and documents,” 2) “a distance measure consistent with the SVD formalism is defined on the resulting vector space,” and 3) “the singular vectors are clustered according to this metric.” (*Bellegarda*, 173.) Essentially, given samples of collections of words (“documents”) for training data, *Bellegarda* determines words which are less common across all the samples, but found in some subset of the examples, which are then clustered together. Given this clustering, a new sample can be determined to be most closely aligned with one of the previously classified clusters.

2. Motivation to Combine

138. A POSITA would have been motivated to combine *Ditzik, Tchorzewski*, and *Roberts* with *Bellegarda*. *Ditzik, Roberts*, and *Tchorzewski*

disclose speech recognition systems that use a language model, and *Bellegarda* discloses well-known implementation details for such language models, including using both syntactic and semantic models together. It was known conventional syntactic language models had drawbacks, such as the lack of contextual analysis and global language constraints. (*'106 patent*, 1:54-61.) It was also known that conventional semantic language models had drawbacks, like “the inherent lack of tight local word order constraints” that “prevent their widespread acceptance and use.” (*Id.*, 2:51-55.) Thus, as *Bellegarda* discloses, a POSITA would have found it desirable to use a semantic model as a “complement to conventional class-based statistical language modeling” to improve speech recognition, remedying each other’s shortcomings and optimizing accuracy. (*Bellegarda*, 172; *'106 patent*, 2:55-58.) A POSITA would have used such a language model in the *Ditzik*, *Tchorzewski*, *Roberts* combined system. Furthermore, a POSITA would have understood that the clustering technique recited in claim 10 is merely a conventional clustering technique used in a semantic language model.

139. A POSITA would have used such a language model in the *Ditzik*, *Tchorzewski*, *Roberts* combined system with a reasonable expectation of success. This would have involved combining known elements according to known speech recognition methods to yield predictable results, namely, improving speech recognition accuracy. (*Bellegarda*, 172.)

3. Claim 9

- a. “The system of claim 1, wherein the system adjusting the language model includes adjusting a syntactic model and a semantic model included in the language model.”**

140. *Ditzik, Tchorzewski, Roberts, and Bellegarda* render obvious this additional limitation. (See IX.A.4.)

141. As explained for claim 1[d] above, *Ditzik* and *Roberts* render the adjusting the language model obvious. (§ IX.A.4.e.)

142. It was well known at the time that a language model would have included a syntactic model and a semantic model. (*Bellegarda*, 172.) Indeed, the ’737 patent admits that a “language model may be of a conventional type, for example consisting of an N-gram syntactic model,” and that a semantic model is “disclosed in U.S. Pat. No. 5,828,999 to Bellegarda et al.” (’737 patent, 22:57-59, 26:16-19.) A POSITA would have understood that the adjusted language model in the combined *Ditzik, Tchorzewski, Roberts* system would include such well known syntactic and semantic models, and that adjusting the language model would mean adjusting the constituent syntactic and semantic models.

143. Indeed, this would have been obvious over *Bellagarda. Bellegarda* discloses “conventional class-based modeling techniques rely on short span relationships, and therefore tend to produce syntactically-oriented clusters.” (*Bellegarda*, 175.) It also discloses a “latent semantic analysis” that “seeks to relate

one another those words which are found to be linked from the evidence presented in the training text database, without regard to the particular syntax used to express that semantic link.” (*Bellegarda*, 172, 173.) These syntactic and semantic models comprising the language model in *Bellegarda* are adjusted, including by using “the available training data by simply keeping track of which word appeared in the document,” “in marked contrast with *n*-gram modeling,” which “ignores word order.” (*Bellegarda*, 173.)

144. A POSITA would have been motivated to combine *Ditzik*, *Roberts*, and *Tchorzewski* with *Bellegarda*. Each of *Ditzik*, *Roberts*, and *Tchorzewski* discloses a speech recognition system that uses a language model, and *Bellegarda* discloses well-known implementation details for the language model. It was well known at the time that the conventional syntactic language models had several drawbacks, such as the lack of contextual analysis and global language constraints. (*'106 patent*, 1:54-61.) It was also well known that the conventional semantic language models had several drawbacks, such as “the inherent lack of tight local word order constraints in such models,” which “may ultimately prevent their widespread acceptance and use.” (*'106 patent*, 2:51-55.) Thus, as *Bellegarda* itself noted, a POSITA would have understood found it desirable to use the semantic technique as a “complement to conventional class-based statistical language modeling technique.” (*Bellegarda*, 172.) Thus, a POSITA would have been

motivated to combine these known language models to complement each other's drawbacks and optimize speech recognition. (*'106 patent*, 2:55-58.)

145. A POSITA would have done so with a reasonable expectation of success. Combining *Ditzik's* system as modified with *Roberts*, and *Tchorzewski* with *Bellegarda's* known syntactic and semantic language models would have involved combining prior art elements according to known speech recognition methods to yield predictable results of improving speech recognition accuracy. (*Bellegarda*, 172.)

4. Claim 10

- a. **10[pre]: “The system of claim 9, wherein the semantic model comprises at least one semantic cluster selected from the group consisting of:”**

146. In my opinion, *Ditzik*, *Roberts*, *Tchorzewski*, and *Bellegarda* render obvious claim 9. (§ IX.C.3.) *Bellegarda* discloses “a latent semantic analysis” applying “familiar clustering techniques,” which are “semantic in nature,” and can act “as a complement to conventional class-based statistical language modeling techniques.” (*Bellegarda*, 172.) Although it is not required that the combination disclose all three limitations 10[a]-10[c], *Bellegarda* discloses all three for the reasons explained below.

- b. 10[a]: “a semantic cluster of a first type that is a mathematically derived group of vectors representing text segments;”**

147. In my opinion, *Bellegarda* discloses 10[a]. *Bellegarda* discloses detailed steps of creating “vector representation” of text segments. (*Bellegarda*, 173-74.) For example, *Bellegarda* discloses that its latent semantic analysis uses “a parsimonious vector representation of each word in a suitable vector space, where familiar clustering techniques can be applied.” (*E.g.*, *Bellegarda*, 172.) *Bellegarda* further discloses that “a singular value decomposition (SVD) is performed on” “a matrix of co-occurrences between words and documents (articles)” to produce “singular vector representations for words and documents.” (*Bellegarda*, 173.) Then, “a distance measure consistent with the SVD formalism is defined on the resulting vector space.” (*Id.*) Thereafter, “the singular vectors are clustered according to this matrix.” (*Id.*)

- c. 10[b]: “a semantic cluster of a second type that is a discrete grouping of text having a particular context; and”**

148. In my opinion, *Bellegarda* discloses 10[b]. *Bellegarda* discloses “group[ing] words into classes and accumulate statistics at the class level rather than the word level” “to better estimate the conditional probability of a word given some context by taking advantage of observations of other words that behave ‘like’

this word in this particular context.” (*Bellegarda*, 172.) *Bellegarda* discloses the following equation for clustering,

$$PP = \exp\left(-\frac{1}{Q} \sum_{i=1}^Q \log \Pr(w_i | \mathcal{C}, \mathcal{S})\right), \quad (8)$$

where C denotes “context considered,” indicating that its semantic clustering uses a discrete grouping of text having a particular context. (*Bellegarda*, 174.)

Bellegarda further teaches that the probability can be expanded as,

$$\Pr(w_i | \mathcal{C}, \mathcal{S}) = \Pr(w_i | C_i) \Pr(C_i | \mathcal{C}), \quad (9)$$

where “ $\Pr(C_i | \mathcal{C})$ can be obtained from the relative frequency of cluster in the context C .” (*Bellegarda*, 174.)

- d. **10[c]: “a semantic cluster of a third type that is defined with respect to an individual reference word that is indicative of a particular context.”**

149. *Bellegarda* discloses 10[c]. *Bellegarda* discloses that its latent semantic framework uses “co-occurrence analysis” that “take[s] place across larger contexts (longer spans) than traditionally considered,” “where two words are said to co-occur if they tend to appear in the same kind of documents, whether or not they actually occur within identical word contexts in those documents.”

(*Bellegarda*, 172-173.) A POSITA would have understood that the co-occurring words act as an individual reference word for each other indicative of a particular context. For example, the words in “Cluster 1” disclosed in *Bellegarda* are

individual reference words indicative of art, and the words in “Cluster 2” are

individual reference words indicative of law. (*Bellegarda*, 175.)

D. Ground 4: Claims 7 and 8 Would Have Been Obvious Over *Ditzik*, *Tchorzewski*, *Roberts*, and *Campbell*

1. Overview of *Campbell*

150. *Campbell* is a “tutorial on design and development of automatic speaker recognition system,” which uses “a machine to recognize a person from a spoken phrase” “to identify a particular person or to verify a person’s claimed identity.” (*Campbell*, 1437.)

2. Claim 7: “The system of claim 1, wherein the system is configured to select the particular user message model based at least in part on an identifying signal received from the microphone.”

151. In my opinion, *Ditzik*, *Roberts*, *Tchorzewski*, and *Campbell* render obvious this additional limitation.

152. As explained for 1[c], *Ditzik* combined with *Tchorzewski* renders obvious selecting a particular user message model from a plurality of user message models. (§ IX.A.4.d, *supra*.) A POSITA would have found it obvious to combine *Ditzik*’s system as modified by *Tchorzewski* with *Campbell* to select the user message model based at least in part on an identifying signal received from the microphone. For example, *Campbell* discloses “automatic speaker recognition,” which “recognize[s] a person from a spoken phrase” “to identify a particular

person or to verify a person's claimed identity." (*Campbell*, 1437.) This system is shown in Figure 1 below:

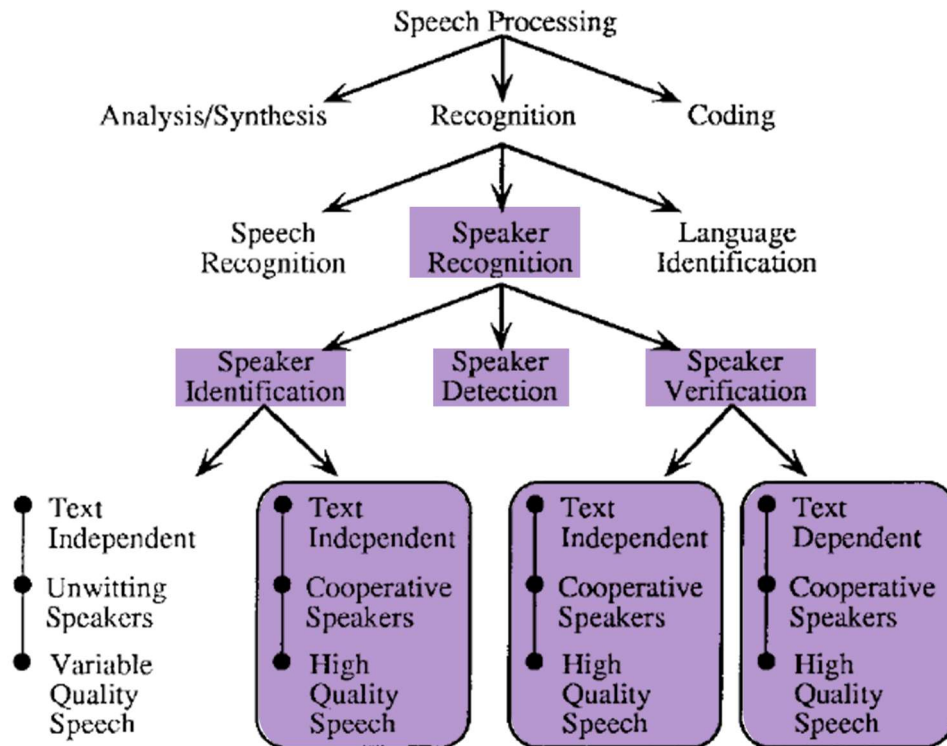


Fig. 1. Speech processing.

(*Campbell*, FIG. 1 (annotated).)

153. A POSITA would have found it obvious to use such a speaker recognition system to identify the user at a tablet and use that information to select the user's message model to improve speech recognition accuracy. *Ditzik, Roberts, Tchorzewski*, and *Campbell* are in the same field of endeavor as each other and the '737 patent, all of which all relate to improving the accuracy of speech recognition systems. A POSITA would have understood adding *Campbell's* design to the *Ditzik, Roberts, Tchorzewski* combination would have improved accuracy because

its automatic speaker recognizer “performs with 98.9% correct identification,” which is far more accurate and secure than pre-existing methods using IDs and passwords. (*See Campbell*, 1437-38.) Correct identification of user is important for training the recognition system and ultimately for accurate speech recognition. A POSITA would have found it desirable to use *Campbell*’s automatic speaker recognition to improve convenience and efficiency by obviating the need for a manual input of a user’s identification information.

154. A POSITA would have done so with a reasonable expectation of success. Applying *Campbell*’s automatic speaker recognition would have simply involved combining known elements according to known methods to yield the predictable result of improved speech recognition. *Campbell* too states that a “speaker-recognition system using methods presented here is practical to implement in software on a modest personal computer” and a POSITA would have found it combinable with the combined *Ditzik, Roberts, Tchorzewski* system, which uses standard tablet computer programs. It also would have been obvious to try, simply choosing from a finite number of identified predictable solutions (e.g., manual speaker identification v. automatic speaker identification) with a reasonable expectation of success.

3. Claim 8: “The system of claim 7, wherein the system is configured to access the particular user message model via a wide area network.”

155. As explained for claim 15, *Ditzik*, *Roberts*, and *Tchorzewski* would have rendered this limitation obvious. (§ IX.A.8.) A POSITA would have understood that the Internet and cable networks disclosed in *Ditzik* are wide area networks, which would be used to load the templates.

E. Ground 5: Claims 1, 2, 4, 12, 13, 16, 19 Would Have Been Obvious Over *Root*, *Roberts*, and *Lewis*

1. Overview of *Root*

156. *Root* discloses a computer system including (1) a display that accepts handwriting input via stylus and (2) interfaces for accessing a portable personality database that contains user preferences. (*Root*, Abstract, 1:10-13, 3:44-47.) The interfaces are “particularly adapted to a user’s habits,” for example, “a handwriting recognition type of interface must have information about the user’s handwriting style,” and “a voice recognition user interface must have specific information about the user’s voice such as pitch, intonation, and accent.” (*Root*, 1:38-45.) This user-specific handwriting and speech information is stored in a database, for example, a PCMCIA card, which can be inserted into a computer to transfer the user-specific information. (*Root*, 5:28-42.)

2. Overview of *Lewis*

157. *Lewis* discloses “conventional machine recognition systems” and explains that such conventional systems “recognize input data by decoding that data using a plurality of trained models.” (*Lewis*, [0004].) For example, *Lewis* explains “speech recognition systems decode input speech utterances using trained speech models to recognize spoken words,” and “handwriting recognition systems recognize input handwriting data using trained character models to convert the handwriting data to machine printed text.” (*Id.*)

3. Claim 1

a. 1[pre] “A system for generating text responsive to voice and handwriting input, comprising:”

158. If the Board finds the preamble limiting, *Root* discloses it. *Root* discloses “a personalized computer user interface system,” including “a trained voice recognition system” that “recognize[s] the spoken commands and words of a particular user” and a “handwriting recognition” that “recognize[s] words written by that particular user.” (*Root*, 3:66-4:4, 4:9-15.) The “[v]oice recognition interfaces translate a user’s spoken words into commands or words that the computer system can interpret.” (*Root*, 6:29-31.) *Root* discloses a personality database of a user’s handwriting where the user’s handwriting pattern is associated with computer-generated text, indicating that the user’s handwriting input is used for generating text. (*Root*, FIG. 3, 5:49-56.) This is shown in Figure 3:

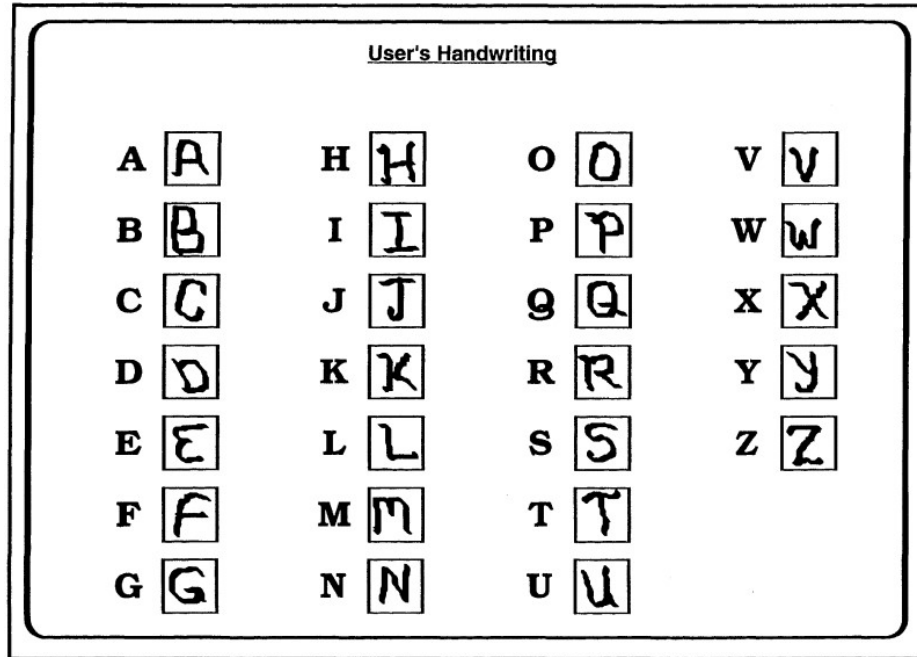


FIG. 3

(*Root*, FIG. 3.)

159. Moreover, “generating text” in response to handwriting and speech input would have been obvious in view of *Lewis*, which provides specific implementation details for the conventional speech and handwriting system used in *Root*. As explained, *Lewis* discloses conventional “speech recognition systems decode input speech utterances using trained speech models to recognize spoken words,” and “handwriting recognition systems recognize input handwriting data using trained character models to convert th[at] handwriting data to machine printed text.” (*Lewis*, [0004].) This would have been an obvious feature to have in *Root*, for the purpose of *Root*’s system is to convert speech and handwriting input into computer text. (*Ditzik*, Abstract; *Mitchell*, 5:59-67.) A POSITA would have

done so with a reasonable expectation of success, simply using known systems according to known methods to achieve predictable results.

- b. 1[a] “a tablet surface, wherein the system is configured to receive freehand input based at least in part on manipulation of a stylus relative to the tablet surface; and”**

160. In my opinion, *Root* renders obvious 1[a]. *Root* discloses that “a number of pen based computer systems have been introduced that allow a user to record information by writing with an **electronic pen** on a **tablet screen**.” (*Root*, 1:24-28.) As shown in Figure 2, *Root* discloses a **tablet surface** receiving **freehand input** based on manipulation of “an **electronic pen 120**” relative to the **tablet surface**. (*Root*, 3:49-52.)

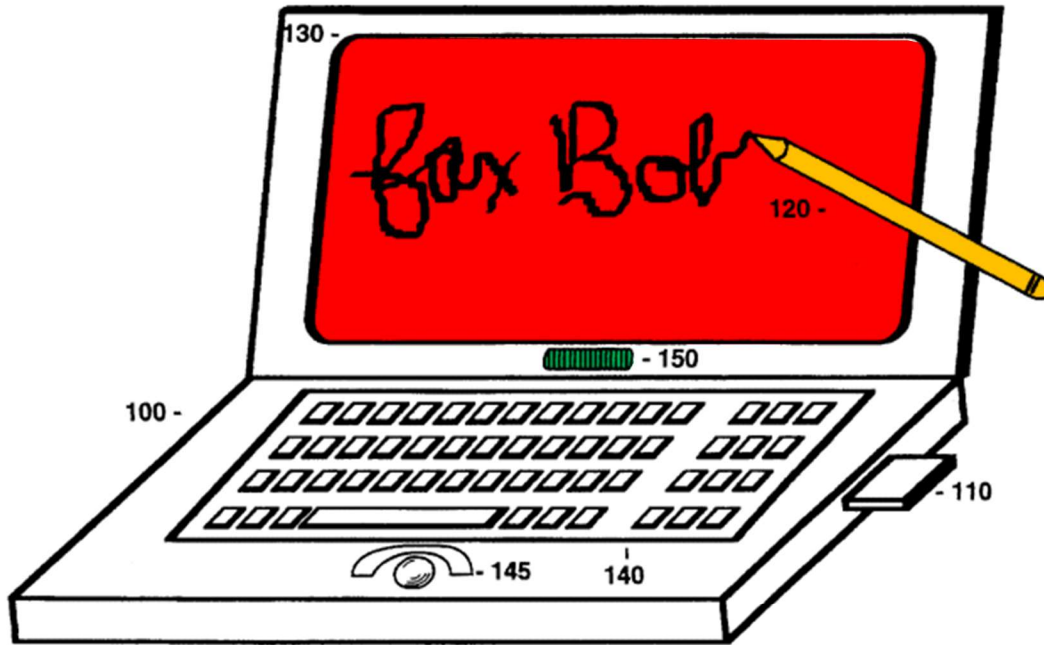


FIG. 1

(*Root*, FIG. 1 (annotated).)

c. 1[b] “a microphone for receiving voice input;”

161. *Root* discloses 1[b]. *Root* discloses that its portable computer system 100 includes a **microphone** for receiving voice input. (*Root*, 3:49-52.)

d. 1[c] “wherein the system is configured to: select a particular user message model from a plurality of user message models;”

162. *Root* alone or in combination with *Roberts* renders 1[c] obvious. *Root* discloses that “[i]f more than one user uses a particular computer system, an operating system with the personalized user interface must have a **personal user**

interface database for each and every user that may use the computer system,” and

“if there are several different computer systems that a user may use, every computer system must contain a copy of that user’s personal preferences.” (*Root*, 1:52-60.) To that end, “a handwriting recognition type of interface must have information about the user’s handwriting style” and “a voice recognition user interface must have specific information about the user’s voice.” (*Root*, 1:41-46.) This system “require[s] a comprehensive database containing information about the specific user who will be using the user interface.” (*Root*, 1:46-48.)

163. *Root* discloses selecting a particular user message model from the collection of user message models for “each and every user” using a “PCMCIA type flash memory card,” that stores “personality profiles,” including a user message model, for each user. (*Root*, 4:50-5:42, 6:25-35, FIG. 2.) When a user inserts the PCMCIA card into the computer, the computer reads “the specific handwriting recognition parameters” and “the set of known test words to learn the user’s particular voice characteristics.” (*Root*, 5:28-42, 6:27-35.) When a different user’s PCMCIA card is inserted, the computer reads a different personality profile for speech and handwriting recognition. (*Id.*)

164. *Root* explains that these personalized models include a handwriting model that “recognize[s] words written by that particular user.” (*Root*, 3:66-4:4, 4:9-15.) *Root* also discloses a “trained voice recognition system” that must “be

trained by the user that will use[] the system.” (*Root*, 4:7-13.) “The user trains the system by speaking a series of known example words,” and the system “uses the information obtained during the training to decipher the user’s speech.” (*Root*, 4:13-15.) A POSITA would have understood this training “information” includes user-specific language models and acoustic models. Indeed, it was well known that a speech and handwriting recognition system includes a language model for satisfactory speech and handwriting recognition. It is well known that the recognizers disclosed by *Root* are augmented to include a language model, so the recognizer knows which words to recognize and how they may be combined, in order to achieve satisfactory recognition performance. It would have been obvious to a POSITA to include a user’s personal speaking and writing patterns in a language model to be stored as part of the personality profile. A POSITA would have found such language model obvious in view of *Roberts*. *Root* discloses a “trained voice recognition system” that “must initially be trained by the user that will use[] the system.” (*Root*, 4:7-13.) “The user trains the system by speaking a series of known example words,” which “uses the information obtained during the training to decipher the user’s speech.” (*Root*, 4:13-15.) A POSITA would have understood that this “information” includes user-specific language models and acoustic models.

165. To the extent *Root* does not expressly disclose its user message model includes a language model and an acoustic and/or handwriting model, however, this would have been obvious in view of *Roberts*. *Roberts* discloses customizing an acoustic model and a language model for each speaker through training. (*Roberts*, 1:65-67.) *Roberts* discloses that the **acoustic model** is made “from one or more utterances of a word” (*Roberts*, 2:37-41) and is updated through a user input, e.g., correcting a misrecognized word by picking an alternative word from a menu of alternative word choices. (*Roberts*, 2:41-47.) *Roberts* also discloses a **language model**, specifically a bigram language model, which “indicates word probabilities based on the language context of the word to be recognized.” (*Roberts*, 4:59-62; *see also id.*, 19:23-31.) *Roberts* discloses that the language model is updated based on user correction of misrecognized words. (*Roberts*, 4:51-59, 13:3-60.) A POSITA would have understood that *Root* as modified by *Roberts* stores a plurality of user message models containing user-specific acoustic models, language models, and handwriting models, and selects a particular user message model for a specific user.

166. A POSITA would have been motivated to combine *Root* and *Roberts*. First, *Root* and *Roberts* are in the same field of endeavor, where speech recognition is combined with some type of user input (e.g., handwriting input in *Root* and user selection of correct words in *Roberts*) to improve overall recognition accuracy.

Second, *Root* already discloses storing user-specific models for speech and handwriting recognition, and *Roberts* simply provides implementation details for *Root*'s speech recognition system and explains that speech recognition systems known at the time used acoustic models and language models.

167. A POSITA would have done so with a reasonable expectation of success. Combining *Root*'s trainable speech and handwriting recognition with *Roberts*'s user-specific acoustic models and language models would have involved combining known elements according to known speech and handwriting recognition methods to yield more accurate recognition systems tailored to each user. It also would have involved using a known speech recognition technique disclosed in *Roberts* to improve similar speech and handwriting recognition system disclosed in *Root*.

- e. **1[d] “adjust a language model of a local message model based at least in part on the particular user message model, wherein the local message model includes the language model, a handwriting model, and an acoustic model;”**

168. In my opinion, *Root* and *Roberts* render this limitation obvious.

169. *Root* and *Roberts* render obvious a local message model including a “language model, a handwriting model, and an acoustic model.” (§ IX.E.3.d.) *Root* discloses a trained voice recognition system and a trained handwriting recognition system, which are trained to learn a user's “spoken commands and words of a

particular user” and handwriting. (*Root*, 3:66-4:15.) *Root* discloses that the local computer system 200 includes a speech and handwriting recognition software that downloads the personality profile from the PCMCIA card inserted into the computer. (*Root*, 5:29-42, 6:25-63.) A POSITA would have understood that the software includes a local message model for recognizing speech and handwriting without inserting the PCMCIA card, which was a well-known feature in the art. (*Ditzik*, 3:43-48, 3:53-56, 5:33-37; *Roberts*, 4:51-62.) In whatever format the user message model is stored on the PCMCIA card, the two recognizers must be able to read them and incorporate them into their run-time recognition software; in other words, the recognizers are constructed to operate with a user specific acoustic and handwriting model, so they have a default model, into which the user specific model is loaded.

170. Next, *Root* and *Roberts* render obvious adjusting the language model of that local message model based in part on the particular user message model selected in limitation 1[c]. *Roberts* discloses adjusting “a language model of a local message model based at least in part on the particular user message model” for the reasons explained for 1[d] in Ground 1 (§ IX.E.3.e.)

171. A POSITA would have been motivated to apply *Roberts*’s language model updating step to *Root*’s recognition system, such that *Root*’s language models are trained to increase recognition accuracy. As explained for 1[c]

(§ IX.E.3.d), *Root* and *Roberts* are in the same field of endeavor. Further, *Root* discloses adapting its recognition user interface “to a user’s habits,” including speech habits and handwriting habits (e.g., *Root*, 1:38-48), and *Roberts* simply discloses implementation details for this training process, including adjusting the language model. A POSITA would have made such an adjustment with a reasonable expectation of success because combining *Root* and *Roberts* this way would have involved combining known elements according to known speech and handwriting recognition methods to yield a predictable recognition system that is trainable using user-specific information.

f. 1[e] “generate text data based at least in part on the freehand input, the adjusted language model, and the handwriting model; and”

172. In my opinion, *Root*, *Roberts*, and *Lewis* render this limitation obvious.

173. *Root* discloses generating text data because ““text can be written on the screen with an electronic pen and a computer will decipher what words have been written on the screen.” (*Root*, 4:3-6.) A POSITA would have understood “deciphering” to mean recognizing handwriting input and displaying it on the display. Indeed, Figure 3 of *Root* discloses “a screen display of handwriting information” displaying a user’s handwriting input in boxes and corresponding computer-generated text on the left of each box. (*Root*, FIG. 3, 5:49-56.)

174. The generated text is based at least in part on the freehand input (*Root*, FIGS 1, 3, 4, 3:66-4:3, 5:53-56.) and handwriting model (“handwriting information” described in *Root*, 5:49-53, 5:57-60.) As explained for 1[c], *Root* combined with *Roberts* would have rendered the adjusting a language model of a local message model obvious. (§ IX.E.3.d.) A POSITA would have understood that the combination of *Root* and *Roberts* outputs text representing the user’s input.

175. It would have been obvious to a POSITA that the handwriting recognizer of the *Root* and *Roberts* combination would output text representing the user’s input, with that text being available for further processing by the system. Alternatively, it would have been obvious further in view of *Lewis*. *Lewis* discloses a handwriting recognition system including a “writer prototypes store 26” that “contains one or more writer prototypes (models)” and a “word list/grammar store 24” containing “a plurality of word lists and associated grammars which may be applied to the recognition engine 22.” (*Lewis*, [0019]-[0020].) A POSITA would have understood that the writer prototypes and the word list and associated grammars constitute the language model. “In the default recognition mode, the recognition engine 22 will decode the user’s handwriting data using the writer-dependent prototype 26 associated with the user, as well as the relevant default word lists/grammars 24.” (*Lewis*, [0021].)

176. A POSITA would have been motivated to combine *Root* as modified by *Roberts* with *Lewis*. First, *Lewis* is in the same field of endeavor as *Root* and *Roberts*—a trainable recognition system with user-specific information. Second, *Root* as modified by *Roberts* already discloses a user-trainable handwriting recognition, and *Lewis* provides implementation details for using adjusted language models in *Root*'s handwriting recognition system. Third, a POSITA would have found it desirable to implement *Lewis*'s trainable language models in *Root*'s handwriting recognition system to customize based on the user's handwriting information, such as grammatical habits, to improve accuracy.

177. A POSITA would have done so with a reasonable expectation of success. It would have involved combining prior art elements (using the adjusted language model in handwriting recognition) according to known handwriting recognition techniques to achieve a trainable handwriting recognition system with improved accuracy.

- g. 1[f] “generate text data based at least in part on the voice input, the adjusted language model, and the acoustic model.”**

178. In my opinion, *Root* discloses generating text data based in part on the voice input because it “uses the information obtained during the training to decipher the user's speech” into “words that the computer system can interpret.” (*Root*, 4:13-15, 6:28-33.) As explained in 1[c]-1[d], *Root* in combination with

Roberts discloses a speech recognition system that uses the adjusted language model and the acoustic model to generate text data. (§§ IX.E.3.d-IX.E.3.e.)

4. Claim 2

a. 2[pre] “The system of claim 1, wherein the system is further configured to:”

179. As explained for 1[pre], *Root*, *Roberts*, and *Lewis* render 2[pre] obvious. (§ IX.E.3.a.)

b. 2[a] “identify information relating to a particular user;”

180. *Root* discloses 2[a]. *Root* discloses that “sophisticated personalized user interfaces require a comprehensive database containing information about the specific user who will be using the user interface,” using user interface software that contains “information that uniquely identifies and describes the user.” (*Root*, 1:38-48.) *Root* uses a “PCMCIA card” to identify a user. (*Root*, 5:28-42, 6:38-63.)

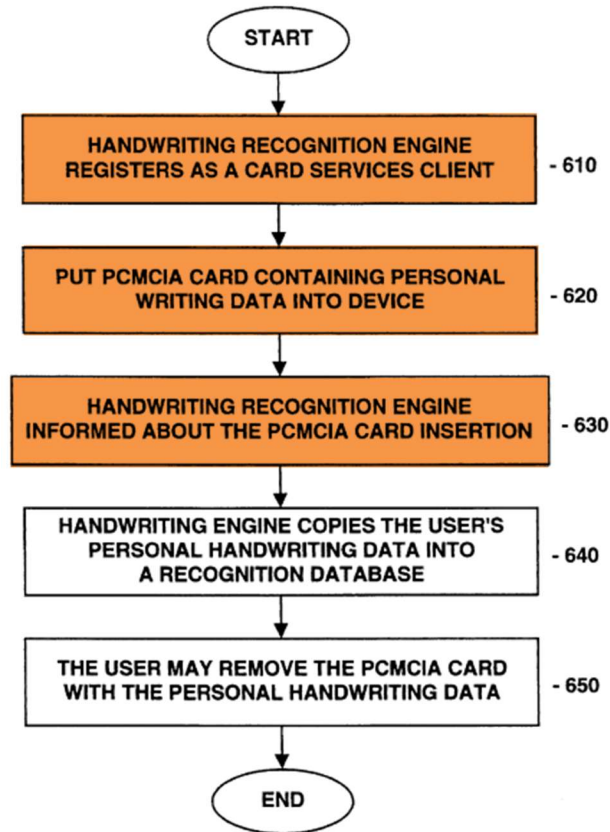


FIG. 6

(*Root*, FIG. 6 (annotated).)

- c. **2[b] “wherein the particular user message model is selected based at least in part on the information relating to the particular user; and,”**

181. *Root* discloses 2[b]. *Root* discloses that when a PCMCIA card for a user is inserted into the computer, *Root*’s “operating system and/or application programs . . . will register as a card services client as mentioned in step 610 of FIG. 6.” (*Root*, 6:39-43.) *Root* further discloses that “any client that need information from the personality profile database can read that information from the personality profile database stored on the PCMCIA card.” (*Root*, 6:53-58.) As explained for

1[c], the personality profile database includes a user message model. (§ IX.E.3.d.)

Thus, *Root*'s system uses the PCMCIA card to select a particular user message model based on the user identification information stored in the card. (*Id.*)

d. 2[c]: “wherein the plurality of user message models includes two or more user message models corresponding to users that are different from the particular user.”

182. In my opinion, *Root* and *Roberts* render obvious this limitation.

183. *Root* discloses “[i]f more than one user uses a particular computer system, an operating system with the personalized user interface must have a personal user interface database for each and every user that may use the computer system.” (*Root*, 1:52-56.) Further, *Root*'s system has access to two or more user message models (comprising a language model, an acoustic model, and a handwriting model in view of *Roberts*) corresponding to different users because it can receive two or more PCMCIA cards, each of which is user specific and carries a particular user's personality profile for speech and handwriting recognition. (*Root*, 5:18-42, 6:25-35.)

5. Claim 4: “The system of claim 1, wherein the system is further configured to train the particular user message model by updating the acoustic model.”

184. *Root*, *Roberts*, and *Lewis* render obvious the system of claim 1.

(§ IX.E.3.) As explained for claim 4 in Ground 1, *Roberts* discloses the additional limitation. (§ IX.A.5.)

185. A POSITA would have been motivated to combine *Roberts*'s process for updating the acoustic model with *Root*'s system, so that the user model is continuously trained and refined to maximize recognition accuracy. (§ IX.E.3.e.) A POSITA would have understood that the accuracy of the user model increases with more input training data, for example, training data for the acoustic model. (*Roberts*, Abstract.) A POSITA would have also had a reasonable expectation of success doing so. Updating the acoustic model was a known concept at the time. (*Mitchell*, 5:48-52.) A POSITA would have had a reasonable expectation of success because combining *Root*'s system *Roberts*'s process would have involved combining known elements according to known recognition methods to yield predictable results of increasing recognition accuracy.

6. Claim 12: “The system of claim 1, wherein the particular user message model includes user-specific acoustic model information and user-specific handwriting model information.”

186. In my opinion, *Root*, *Roberts*, and *Lewis* render the system of claim 1 obvious. (§ IX.E.3.) As explained for limitation 1[c] in Ground 5, *Root* discloses the additional limitation of claim 12. (§ IX.E.3.d .)

187. As explained for 1[c], *Root* discloses the particular user message model. (§ IX.E.3.d.) *Root* discloses “a trained voice recognition system” that “recognize[s] the spoken commands and words of a particular user” and “uses the information obtained during the training to decipher the user’s speech” (user-

specific acoustic model information). (*Root*, 4:7-15.) *Root* also discloses “storing personalized handwriting information,” for example, as shown in Figure 3, “to decipher handwriting” (user-specific handwriting model information). (*Root*, 5:46-56.)

7. Claim 13

a. 13[pre]: “A method, comprising:”

188. Independent claim 13 is substantially similar to claim 1, except that it is directed to a method. *Root*, *Roberts*, and *Lewis* render obvious the claimed method for the same reasons discussed above for claim 1. (§ IX.E.3.)

b. 13[a]: “receiving, at a device, freehand input based at least in part on manipulation of a stylus;”

189. This limitation is substantially identical to limitation 1[a].
(§ IX.E.3.b.) *Root* discloses this limitation as explained for limitation 1[a].
(§ IX.E.3.b.) Moreover, *Root* discloses receiving freehand input at a laptop, which is a device. (*Id.*)

c. 13[b]: “receiving, at the device, audio input based at least in part on information received at a microphone;”

190. This limitation is substantially identical to limitation 1[b].
(§ IX.E.3.c.) *Root* discloses 13[b] as explained for 1[b]. (§ IX.E.3.c.) *Root* discloses receiving audio input through a microphone connected to the laptop. (*Id.*)

- d. **13[c]: “selecting, via the device, a selected user message model from a plurality of user message models;”**

191. This limitation is substantially identical to limitation 1[c].

(§ IX.E.3.d.) *Root* alone or with *Roberts* also renders 13[c] obvious as explained for 1[c]. (§ IX.E.3.d.) A POSITA would have understood that *Root* uses user identification information obtained from a PCMCIA card via the laptop device to select the corresponding user’s user message model from a plurality of user message models for different users.

- e. **13[d]: “adjusting, via the device, a language model of a local message model, wherein the local message model includes the language model, a handwriting model, and an acoustic model, wherein the adjusting the language model is based at least in part on the selected user message model;”**

192. This limitation is substantially identical to limitation 1[d].

(§ IX.E.3.e.) *Root* and *Roberts* render 13[d] obvious as explained for 1[d]. (§ IX.E.3.e.) A POSITA would have understood *Root*, as modified by *Roberts*, adjusts a language model of a local message model via *Root*’s laptop because the laptop receives acoustic and handwriting inputs, which are used to update (adjust) the language model.

- f. **13[e]: “generating, by the device, converted handwriting text data based at least in part on the**

**freehand input, the adjusted language model, and the
handwriting model; and”**

193. This limitation is substantially identical to limitation 1[e]. (§ IX.E.3.f.)

Root, Roberts, and Lewis render 13[e] obvious as explained for 1[e]. (§ IX.E.3.f.)

Root’s laptop device converts handwriting input into computer generated text and displays the text on its display. (*Root*, 6:28-31.)

**g. 13[f]: “generating, by the device, converted speech
text data based at least in part on the audio input, the
adjusted language model, and the acoustic model.”**

194. This limitation is substantially identical to limitation 1[f].

(§ IX.E.3.g.) *Root* discloses this 13[f] as explained for 1[f]. (§ IX.E.3.g.) *Root*’s laptop device “recognize[s] the spoken commands and words of a particular user” and “uses the information obtained during the training to decipher the user’s speech” and generate text. (*Root*, 4:9-15.)

**8. Claim 16: “The method according to claim 13, further
comprising training the selected user message model.”**

195. This limitation is substantially identical to claim 4, which requires training a particular message model by updating the acoustic model. (§ IX.E.5.)

Roberts discloses training a selected user message model for the reasons explained for claim 4. (§ IX.E.5.) As such, *Root* and *Roberts* render obvious the additional limitation of claim 16. (§ IX.E.5.)

9. Claim 19

- a. 19[pre]: “An article of manufacture including a non-transitory computer-readable medium having stored thereon instructions executable by a computing device to cause the computing device to perform operations comprising:”**

196. Independent claim 19 is substantially similar to claim 1, but it is directed to “[a]n article of manufacture including a non-transitory computer-readable medium having stored thereon instructions executable by a computing device to cause the computing device to perform operations.” *Root* discloses the preamble as noted for 1[pre]. (§ IX.E.3.a.) Indeed, *Root*’s portable computer system (e.g., tablet, *Root*, 1:24-28) includes a “processor” (a computing device) and a “memory” (a non-transitory computer-readable medium). (*Root*, Abstract, 3:44-52, Claims 1, 8.) The memory stores instructions for the processor to perform the instructions for the operations, as explained in 19[a]-19[f] below. (§§ IX.E.9.b-IX.E.9.g.)

- b. 19[a]: “receiving freehand input based at least in part on manipulation of a stylus;”**

197. As discussed above for limitation [19pre], *Root* discloses a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input. (*Root*, Abstract, 3:44-52, Claims 1, 8.) As discussed for limitations 1[a] and 13[a], *Ditzik* also discloses its computer receiving freehand input based on manipulation of a

stylus. (§ IX.E.3.b; § IX.E.7.b.) *Root* thus discloses its instructions cause the tablet microprocessor to perform this limitation.

c. 19[b]: “receiving audio input based at least in part on information received at a microphone;”

198. As discussed above for limitation [19pre], *Root* discloses a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input. (*Root*, Abstract, 3:44-52, Claims 1, 8.) As discussed for limitations 1[b] and 13[b], *Root* also discloses its tablet receiving audio input (voice input) based on information received at a microphone. (§ IX.E.3.c; IX.E.7.c.) *Root* thus discloses the instructions cause the tablet microprocessor to perform this limitation.

d. 19[c]: “selecting a particular user message model from a plurality of user message models;”

199. As discussed above for limitation [19pre], *Root* and *Roberts* render obvious a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input. (*Root*, Abstract, 3:44-52, Claims 1, 8.) As discussed for limitations 1[c] and 13[c], *Root* and *Roberts* render obvious selecting a particular user message model from a plurality of user message models. (§ IX.E.3.d; § IX.E.7.d.) *Root* and *Roberts* thus render obvious the instructions causing the tablet microprocessor to perform this limitation.

- e. **19[d]: “adjusting a language model of a local message model based at least in part on the particular user message model, the local message model including the language model, a handwriting model and an acoustic model;”**

200. As discussed above for limitation [19pre], *Root* discloses a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input. (*Root*, Abstract, 3:44-52, Claims 1, 8.) As discussed for limitations 1[d] and 13[d], *Root* and *Roberts* render obvious adjusting a language model of a local message model based at least in part on the particular user message model, the local message model including a language, handwriting, and acoustic model. (§ IX.E.3.e; § IX.E.7.e.) *Root* and *Roberts* thus render obvious the instructions causing the tablet microprocessor to perform this limitation.

- f. **19[e]: “generating converted handwriting text data based at least in part on the freehand input, on the adjusted language model, and on the handwriting model; and”**

201. As discussed above for limitation [19pre], *Root* discloses a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input. (*Root*, Abstract, 3:44-52, Claims 1, 8.) As discussed for limitations 1[e] and 13[e], *Root* and *Roberts* render obvious generating converted handwriting text data based on the freehand input, adjusted language model, and handwriting model. (§ IX.E.3.f;

§ IX.E.7.f.) *Root* and *Roberts* thus render obvious the instructions causing the tablet microprocessor to perform this limitation.

- g. 19[f]: “generating converted speech text data based at least in part on the audio input, on the adjusted language model, and on the acoustic model.”**

202. As discussed above for limitation [19pre], *Root* discloses a tablet with a memory storing instructions to be executed by the microprocessor to perform operations that generate text data based on speech and handwriting input. (*Root*, Abstract, 3:44-52, Claims 1, 8.) As discussed for limitations 1[f] and 13[f], *Root* and *Roberts* render obvious generating converted speech text data based on the audio (voice) input, adjusted language model, and acoustic model. (§ IX.E.3.g; § IX.E.7.g.) *Root* and *Roberts* thus render obvious the instructions causing the tablet microprocessor to perform this limitation.

F. Ground 6: Claims 3, 5, 6, 11, and 14 Would Have Been Rendered Obvious Over *Root*, *Roberts*, *Lewis*, and *Verrier*

- 1. Claim 3: “The system of claim 2 wherein the microphone is housed in the stylus.”**

203. *Root*, *Roberts*, *Lewis*, and *Verrier* render obvious claim 3. As discussed with respect to claim 3 in Ground 2, *Verrier* discloses the additional limitation of claim 3. (§ IX.B.4.)

204. POSITA would have been motivated to combine *Root*, *Roberts*, and *Lewis* with *Verrier*. Each discloses receiving user input. Indeed, *Root* discloses using a stylus for inputting information to the tablet. (§ IX.E.3.b.) It was well

known at the time that a stylus could be used to input information, like user identifying information or voice and handwriting input information. A POSITA would have found it obvious to add a microphone to *Root*'s stylus to more smoothly transition between speech and handwriting input and avoid clashes between the two types of inputs. (*Verrier*, 3:4-36.) A POSITA would have had done so with a reasonable expectation of success. *Root* already uses a stylus (§ IX.E.3.b); thus, replacing that stylus with *Verrier*'s would have merely involved a simple substitution of one known element for another to obtain predictable results (e.g., personalized speech and handwriting recognition for improved accuracy). It would also have involved combining prior art tablet and personalized stylus elements according to known methods to yield predictable results and adding a known personalized stylus in *Verrier* to a known device in *Root* in the same way.

2. Claim 5: “The system of claim 1, further comprising a proximity sensor configured to sense a proximity of a user to the microphone.”

205. For the same reasons provided in Ground 2, *Root*, *Roberts*, *Lewis*, and *Verrier* render claim 5 obvious. (§ IX.B.5.)

3. Claim 6: “The system of claim 1, further comprising an identification circuit configured to identify an owner of the stylus, wherein the system is configured to select the

particular user message model at least in part in response to the identification circuit.”

206. For the same reasons provided in Ground 2, *Root*, *Roberts*, *Lewis*, and *Verrier* would have rendered claim 6 obvious. (§ IX.B.6.) As explained for limitation 2[b], *Root* and *Roberts* would have rendered obvious selecting the particular user message model based in part on the information relating to the particular user in response to the identification circuit. Using an identification circuit in the stylus would have been an obvious modification a POSITA would implement with a reasonable expectation of success, as discussed for claim 6 in Ground 2. (§ IX.B.6.)

4. **Claim 11: “The system of claim 1, further comprising an activation circuit configured to sense movement of the stylus toward the tablet surface, wherein the system is further configured to receive the freehand input in response to the activation circuit sensing the movement.”**

207. For the same reasons provided in Ground 2, *Root*, *Roberts*, *Lewis*, and *Verrier* render claim 11 obvious. (§ IX.B.7.)

5. **Claim 14: “The method according to claim 13, further comprising switching from receiving audio input to receiving freehand input in response to detecting movement of the stylus towards a tablet.”**

208. For the same reasons provided in Ground 2, *Root*, *Roberts*, *Lewis*, and *Verrier* render claim 14 obvious. (§ IX.B.8.)

G. Ground 7: Claims 9 and 10 Would Have Been Obvious Over *Root*, *Roberts*, *Lewis*, and *Bellegarda*

- 1. Claim 9: “The system of claim 1, wherein the system adjusting the language model includes adjusting a syntactic model and a semantic model included in the language model.”**

209. For the same reasons provided in Ground 3, *Root*, *Roberts*, *Lewis*, and *Bellegarda* render claim 9 obvious. (§ IX.C.3.) A POSITA would have been motivated to combine these disclosures with a reasonable expectation of success for the same reasons explained in Ground 3. (*Id.*) *Bellegarda* discloses known implementation details for language models used in the *Root/Roberts/Lewis* system. (*Id.*)

2. Claim 10

- a. 10[pre]: “The system of claim 9, wherein the semantic model comprises at least one semantic cluster selected from the group consisting of:”**
- b. 10[b]: “a semantic cluster of a second type that is a discrete grouping of text having a particular context; and”**
- c. 10[c]: “a semantic cluster of a third type that is defined with respect to an individual reference word that is indicative of a particular context.”**

210. For the same reasons provided in Ground 3, *Root*, *Roberts*, *Lewis*, and *Bellegarda* render claim 10 obvious. (§ IX.C.4.) A POSITA would have been motivated to combine these disclosures with a reasonable expectation of success for the same reasons explained in Ground 3. (*Id.*)

H. Ground 8: Claims 7, 8, and 15 Would Have Been Obvious Over *Root, Roberts, Lewis, and Campbell*

- 1. Claim 7: “The system of claim 1, wherein the system is configured to select the particular user message model based at least in part on an identifying signal received from the microphone.”**

211. *Root, Roberts, and Lewis* render the system of claim 1 obvious. (See IX.E.3.)

212. As explained for 1[c], *Root* alone or with *Roberts* renders obvious selecting a particular user message model from a plurality of user message models. (§ IX.E.3.d.) A POSITA would have further combined *Root, Roberts, and Lewis*, with *Campbell* with a reasonable expectation of success for the same reasons explained in Ground 4. (§ IX.D.) *Campbell* provides a known method to improve accuracy of the *Root/Roberts/Lewis* system. (*Id.*)

- 2. Claim 8: “The system of claim 7, wherein the system is configured to access the particular user message model via a wide area network.”**

213. As explained for limitation 1[c], *Root* alone or with *Roberts* renders obvious accessing a particular message model. (§ IX.E.3.d.) *Root* further discloses its invention may “comprise a general purpose computer,” (*Root* 3:21-23), which was well known at the time to include network interfaces and access remote servers, filesystems, and databases. To the extent *Root* does not expressly disclose accessing the particular message model via a wide area network, it would have been obvious to do so in view of *Campbell*.

214. *Campbell* discloses “ubiquitous telephone network and microphones bundled with computers.” (*Campbell*, 1438.) A POSITA would have understood that the “telephone network” is a wide area network allowing Internet access.

215. A POSITA would have found it obvious use the Internet to access the user message model in the combined *Root*, *Campbell*, *Lewis* system. *Root* discloses that “sophisticated personalized user interfaces require a comprehensive database containing information about the specific user who will be using the user interface.” (*Root*, 1:46-48.) It was well known at the time that it is desirable to store a large amount of data (e.g., user message models for different users) on a server, for example, a centralized server, which can be queried for a specific user model. (*Tchorzewski*, 2:1-7; *Mitchell*, 12:23-28.) A POSITA would have found it obvious that a client device, such as *Root*’s computer, can access the user message models on the server using the Internet. Storing the user profile on a networked server would also have been advantageous because the system would still work in case the user misplaced or forgot their PCMCIA card, without which the system would have very degraded performance. In the absence of the PCMCIA card, a user would enter a user name and password and, once verified, the user’s personal profile would be loaded from the network without need to carry the physical device. These were already well-known techniques for authenticating a user and access computer resources that were not physically present on that computer.

216. A POSITA would have done so with a reasonable expectation of success. Combining *Root*'s system as modified by *Roberts* and *Lewis* with *Campbell* to access the user model via the wide area network would have involved combining prior art elements according to known computer networking methods to yield a speech recognition system that can efficiently access a large database of user message models. It would also have been obvious to try given that the user message models can be stored either locally or elsewhere, and a POSITA would have understood that it would be more efficient to store a large amount of user message model data on a server rather than on a multitude of PCMCIA cards.

3. Claim 15: “The method according to claim 13, further comprising accessing the selected user message model via a network.”

217. *Root*, *Roberts*, and *Lewis* render the method of claim 13 obvious. (§ IX.E.7.) For the reasons explained for claim 8, this additional limitation would have further been obvious in view of *Campbell*. (§ IX.H.2.)

X. Conclusion

218. For the reasons set forth above, I believe claims 1-16 and 19 of the '737 patent would have been obvious in view of the prior art.

219. In signing this Declaration, I understand it will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I understand I may be subject to cross-examination in this case, and that any cross-examination will take place in the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

220. I declare that all statements made herein of my knowledge are true, and that all statements made on information and belief are believed to be true, and that those statements were made with the knowledge that willful false statements and the like so made are punishable by fine, imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Respectfully submitted,

Date: September 11, 2023

By: 
Christopher Schmandt