

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

APPLE INC.
Petitioner

v.

ADVANCED VOICE RECOGNITION SYSTEMS, INC.
Patent Owner

Case No. 2019-01352
U.S. Patent No. 7,558,730

**PETITION FOR *INTER PARTES* REVIEW
OF U.S. PATENT NO. 7,558,730**

TABLE OF CONTENTS

I. INTRODUCTION	1
II. SUMMARY OF THE '730 PATENT	1
A. DESCRIPTION OF THE ALLEGED INVENTION OF THE '730 PATENT.....	1
B. SUMMARY OF THE PROSECUTION HISTORY OF THE '730 PATENT	4
C. LEVEL OF SKILL OF A PERSON HAVING ORDINARY SKILL IN THE ART.....	6
III. REQUIREMENTS FOR INTER PARTES REVIEW UNDER 37 C.F.R.	
§ 42.104	6
A. GROUNDS FOR STANDING UNDER 37 C.F.R. § 42.104(A)	6
B. IDENTIFICATION OF CHALLENGE UNDER 37 C.F.R. § 42.104(B) AND RELIEF REQUESTED	7
C. CLAIM CONSTRUCTION UNDER 37 C.F.R. § 42.104(B)(3)	8
IV. THE CHALLENGED CLAIMS OF THE '730 PATENT ARE UNPATENTABLE	12
A. <u>GROUND 1</u> : <i>SKLADMAN</i> IN VIEW OF <i>OBILISSETTY</i> RENDERS CLAIMS 15 AND 17 OBVIOUS 12	
B. <u>GROUND 2</u> : <i>SKLADMAN</i> IN VIEW OF <i>OBILISSETTY</i> IN FURTHER VIEW OF <i>SCHRAGE</i> RENDERS CLAIMS 15 AND 17 OBVIOUS	44
C. <u>GROUND 3</u> : <i>COHN</i> IN VIEW OF <i>OBILISSETTY</i> RENDERS CLAIMS 15 AND 17 OBVIOUS ...	50
V. CONCLUSION	73
VI. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8(A)(1)	74
A. REAL PARTY-IN-INTEREST	74
B. RELATED MATTERS.....	74
C. LEAD AND BACK-UP COUNSEL.....	74

I. INTRODUCTION

Petitioner Apple Inc. (“Petitioner”) requests an *Inter Partes* Review (“IPR”) of claims 15 and 17 (collectively, the “Challenged Claims”) of U.S. Patent No. 7,558,730 (“the ’730 Patent”). *’730 Patent* (EX1001).

II. SUMMARY OF THE ’730 PATENT

A. Description of the alleged invention of the ’730 Patent

The ’730 Patent generally describes systems and methods for “speech recognition and transcription among users employing incompatible protocols for generating, transcribing, and exchanging speech.” *’730 Patent* (EX1001) at Abstract. The ’730 Patent purportedly solves the problem of incompatible protocol use within prior speech recognition/transcription systems by utilizing a uniform system protocol for routing by a System Transaction Manager and transcription at a Speech Recognition and Transcription Engine. The system enables requesting and recipient users each to communicate between systems using otherwise incompatible protocols. *Id.*

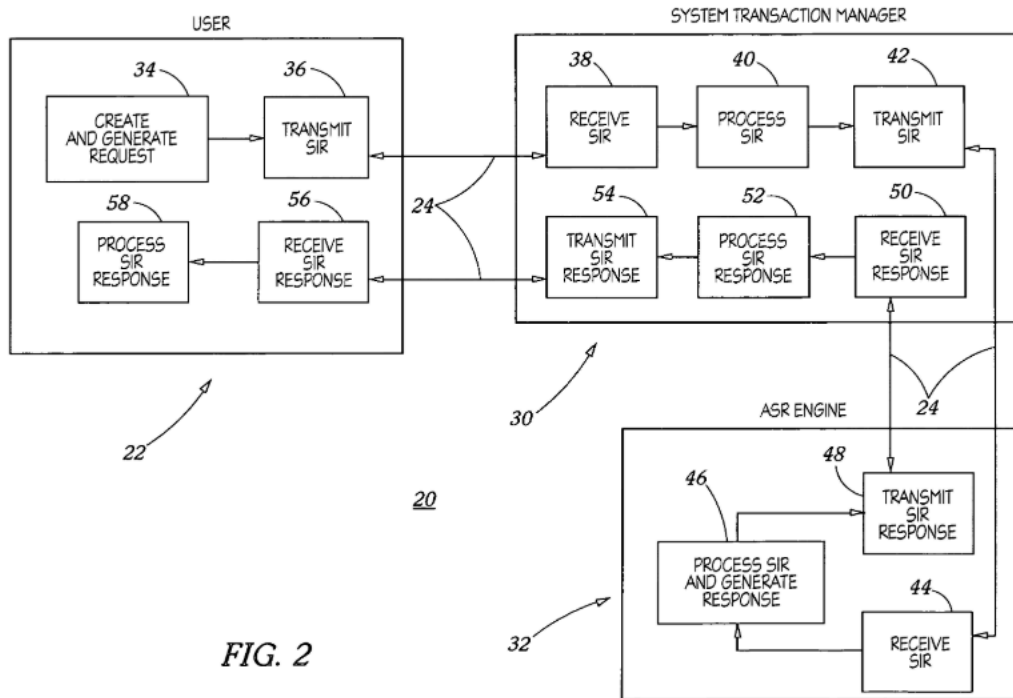


FIG. 2

As illustrated above in Fig. 2, a Speech Information Request (SIR) is generated and includes spoken text in a particular format for the system and optional formatted spoken embedded commands. *Id.* at 10:12-14, 10:32-35. Receipt of the SIR at the System Transaction Manager 30 via communications link 24 activates the System Transaction Manager 30 to perform functions such as decryption, validation, user authentication, and subsequent processing. *Id.* at 11:66-12:8, 12:23-35, 12:45-48. The System Transaction Manager 30 creates a data record, called a Job, by extracting speech data (such as digitized audio) from the formatted message portion of the Request. *Id.* at 12:9-22. This may be done, for example, by parsing header and routing data from the SIR to locate and retrieve the speech data. *Id.* at 13:10-22. The System Transaction Manager may also optionally

store the SIR in one or more bins and/or assign a priority level to the SIR before transmitting it to a Speech Recognition and Transcription Engine. *Id.* at 12:45-13:40.

Receipt of the Request activates the Speech Recognition and Transcription Engine 32. *Id.* at 14:12-14. If Engine 32 can process the Request, then the Request is sent to the Engine 32 for processing and generation of a Response. *Id.* at 14:15-22. The Speech Recognition and Transcription Engine 32 may further access any additional data needed to respond to the SIR, for example, transcribe spoken text or execute spoken commands. *Id.* at 13:65-14:2.

Speech Recognition and Transcription Engine 32 processes the SIR, transcribes the speech therein, and generates a Response to the SIR in the form of a formatted transcription of the speech. *Id.* at 14:49-53. Speech Recognition and Transcription Engine 32 then transmits the Response to the System Transaction Manager via communications link 24. *Id.* at 15:4-7.

The System Transaction Manager 30 transmits the completed Response to its intended Recipients, which may include the requesting user, non-requesting users, and/or passive users. *Id.* at 16:15-18. The system 20 may employ a User Application Service Adapter 80 (“UASA”) to provide an interface between the System Transaction Manager 30 and recipients. *Id.* at 16:23-28. After receiving a

compatible Response, the requesting user 22 (or any recipient) may process the Response as necessary for internal use. *Id.* at 16:42-44.

B. Summary of the prosecution history of the '730 Patent

The '730 Patent issued on an application (the "'794 Application") filed July 3, 2007. *See* '730 Patent (EX1001). The '730 Patent is a continuation of an application filed on November 27, 2001. *Id.* For purposes of this IPR, it is assumed that all Challenged Claims are entitled to a priority date of November 27, 2001.

Prosecution of the '730 Patent included only one Office Action dated November 27, 2007, which included a final rejection under 35 U.S.C. §103(a) of claims 1-17 as being unpatentable over U.S. Patent No. 6,434,526 to Cilurzo ("Cilurzo") and a textbook entitled "Computer Networks" by Tanenbaum ("Tanenbaum").¹ *See, Cilurzo* (EX1017) and *Tanenbaum* (EX1018).

Applicants responded to the §103(a) rejection with an Amendment dated September 15, 2008, in which claims 1, 2, 8, 12, 14, and 16 were amended and

¹ The November 27, 2007, Office Action was made final because the Office determined that all claims as-filed in the '749 Application were drawn to the same invention claimed in the parent application and could have been finally rejected therein based upon the same prior art. The prosecution history of its parent application (the "'849 Application") is included for reference as Exhibit 1019. After multiple examiner interviews, the Office withdrew finality of the prior Action by letter dated April 14, 2008. Applicants replied to the §103(a) rejection by Amendment on September 15, 2008, which is discussed briefly in the body of this Petition.

new claims 18 and 19 were added. Representative amendment to claim 1 clarified that the system transaction manager uses “a uniform system protocol,” that first and second users employ a first and second “user legacy” protocol, that second user legacy protocol is “different than the first user legacy protocol,” and that the system includes “an application service adapter configured to provide bi-directional translation between the first user legacy protocol and the uniform system protocol and between the second user legacy protocol and the uniform system protocol.” See *'730 File History* (EX1002) at pp. 123-129. Applicants also distinguished *Cilurzo* as not:

teach[ing] that the speech server 300 uses a uniform system protocol, or that a first computer system 100 and a second computer system 100 utilize first and second user legacy protocols, respectively, where the first and second user legacies are different from one another[or]

* * * *

an application service adapter (“ASA”) configured to provide bi-directional translation between the first user legacy protocol and the uniform system protocol and between the second user legacy protocol and the uniform system protocol.

Id. at p. 131. Applicants also clarified the foregoing limitations all reside in the Application System Adapter by emphasizing Paragraph [0031] of the Specification as filed. *Id.* at p. 131-132. Applicants further cited Toro Co. v.

White Cons. Indus., 199F.3d 1295, 1301 53 USPQ2d 1065, 1069 (Fed. Cir. 1999)

for the proposition that “[w]here an explicit definition is provided by an applicant for a term, that definition will control interpretation of the term as it is used in the claim.” *Id.* at p. 132.

Allowance was granted on March 6, 2009, with reasons provided only as “The arguments submitted by the applicant are persuasive and claims 1-8 and 10-19 are deemed allowable over the cited prior art of record.” *Id.* at p. 149.

C. Level of skill of a person having ordinary skill in the art

A person having ordinary skill in the art (PHOSITA) at the time of the ’730 Patent would have at least a bachelor’s degree in electrical engineering, computer engineering, computer science, electrical and computer engineering, or in some other field related to information sciences and one year of experience in the area of telecommunications or a related field. Additional industry experience or technical training may offset less formal education, while advanced degrees or additional formal education may offset lesser levels of industry experience. *Creusere Dec.* (EX1003) at ¶¶30-32.

III. REQUIREMENTS FOR INTER PARTES REVIEW UNDER 37 C.F.R. § 42.104

A. Grounds for standing under 37 C.F.R. § 42.104(a)

Petitioner certifies that the ’730 Patent is available for IPR and that the Petitioner is not barred or estopped from requesting IPR challenging the claims of

the '730 Patent. Specifically, Petitioner states: (1) Petitioner is not the owner of the '730 Patent, (2) Petitioner has not filed a civil action challenging the validity of any claim of the '730 Patent, and (3) this Petition is filed less than one year after the Petitioner was served with a complaint alleging infringement of the '730 Patent.

B. Identification of challenge under 37 C.F.R. § 42.104(b) and relief requested

In view of the prior art and evidence presented, claims 15 and 17 of the '730 Patent are unpatentable and should be cancelled. 37 C.F.R. § 42.104(b)(1). Further, based on the prior art references identified below, IPR of the Challenged Claims should be granted. 37 C.F.R. § 42.104(b)(2).

Proposed Grounds of Unpatentability	Exhibits
<u>Ground 1:</u> Claims 15 and 17 are obvious under § 103(a) over U.S. Patent No. 6,487,278 to Skladman et al. (" <i>Skladman</i> ") in view of U.S. Patent No. 7,558,735 (" <i>Obilisetty</i> ") ²	EX1004, EX1005
<u>Ground 2:</u> Claims 15 and 17 are obvious under § 103(a) over U.S. Patent No. 6,487,278 to <i>Skladman</i> in view of <i>Obilisetty</i> in further view of U.S. Patent No. 6,850,609 to Schrage (" <i>Schrage</i> ").	EX1004 EX1005 EX1022
<u>Ground 3:</u> Claims 15 and 17 are obvious under § 103(a) over U.S.	EX1006, EX1005

² Though the patent numbers of *Obilisetty* and the '730 Patent are very close, *Obilisetty* still qualifies as prior art because it issued on July 7, 2009, based on an application filed December 28, 2000. Therefore qualifies as prior art with regard to the '730 Patent under 35 U.S.C. §102(e).

Proposed Grounds of Unpatentability	Exhibits
Patent No. 6,064,723 to Cohn et al. (“ <i>Cohn</i> ”) in view of <i>Obilisetty</i>	

Section IV identifies where each element of the Challenged Claims is found in the prior art. 37 C.F.R. § 42.104(b)(4). The exhibit numbers of the supporting evidence relied upon to support the challenges are provided above and the relevance of the evidence to the challenges raised are provided in Section IV. 37 C.F.R. § 42.104(b)(5). **Exhibits EX1001–EX1022** are also attached.

C. Claim construction under 37 C.F.R. § 42.104(b)(3)

In an *inter partes* review proceeding, a claim of a patent . . . shall be construed using the same claim construction standard that would be used to construe the claim in a civil action under 35 U.S.C. 282(b), including construing the claim in accordance with the ordinary and customary meaning of such claim as understood by one of ordinary skill in the art and the prosecution history pertaining to the patent.” 37 C.F.R. §42.100(b) (as amended, effective Nov. 13, 2018); *see also, Phillips v. AWH Corp.*, 415 F.3d 1303, 1312 (Fed. Cir. 2005). For all of the claim terms not expressly construed below, Petitioner applies the ordinary and customary meaning of the claim terms as understood by a PHOSITA.

In the co-pending litigation, the Court has issued a claim construction ruling adopting a series of constructions that had been stipulated by the parties. Claim Construction Ruling (EX1021). Petitioner respectfully submits that these agreed-

upon constructions should be applied in this proceeding. Many constructions, as indicated below, are taken directly from a set of explicitly defined terms found in the Specification. *See, e.g., '730 Patent* (EX1001) at 5:50-8:29 (“System Nomenclature”).

1. “uniform system protocol”

The Court has construed “uniform system protocol” to mean “a protocol that processes information expressed in a normalized data format.” *See, e.g., '730 Patent* at 17:22-24 (“Like the System Transaction Manager 30’, the Uniform System Protocol 90 processes information expressed in the Normalized Data Format.”); 6:50-53 (construing “normalized data format” as “a uniform internal data format used for handling Speech Information Requests and Responses with System components within the Speech Recognition and Transcription System.”).

2. “first protocol”

The Court has construed “first protocol” to mean “a first group of processes that a User and/or an ASR employs to directly support some business process or transaction and is accessed using a Native Communications Protocol.” *See, e.g., id.* at 6:61-64 (providing explicit definition of “protocol”); 6:44-49 (defining “native communications protocol” to be a protocol used in a legacy system).

3. “second protocol”

The Court has construed “second protocol” to mean “a second group of processes that a User and/or an ASR employs to directly support some business process or transaction and is accessed using a Native Communications Protocol.” *See, e.g., id.* at 6:61-64 (providing explicit definition of “protocol”).

4. “application service adapter”

The Court has construed “application service adapter” (“ASA”) to mean “an application layer within the Speech Recognition and Transcription System that provides an interface among Users, Speech Recognition and Transcription Engines, the System Transaction Manager, and other system components by allowing a User’s existing application and/or a system components application to communicate with the Transaction Manager.” *See, e.g., id.* at 5:60-66 (providing explicit definition).

5. “user application service adapter”

The Court has construed “user application service adapter” (“UASA”) to mean “a specific Application Service Adapter that handles formatting and routing of Speech Information Requests and Responses to elements of a User’s Protocol within the Speech Recognition and Transcription System.” *See, e.g., id.* at 8:15-19 (providing explicit definition).

6. “speech information request(s)”

The Court has construed “speech information request(s)” to mean “formatted speech, which can be acted upon by system component, including the System Transaction Manager.” *See, e.g., id.* at 7:57-60 (providing explicit definition).

7. “response(s) to [the] speech information request(s)”

The Court has construed “response(s) to [the] speech information request(s)” to mean “a formatted transcription of formatted speech.” *See, e.g., id.* at 7:11-12 (providing explicit definition).

8. “spoken text”

The Court has construed “spoken text” to mean “audible words.” *See, e.g., id.* at 9:3-5 (noting that “spoken text can be acquired using a microphone coupled to an A/D converter. . . .”, which must constitute “audible words” to be captured by the microphone).

9. “system transaction manager”

The Court has construed “system transaction manager” to mean “a server application that provides a central interconnect point (hub) and a communications interface among system components and Users having disparate or heterogeneous protocols; and, an information router (or bridge or switch) within the Speech Recognition and Transcription System.” *See, e.g., id.* at 7:51-56 (providing explicit definition).

10. “speech recognition and transcription engine”

The Court has construed “speech recognition and transcription engine” to mean “a process running on a computer that recognizes an audio file and transcribes that file to written text to generate a transcription of speech.” *See, e.g., Id.* at 7:57-60 (providing explicit definition).

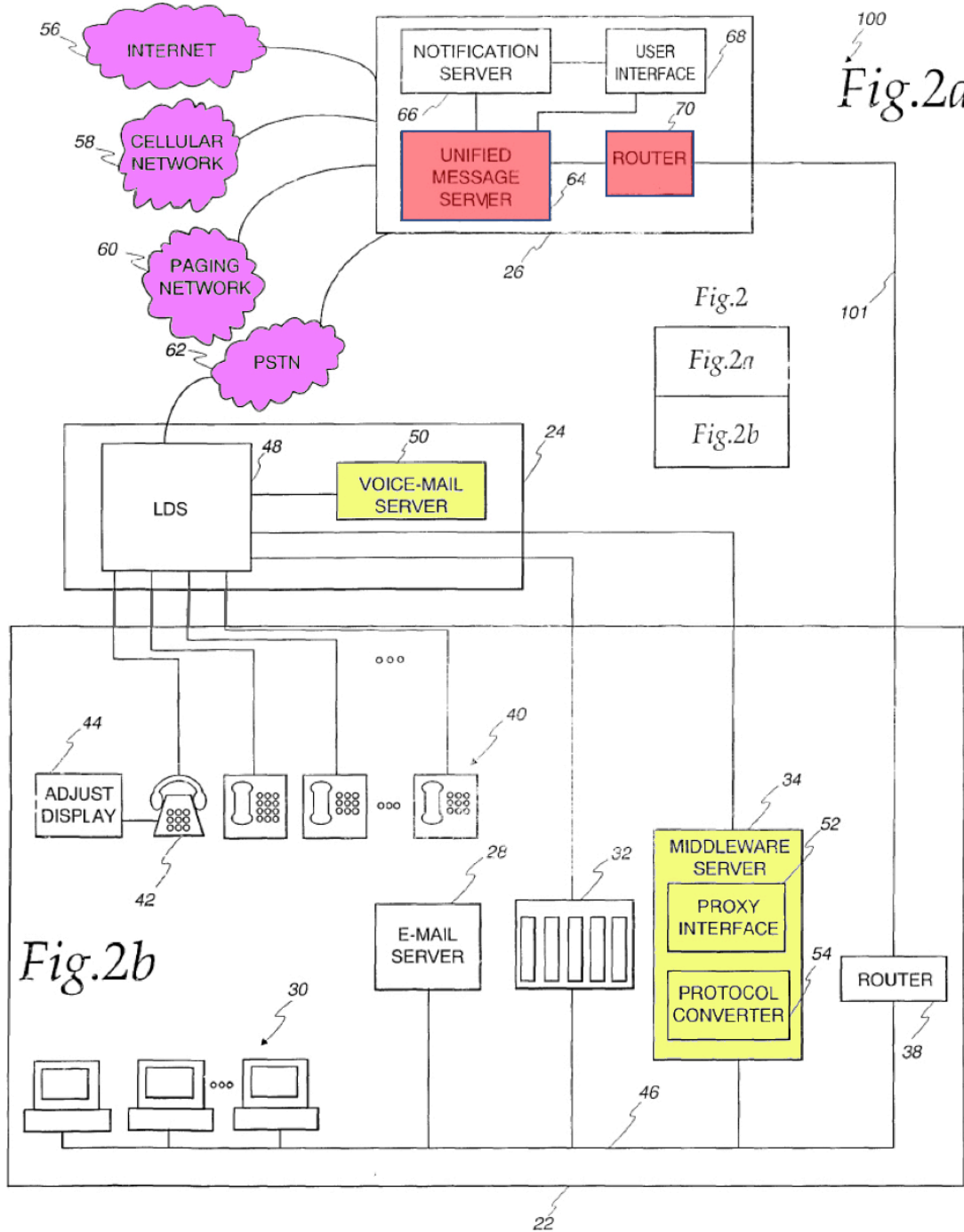
IV. THE CHALLENGED CLAIMS OF THE '730 PATENT ARE UNPATENTABLE

A. Ground 1: *Skladman* in view of *Obilisetty* renders claims 15 and 17 obvious

Skladman was filed February 29, 2000 and issued November 26, 2002. *Skladman* qualifies as prior art under 35 U.S.C. § 102(e) (pre-AIA). *Skladman* (EX1004). *Skladman* was not cited or considered during the original prosecution of the '730 Patent. *Supra*, Section II.B.

Like the '730 Patent, *Skladman* generally teaches a network communications system for enabling communication between messaging systems having disparate capabilities and communications protocols; e.g., connecting a legacy messaging system into a modern unified messaging system without eliminating existing investments into legacy systems. *Id.* at Abstract, 2:4-9; 2:46-3:4; 6:20-23. As illustrated and annotated in Figure 2a and 2b below, *Skladman* provides a Unified Message Server 64 plus a Router 70 (depicted in red) that is centrally located on the Internet and operable to communicate with a legacy messaging system (such as voicemail 50, depicted in yellow) that works in conjunction with a proxy interface

52 and protocol converter 54 (also depicted in yellow) that “translates the retrieved messages [from the legacy system] into one or more predetermined formats usable by the unified message 64” that the user can then access. *Id.* at 4:14-20.



Id. at Figs. 2a and 2b (annotated).

Skladman is therefore from an analogous field to the '730 Patent and reasonably pertinent to the problem to be solved of facilitating communication

between systems employing heterogenous protocols. *See, e.g.*, ‘730 Patent (EX1001) at 3:56-62 (“The present invention provides a system for facilitating speech recognition and transcription among users employing heterogeneous or disparate entity system protocols...”), 7:51-56; *Skladman* (EX1004) at 1:21-24 (“The present invention relates generally to communication systems, and in particular to a communication system for providing unified messaging services for transferring different types of electronic messages.”), 2:7-9 (“Accordingly, there is a need for a method and system of incorporating legacy messaging systems into a modern unified messaging environment.”); *Creusere Dec.* (EX1003) at ¶48.

Obilisetty issued on July 7, 2009, based on an application filed December 28, 2000, and therefore qualifies as prior art with regard to the ‘730 Patent under 35 U.S.C. §102(e). *Obilisetty* was not cited or considered during prosecution of the ‘730 Patent.

Obilisetty discloses systems and methods for “capturing and transcribing dictated information and for delivering the transcribed information to an end user” over a network, including the Internet. *Obilisetty* (EX1005) at 2:45-47. A “voice file containing a digital recording of the dictated information is received via the Internet,” forwarded “to a remote transcription service provider,” and, in turn “delivered to, for example, a facsimile machine, a text server and/or printer, a computer system or Web browser.” *Id.* at 2:53-60.

Like the '730 Patent, *Obilisetty* is directed towards a centralized communications system on the Internet that connects legacy transcription systems to modern transcription systems, that then create a transcription of the speech that is then provided to a user. *Obilisetty* (EX1005), at Abstract; 2:53-3:20; 3:36-50. As illustrated and annotated in Figure 3B below, *Obilisetty* captures voice through its legacy dictation system 320 (depicted in green). *Id.*, at 7:61-64. That legacy dictation system is connected to centralized computer system 390 (depicted in yellow) via the Internet or WWW. *Id.*, at 7:64-8:8. Computer system 390 is then coupled to a centralized transcription service provider 330 (depicted in red) via the Internet or WWW. *Id.* *Obilisetty* teaches the recorded voice captured by the legacy system is transmitted to computer system 390 where it is forwarded through the Internet to a centralized transcription service provider 330. *Id.*, at 8:9-24. The speech is then transcribed and returned to the user via the computer system 390. *Id.*, at 8:41-50.

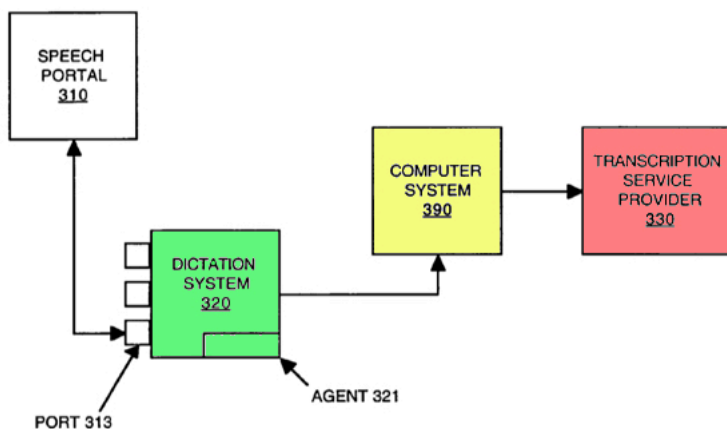


Figure 3B

Id., at Fig. 3B (annotated).

Obilissety is therefore from an analogous field to the '730 Patent and reasonably pertinent to the problem to be solved of facilitating communication and transcription of speech between systems employing heterogenous protocols. *Creusere Dec.* (EX1003) at ¶50.

i. Claim 15

15. A system for facilitating speech recognition and transcription among users, the system comprising:

To the extent that the preamble is limiting, *Skladman* in combination with *Obilissety* teaches a computerized system for facilitating speech recognition and transcription among users. *See, e.g., Skladman* (EX1004), at Abstract (“A system and method for interfacing unified message processing systems with legacy voice mail, e-mail and facsimile systems, located behind corporate firewalls. The system

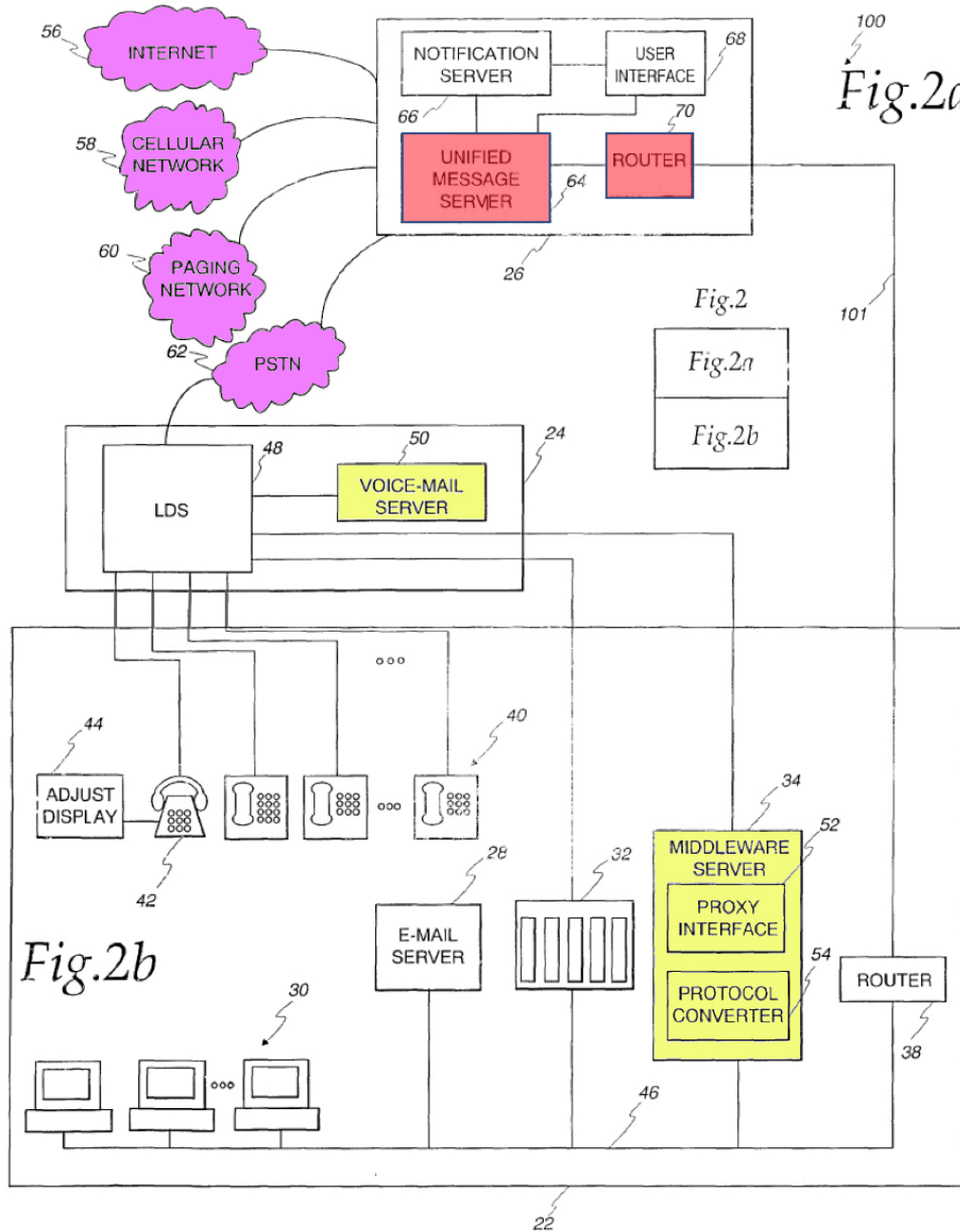
includes a unified message server, a proxy interface and a message protocol convertor. The proxy interface is configured to access the legacy system in response to a request from a unified message server. Messages stored on the legacy system are converted by a protocol convertor to a predetermined format compatible with the unified message server. The converted messages are then transferred to a unified message server which is capable of providing messages from different messaging system, such as voice mail, e-mail and facsimile to users in a predetermined format. The invention permits enterprise wide communication systems to provide unified messaging without abandoning pre-existing legacy messaging system.”); 2:47-3:4; 7:50-56; Figs. 2a and 2b. And *Obilsetty* (as described below) teaches a centralized system for transcription of speech that, when combined with *Skladman*, teaches the preamble’s requirement for a system facilitating speech recognition and transcription among users.

[15(a)(1)] a system transaction manager, the system transaction manager utilizing a uniform system protocol for handling speech information requests and responses to speech information requests,

As discussed above in Section III.C, a “system transaction manager” should be construed to be “[a] server application that provides a central interconnect point (hub) and a communications interface among system components and Users having disparate or heterogeneous protocols and an information router (or bridge

or switch) within the Speech Recognition and Transcription System.” *Supra*,
Section III.C.

As can be seen depicted in red below in Figs. 2a and 2b, *Skladman* includes unified message center 26, that utilizes a unified message server 64 and router 70, to serve as a system transaction manager.



Skladman (EX1004), at Figs. 2a and 2b (annotated).

The unified message center 26 utilizes server 64 and router 70 to serve as the central interconnect point between the legacy voicemail server 50 and the internet 56 access point. It receives and handles speech information requests and responses

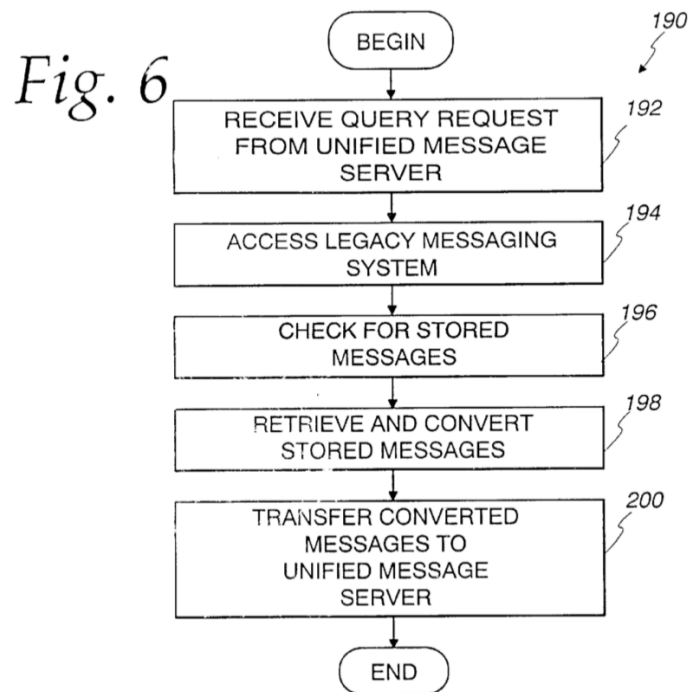
to speech information requests (as explained further below in Section IV.A with regards to limitations [15(a)(2)] and [15(a)(3)]). More specifically, unified message center 64 utilizes its unified message server 64 and router 70 to interface with one user through *Skladman's* unified message server 64 and router 70 serve as a central interconnect point between the unified message center and the legacy systems (including, for example, voice-mail server 50, proxy interface 52, and protocol converter 54) and is collectively operable to handle both speech information requests and responses to speech information requests:

To integrate the legacy systems, the system of the present invention can include a unified message server, a proxy interface, and a message protocol converter. The proxy interface is configured to access the legacy systems **in response to a request from the unified message server**. Messages stored on the legacy system are then converted by a protocol convertor to a predetermined format usable by the unified message server. **The converted messages are then transferred to the unified message server**, which is capable of providing messages from different messaging systems to users in a predetermined format. This arrangement is advantageous in that it permits enterprise-wide communication systems to provide unified messaging without abandoning pre-existing legacy messaging systems.

Skladman (EX1004), at 2:58-3:4 (emphasis added); *see also* 3:15-19; 4:17-20 (“The protocol converter 54 translates the retrieved messages into one or more predetermined formats usable by the unified message [server] 64 and the[n]

transfers the converted messages to the [unified message] server 64.”). Unified message center 64 then interfaces with another user through Internet 56 to provide the message to that user in a different format. *Id.*, at 6:20-59.

Figure 6 of *Skladman* also discloses that the unified message server 64 handles requests and responses:



Skladman (EX1004), Fig. 6; 9:10-32.

The unified message center of *Skladman* also utilizes a uniform system protocol, or “a protocol that processes information expressed in a normalized data format.” *Supra*, Section III.C. A specific example protocol used in *Skladman* is TCP/IP, which is expressly called out as an example of a “unified system protocol” in the ’730 Patent. *Id.* at 7:64-8:6, 8:27-32; ’730 Patent (EX1001) at 5:60-6:3;

Skladman (EX1004), at Abstract (“Messages Stored on the legacy system are converted by a protocol convertor to a predetermined format compatible with the unified message server”); 2:63-3:4 (same). As noted above, the purpose of *Skladman* is to convert messages from disparate messaging systems (or legacy messaging systems) into a predetermined format that is usable by the unified message system. Thus, *Skladman* is able to interface among systems having disparate or heterogeneous protocols as required by the claim construction.

Therefore, *Skladman* teaches a system transaction manager within the appropriately understood construction of that term.

[15(a)(2)] the speech information requests ... comprising ... formatted spoken text and;

The parties have agreed that a “speech information request” is “formatted speech, which can be acted upon by system components, including the System Transaction Manager.” *Supra*, Section III.C. *Skladman* teaches that a speech information request is formed when a user creates a voicemail on the system requesting that it be delivered to another user.

Specifically, *Skladman* teaches that “legacy messaging systems can include voice mail and e-mail servers using industry standard or proprietary message protocols and formats.” *Skladman* (EX1004) at 2:52-54. The voicemail system of *Skladman* is “a commercially-available voice mail platform, such as the Octel Messaging Server from Lucent Technologies, Inc. or Meridian mail from Nortel,

Inc.” *Id.*, at 3:38-41. Further, as noted in Fig. 1, the voicemail system of *Skladman* resides within the telephone company’s central office, but the “voice mail server 50 can be provided by privately owned equipment residing within the enterprise system 22, such as a privately owned voice mail server connected to a private branch exchange (PBX).” *Id.*, at 3:42-47. *Skladman* teaches that the middleware server (including proxy interface 52 and protocol converter 54) works in conjunction with the voicemail server to format the voicemail file by converting it into a digitized audio file, route that formatted voicemail message (e.g., the spoken text), and deliver the message to the intended user/recipient, which collectively constitutes the speech information request. *Id.*, at 5:5-6 (describing the formatting of the spoken text as “[t]he proxy interface 52 then records and digitizes the messages as they are played back.”); 4:37-41 (describing the routing information contained within the voicemail information as “[t]he envelop[e] information can include data such as the identification of the intended recipient(s), the identification of the sender, the date and time the message was received, or information of the same nature.”). The voicemail message in *Skladman* also is formatted in a manner that provides an audio file to the user for replaying through the legacy voicemail system. *Skladman*, (EX1004) at 5:5-11. A PHOSITA at the time of the invention would have also recognized that the message on the voicemail system would be formatted speech intended to be delivered to the user

and played as an audio file by the voicemail system. *Creusere Dec.* (EX1003) at ¶54.

Finally, *Skladman* teaches that the voicemail message is “to be acted upon by the system components, including the System Transaction Manager.” *Supra*, Section III.C. As noted above, the system in *Skladman* is intended to obtain the messages on the legacy system and “act on them” by obtaining them from the legacy system, converting them into the appropriate predetermined (e.g., normalized/unified) format through the protocol converter 54, and then delivering them to the unified message center 26, where they can be used by the unified message server 64 (which is a part of the unified message center 26, which is mapped as the “system transaction manager”). *Skladman* (EX1004) at 2:49-3:4; 6:21-24 (“the unified message server 64 can be a PC server configured to receive various types of messages from disparate messaging systems. The unified message server 64 permits users to receive their messages by way of a single interface.”); 6:35-51; 9:10-27; Fig. 3a, 3b, 6. Thus, *Skladman* includes a speech information request as appropriately understood by the construction of that term. *Supra*, Section III.C.

[15(a)(3)] the speech information ... responses comprising ... formatted transcriptions of the formatted spoken text;

The parties have agreed that “response(s) to [the] speech information request(s)” should be construed as “[a] formatted transcription of formatted

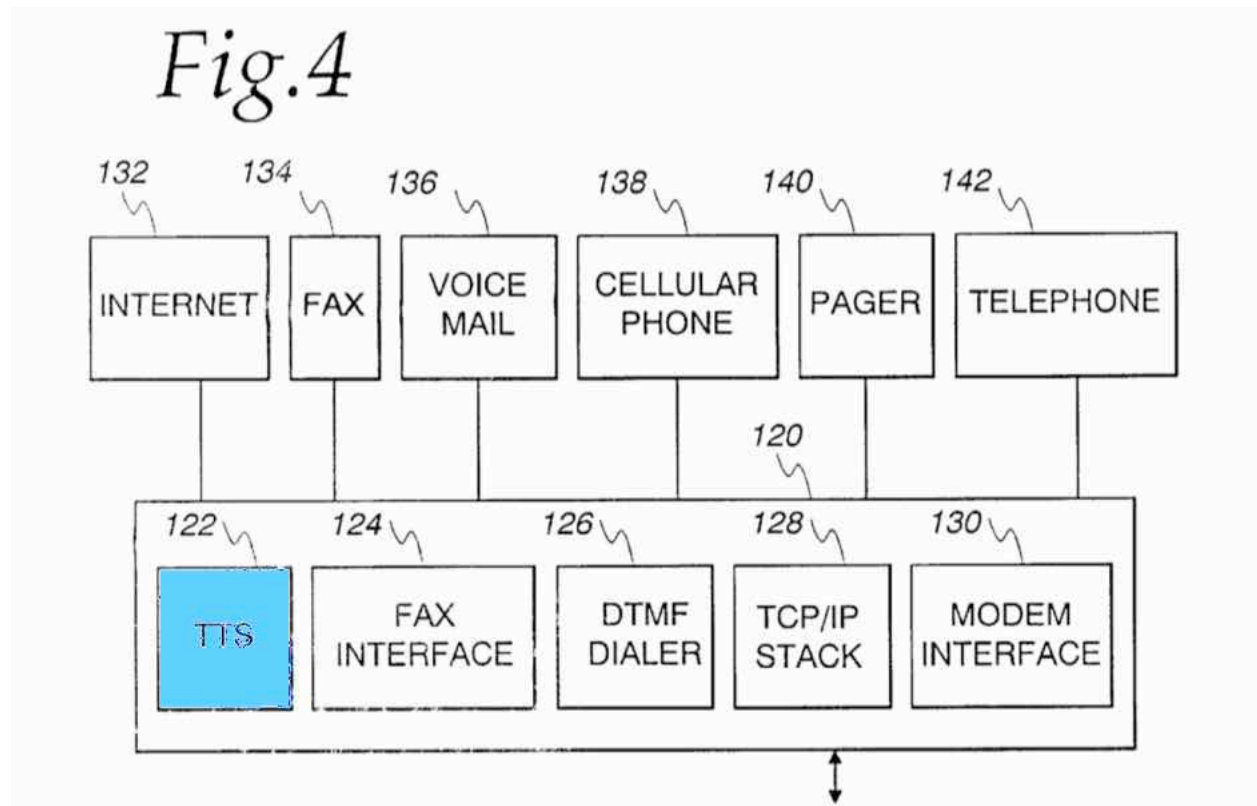
speech.” *Supra*, Section III.C. As noted above, *Skladman* teaches that a legacy voicemail message will be formatted and routed to the unified message server 64 by the voicemail server 50 working in conjunction with the middleware server, including proxy interface 52 and protocol converter 54. First, the voicemail is collected from the voicemail server 50 by the proxy interface 52 (at the request of the unified message server 64). *Skladman* (EX1004) at 4:14-44. Once the proxy interface has connected to the voicemail server 50, the proxy interface 52 will obtain the voicemail message and then the voicemail will be “convert[ed] [from] audio playback to digitized voice mail messages.” *Skladman* (EX1004) at 5:8-10. This digitized voicemail message will then be passed by the proxy interface 52 to the protocol converter 54 where its protocol (or “format”) will be converted from the legacy format into the unified format used by the *Skladman* unified message server 64:

The digitized voice-mail messages are then converted and compressed by the protocol converter 54 using standard digital audio compression formats such as .wav or 0.32 k adpcm.

Id. at 5:11-14. These converted and re-formatted messages are then provided to the unified message server 64 for delivery to the user, who is the intended recipient of the voicemail. *Id.* at 5:13-15; 2:65-3:1; 6:22-34; 6:48-59.

Skladman also discloses a similar process for emails stored on the user’s legacy email server. *Id.* at 5:18-52. After the format of the legacy emails is

converted by the protocol converter 54, the email is provided to the unified message server 64 where the user/recipient of the email may then receive the email from the *Skladman* system through an email account, cellular phone, pager, or conventional telephone. *Id.* at 5:18-52; 6:21-34. In a situation where the converted email message is being provided via telephone, *Skladman's* unified message server 64 “can be interfaced to a text-to-speech (TTS), generator to convert the e-mail message to audio messages that can be heard over the phone.” *Id.* at 6:52-59; Fig. 4; 7:57-8:26. The TTS that interfaces with unified message server 64 is depicted in blue in Fig. 4 below:



Id. at Fig. 4 (annotated).

Despite teaching that legacy emails can be converted and formatted to be compatible with the unified message server and then further converted to speech, *Skladman* does not expressly teach the inverse scenario in which a voicemail message is converted and formatted to a transcription of the formatted speech. However, this functionality was well known at the time. *Creusere Dec.* (EX1003) at ¶¶35-47, 54; *see also* '730 Patent (EX1001) at 2:18-35 (noting “[t]here are several speech recognition systems currently on the market that can operate on a desktop computer” and identifying Dragon Dictate and IBM ViaVoice). Another example is *Obilisetty*, which teaches a method and system for capturing and transcribing dictated information and for delivering the transcribed information to an end user. *Obilisetty* (EX1005) at Abstract. Specifically, *Obilisetty* teaches:

A voice file containing a digital recording of the dictated information is received via the Internet from a remote device. The voice file is forwarded via the Internet to a remote transcription service provider. The transcribed information is received from the transcription service provider and in turn delivered to, for example, a facsimile machine, a text server and/or printer, a computer system or Web browser.

Id. *Obilisetty* also teaches that it can obtain voice files from a different location (such as the Internet), and “send voice files and transcript files (that is, a file containing the transcribed version of a voice file)” to a user. *Id.* at 6:58-64, Fig. 3A.

Obilisetty teaches a similar arrangement of network components to that in *Skladman* to accomplish its transcription. Namely, the voice is recorded on a legacy system and then transcribed through a centralized server system on the Internet:

In accordance with the present embodiment of the present invention, recorded information is digitized (e.g., analog-to-digital conversion) by ICS 315. The resultant voice file is forwarded to computer system 390 via the Internet.

Id., at 7:39-42.

Computer system 390 then sends the voice file to transcription service provider 330, either automatically or in response to a request from transcription service provider 330.

Id., at 7:46-48.

Referring still to FIG. 4, transcription service provider 330 returns to computer system 390 a transcript file providing a transcribed version of the voice file.

Id., at 8:41-43.

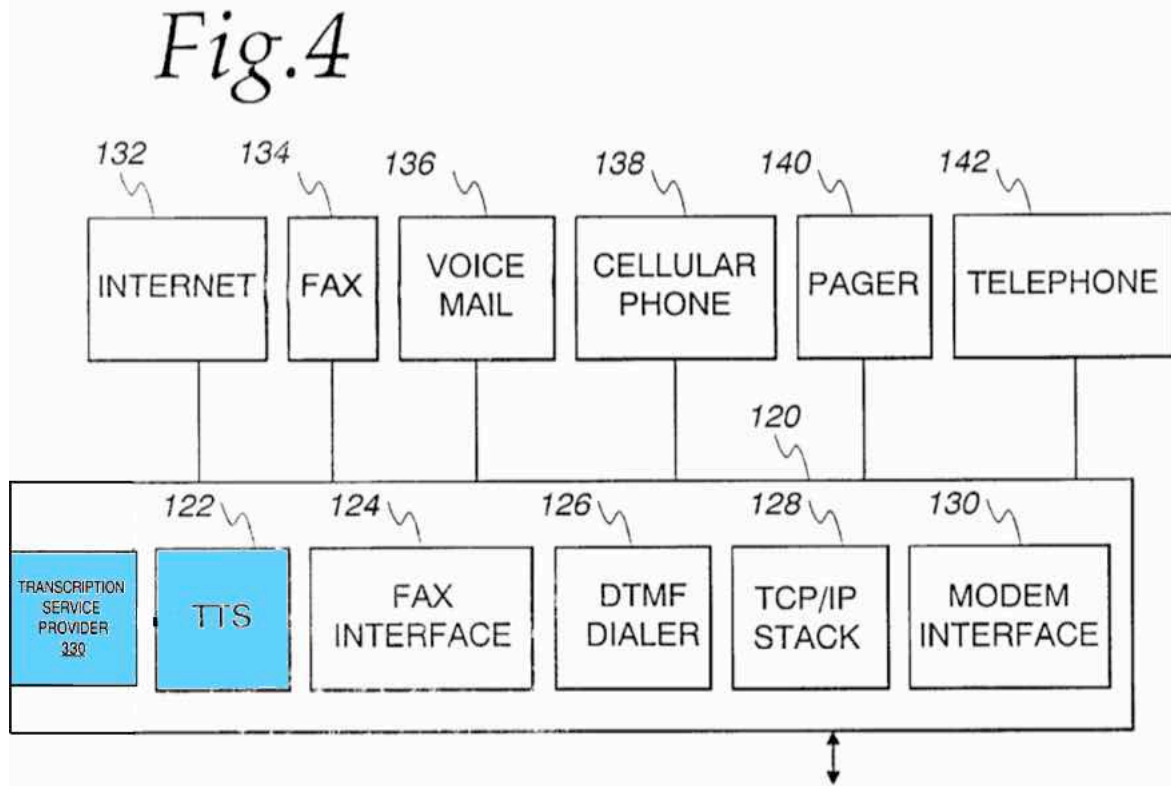
The transcribed information can be provided by computer system 390 to the end user in a variety of different ways.

Id., at 8:49-50; *see also* 8:50-62 (describing the use of facsimile, text, printer, or Web browser to deliver the transcribed information); Fig. 3A, 3:19-50, 6:34-36, 8:25-40. *Obilisetty* teaches that the transcription service provider 330 “is a computer system or other such device” and, therefore, the transcription is done by a system on a computer as required by the claim construction. *Id.*, at 6:58-64.

Obilissety further recognizes that its transcription system “is well-suited for use with any number of speech portals, intelligent communication servers operated by different CLECs, computer systems, Web servers, and transcription service providers.” *Id.* at 6:15-21; 8:43-48 (“In various embodiments, the transcript files can be formatted according to conventional word processing applications (such as Microsoft Word), XML (Extensible Markup Language), or HL7. Thus, *Obilissety* teaches a formatted transcription of formatted speech. *Creusere Dec.* (EX1003) at ¶54.

Based on the teachings of *Obilissety*, it would have been obvious to a PHOSITA to enable the system of *Składman* to transcribe the voicemail messages into a formatted transcription of the formatted speech contained therein. *Creusere Dec.* (EX1003) at ¶¶55-57. A PHOSITA would have been motivated to enable the system of *Składman* to transcribe voicemail messages in order to accommodate different methods of providing the message to the user in a manner similar to the TTS system 122 depicted in Fig. 4. *Id.* A PHOSITA would have recognized that this would have been both feasible and desirable because the system of *Składman* already provided users text-based message access (e.g., email) and audio message (e.g., voicemail) access, and performs the inverse scenario of providing audio-based access to text-based messages. *Id.* It also would have been desirable for a user to receive their voicemails via email without accessing the voicemail system.

Id. This is because reading a transcript of a voicemail in an email has all sorts of advantages over listening to a voicemail. *Id.* For example, it can be done discretely, it can be done in a noisy environment, it avoids the problem of privacy concerns with someone listening to the voicemail, and reading the transcript will be far quicker than dialing in to the voicemail server and listening to a lengthy message. *Id.* A PHOSITA would recognize that users would have both desired and expected speech-to-text functionality (e.g., receiving voicemails as email) is an obvious extension of the current teachings that would be easily supported with a minor addition to the existing *Skladman* infrastructure. *Id.* This modification would have predictably resulted in *Skladman's* unified message server 64 being able to interface with *Obilisetty's* transcription system to provide the unified message server 64 with a formatted transcription of the formatted voicemail message. *Id.* In this scenario, A PHOSITA would have understood that one possible combination would have resulted in the combination seen in modified Fig. 4 as follows (combined “transcription service provider 330” from *Obilisetty* shown next to the TTS of *Skladman*, both noted in blue below):



Skladman (EX1004) at Fig. 4 (modified to include Transcription Service Provider 330 from *Obilissety*); *Creusere Dec.* (EX1003) at ¶55-57.

Further, because the system of *Obilissety* was designed to interface a legacy system with a centralized system located on the Internet, it would have been obvious to combine the *Obilissety* transcription system with the *Skladman* system. *Creusere Dec.* (EX1003) at ¶57. This would have predictably enabled the user of the *Skladman* system to receive formatted transcriptions of their voicemail messages through the unified message server 64. *Id.* Moreover, such a combination would not have rendered the *Skladman* system inoperable or changed the principle operation of the system. *Id.* Indeed, because the *Skladman* system already

contemplated a text-to-speech conversion, speech-to-text would have been a natural extension of the capabilities already provided by *Skladman. Id.*

Therefore, when *Skladman's* system is combined with the transcription capabilities of *Obilisetty*, the voicemail message (which is formatted speech) would have been converted into a formatted transcription of the formatted speech, thus constituting the response to the speech information request. *Creusere Dec.* (EX1003) at ¶¶54-57.

[15(b)] a first user application service adapter communicating with at least one user and the system transaction manager, the first user application service adapter configured to generate speech information requests from spoken text produced by the at least one of the users through a first protocol;

As discussed above in Section III.C, an “application service adapter” is “an application layer within the Speech Recognition and Transcription System that provides an interface among Users, Speech Recognition and Transcription Engines, the System Transaction Manager, and other System components by allowing a User’s existing application and/or a System components application to communicate with the Transaction Manager,” while a “user application service adapter” should be construed to be “a Specific Application Service Adapter that handles formatting and Routing of Speech Information Requests and Responses to elements of a User’s Protocol within the Speech Recognition and Transcription System.” *Supra*, Section III.C.

As the construction makes clear, the primary purpose of a user application service adapter (“UASA”) is to handle formatting and routing of speech information requests and responses within the overall speech recognition and transcription system. The location of the UASA does not matter—the ’730 Patent explains that the location of the User Application Service Adapter may “reside on the Users’ 22 Workstation/workgroup computer system.” ’730 Patent (EX1001), at 9:51-53, 13:52-55, 16:27-38.

Additionally, the claim limitation makes clear that the first UASA must generate the speech information request from spoken text produced by the at least one of the users through a first protocol.

Skladman teaches that a user will contact the voicemail server 50 and leave a recorded message that consists of that user’s spoken text. *Skladman* (EX1004), at 4:33-47; 4:60-63; 5:5-17. The voicemail server 50 will then create a recording of the user’s message in a first audio format that is intended to be provided to the other user (the recipient of the voicemail message). *Id.* *Skladman* teaches that this voicemail message, upon request of unified message server 64, will be additionally formatted for and routed to the unified message server 64 by the middleware server, including proxy interface 52 and protocol converter 54. First, the voicemail is collected from the voicemail server 50 by the proxy interface 52 (at the request of the unified message server 64). *Skladman* (EX1004) at 4:14-44. Once the proxy

interface has connected to the voicemail server 50, the proxy interface 52 utilizes a standard analog/digital converter to obtain the voicemail message from the voicemail server where it is “convert[ed] [from] audio playback to digitized voice mail messages.” *Skladman* (EX1004) at 5:8-10. This digitized voicemail message will then be passed by the proxy interface 52 to the protocol converter 54 where the protocol converter 54 converts the protocol (or “format”) from the legacy format into the unified format used by the *Skladman* unified message server 64:

The digitized voice messages are then converted and compressed by the protocol converter 54 using standard digital audio compression formats such as .wav or 0.32 k adpcm.

Id. at 5:11-14. These converted and re-formatted messages are then provided to the unified message server 64 for delivery to the user, who is the intended recipient of the voicemail. *Id.* at 5:13-15; 2:65-3:1; 6:22-34; 6:48-59. In this manner, the voicemail server 50 in conjunction with the middleware server 34, including proxy interface 52 and protocol converter 54, generates the Speech Information Request.

Skladman also teaches the voicemail server 50 and middleware server 34 constitute an application layer within the voicemail system that provides an interface between a user who is leaving a message and the user who will receive the message. *Skladman* (EX1004) at 4:33-44 (describing that the voicemail message will include “identification of the sender”); *Creusere Dec.* (EX1003) at ¶58 (concluding that Skladman’s voicemail server 50 and middleware server 34

constitute an “application layer” as that term is used in the proposed construction because they serve as the interface between the user and the central unified server).

The voicemail server 50 in conjunction with the middleware server 34 in *Skladman* will use the routing information from the voicemail to ultimately route the message to the unified message center 26. For example, *Skladman* teaches the “envelope information” is used for the routing to the intended recipient:

In response to a voice mail access request from the unified message server 64, the proxy interface 52 connects to the voice mail server 50 to check “envelope information” for voice mail messages stored at the server 50. The envelop information can include data such as the identification of the intended recipient(s), the identification of the sender, the date and time the message was received, or information of the same nature. To access the envelop information, the middleware server 34 can use a standard communication connection, such as a X.25 connection, to communicate with the voice mail server 50 by way of the LDS 48.

Skladman (EX1004), at 4:33-44.

As noted above, the proxy interface 52 provides the voicemail to the protocol converter 54 within middleware server 34 where it is reformatted into a different, uniform system protocol compatible with unified message server 64:

FIG. 6 is a flow diagram 190 illustrating the operation of the middleware server 34. In step 192, the middleware server 34 receives an access request from the unified message server 64. In response to

this request, the proxy interface 52 of the middleware server 34 accesses a specified legacy messaging system 28,50 (step 194). As described earlier in connection with FIG. 1, the proxy interface 52 can first query the legacy message server to determine the presence of messages. If there are messages stored on the legacy message server, the proxy interface 52 then performs a login emulation to access the stored messages. Upon successfully logging-in to the message server, the proxy interface 52 retrieves stored messages and provides them to the protocol converter 54 (step 198). **The protocol converter 54 then converts the messages to a standard digital format usable by the unified message server 64.** After converting the messages, protocol converter 54 transfers the converted messages to the unified message server 64 (step 200).

Id., at 9:10-27 (emphasis added). Through this formatting, *Skladman* teaches that its voicemail message is converted from a first, native protocol into a different protocol used by the unified message center. *Id.* at 5:5-33.

Therefore, the voicemail server 50 and middleware server 34 of *Skladman* constitutes the first user application service adapter as that phrase is properly construed.

[15(c)] a speech recognition and transcription engine communicating with the system transaction manager, the speech recognition and transcription engine configured to receive speech information requests from the system transaction manager, to generate responses to the speech information requests, and to transmit the responses to the system transaction manager; and

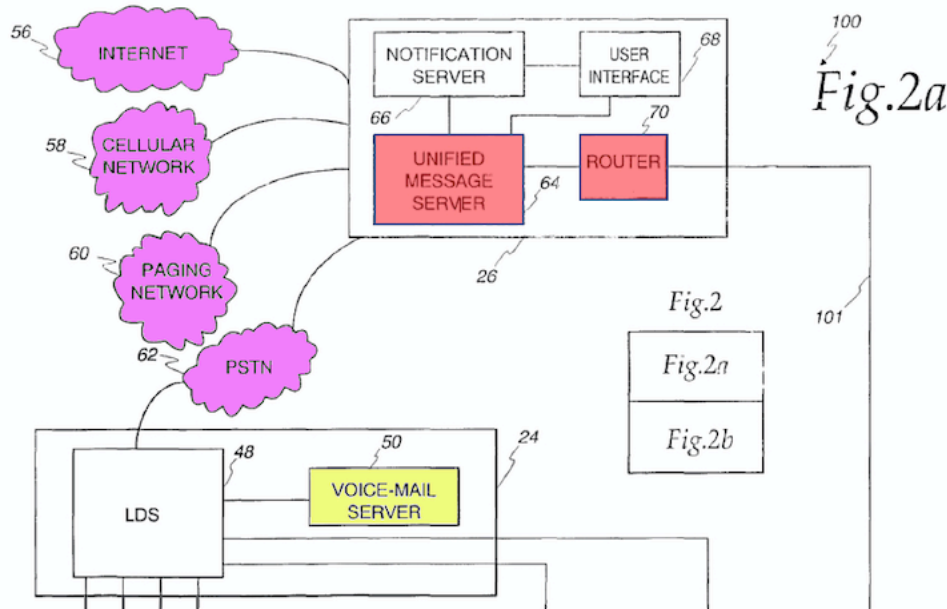
A “speech recognition and transcription engine” is “[a] process running on a computer that recognizes an audio file and transcribes that file to written text to generate a transcription of Speech.” *Supra*, Section III.C.

As discussed in the prior section, *Skladman*'s unified message server 64 receives a formatted voice message from the voicemail server 50 working in conjunction with middleware server 34. As discussed above with regards to limitation [15(a)(3)], *Skladman* discloses a text-to-speech synthesizer 122 that interfaces with the unified message server 64 (e.g., the system transaction manager) to convert text (such as emails) to speech. While *Skladman* does not disclose the inverse system that will transcribe an audio file containing speech into written text, *Obilisetty* does disclose such a system and, according the combination proposed by Petitioner, the *Obilisetty* transcription system would interface with the unified message server 64 to transcribe the audio file provided by the protocol converter 54 containing the voicemail message into written text to generate a transcription. *See* Section IV.A at limitation [15(a)(3)]. The result of this combination would be to generate a response to the speech information request by creating a formatted transcription of the formatted speech contained in the voicemail message. *Creusere Dec.* (EX1003) at ¶56. *Skladman* teaches that this process would occur through communications with the unified message server 64, where it could then be provided to the user. *Skladman* (EX1004), at Fig. 6; 6:52-

59; 7:64-8:32. Thus, the transcription system of *Obilisetty*, working in communication with the *Skladman*'s unified message server 64 would constitute the speech recognition and transcription engine.

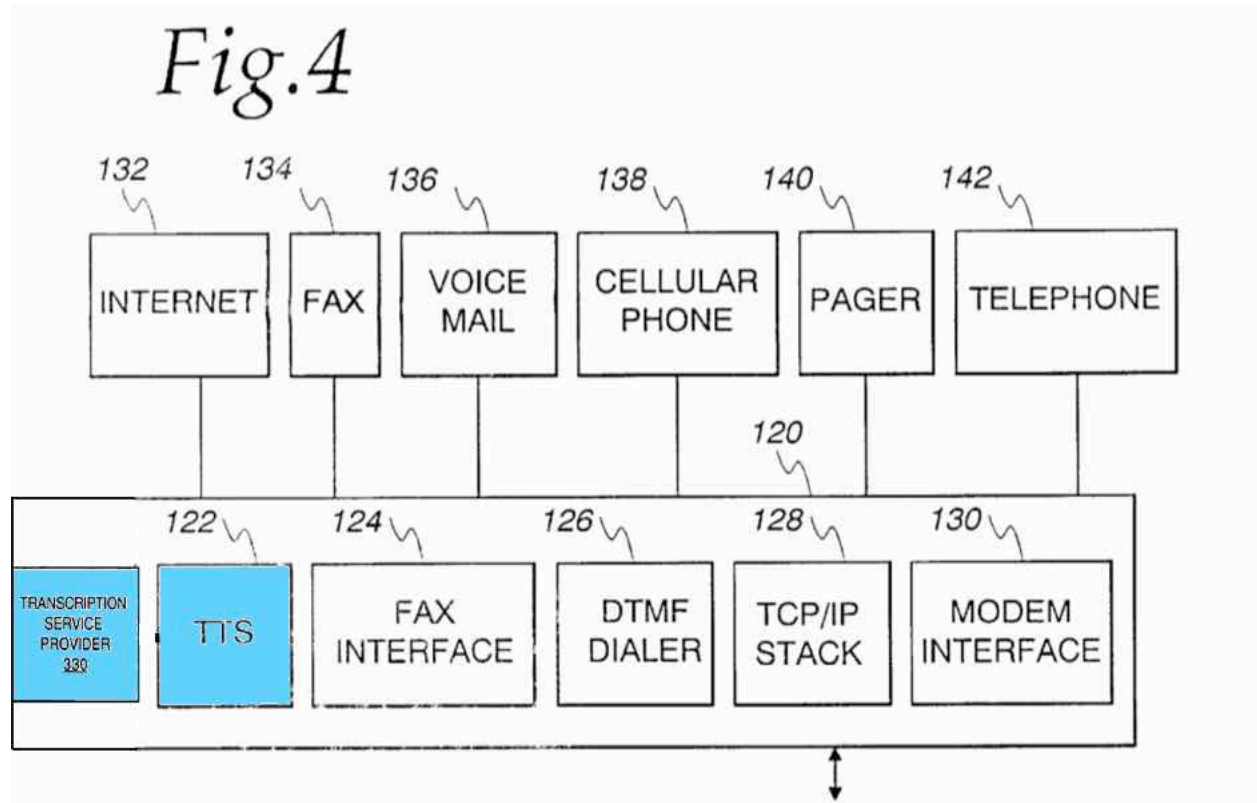
[15(d)] a second user application service adapter communicating with one or more of the users and with the system transaction manager, the second user application service adapter which is different than the first user application service adapter and configured to provide the one or more users with a transcription of the spoken text that is compatible with a second protocol, the second protocol being the same as or different than the first protocol.

Skladman teaches that once the voicemail (or email) from the legacy system has been converted to a normalized, unified format, that voicemail (or email) may then be accessed by a user. *Skladman* (EX1004) at 6:22-23 (“The unified message server 64 permits users to receive their messages by way of a single interface.”). *Skladman* teaches that a user may receive their messages through internet 56, cellular network 58, paging network 60, or conventional telephone 62 (as depicted in purple in Fig. 2a below):



Id., at Fig. 2a (annotated); 6:24-34. For example, a user may utilize internet 56 to obtain messages from the unified message center 26 (and unified message server 64, in particular) using “standard e-mail software on a PC or a web page” where the voicemails from the legacy system are provided to users “as .wav e-mail attachments that can be audibly played back at a user’s PC if the PC includes standard audio .wav decompression software.” *Id.*, at 6:35-51. Under Petitioner’s proposed combination with *Obilissetty*, a PHOSITA would have understood that the formatted transcription of the formatted speech from the voicemail would be provided via email to the user using a user-facing email application that properly formats and presents the transcribed text in a viewable form. *Creusere Dec.* (EX1003) at ¶59.

Skladman teaches an interface 120 for connecting the user to the unified message server 64 (depicted below in Fig. 4 showing the combined system with the transcription service provider from *Obilissety* and the *TTS* from *Skladman* depicted in blue):



Skladman (EX1004) at Fig. 4; 7:57-59 (depicting the modified Fig. 4 from Petitioner’s combination). As noted above, modified Fig. 4 also includes the TTS generator and the transcription service provider from *Obilissety*. These components are utilized for providing either text-to-speech capabilities or speech transcription capabilities to the Unified Message Server. The other interfaces in block 120—fax interface 124, DTMF Dialer 126, TCP/IP Stack 128, and Modem Interface 130—

are utilized by the unified message server for communicating with a user through one of the networks identified in blocks 132-142. *Id.*, at 7:57-8:32. For example, the interface 120 includes a conventional Fax Interface 124, TCP/IP interface 128, and a conventional MODEM interface 130 that “permit the unified server 64 to communicate with each of the various communication networks” such as the internet for sending messages (such as email or transcribed voicemail) to the user. *Id.*, at 7:64-8:26 (noting, for example, that “[t]he fax interface 124 can permit the unified server 64 to transfer messages to subscribers by way of the fax,” and that “the stack 128 permits the server 64 to communicate messages to subscribers over data networks using the TCP/IP protocol, such as the Internet 132.”). “Each interface component can be implemented using commercially available PC peripheral devices configured to communicate with the server 64 using standard APIs in the Windows OS.” *Id.* at 8:3-6; 8:15-18; 8:27-32.

A PHOSITA would have understood that one of these interfaces at 124-130 would be used to connect the user to the unified message server 64 to provide the user with the transcribed, formatted speech. *Creusere Dec.* (EX1003) at ¶60. *Skladman* also teaches that the messages provided to the user are a second, different protocol than the original message. *Skladman* (EX1004) at 2:58-3:4. A PHOSITA would recognize they must be a different protocol to be communicated through one of the interfaces at 124-130. *Creusere Dec.* (EX1003) at ¶60. As

discussed above, the original voicemail message is a different protocol and format than the final transcribed message provided to the user by the unified message server 64. *See* Section IV.A at limitation 15(b).

Thus, the interfaces at 124-130 depicted in *Skladman* Fig. 4 handle routing of the formatted and translated message from the unified message center system of *Skladman* to the user and would constitute a second user application service adapter as construed by the parties. *Supra*, Section III.C.

ii. Claim 17

17. A method of exchanging transcribed spoken text among users, the method comprising:

To the extent that the preamble is limiting, *Skladman* in view of *Obilisetty* teaches this limitation. *See supra* Section IV.A at limitation 15. Additionally, as noted above, the *Skladman* plus *Obilisetty* combination would result in the first user's voicemail containing speech being transcribed into text and exchanged with another user. *See Supra* Section IV.A, at limitation 15(a)(3) and 15(d). Thus, *Skladman* in view of *Obilisetty* also renders obvious a method of exchanging transcribed spoken text among users.

[17(a)] generating a speech information request from spoken text obtained through a first protocol, the speech information request comprised of formatted spoken text generated using a first user application service adapter;

Skladman teaches this limitation. *See supra* Section IV.A at limitation 15(a)-15(b). *Id.*

[17(b)] transmitting the speech information request to a speech recognition and transcription engine via a system transaction manager;

Skladman teaches this limitation. *See supra* Section IV.A at limitation 15(a)-

(c).

[17(c)] generating a response to the speech information request using the speech recognition and transcription engine, the response comprised of a formatted transcription of the formatted spoken text;

Skladman in view of *Obilisetty* teaches this limitation. *See supra* Section IV.A at limitations 15(a), (c).

[17(d)] transmitting the response to a user via the system transaction manager; and

Skladman teaches this limitation. *See supra* Section IV.A at limitation 15(a),

(d).

[17(e)] providing the user with a processed transcription of the spoken text using a second user application service adapter, the processed transcription being compatible with a second protocol that is different than the first protocol.

Skladman in view of *Obilisetty* teaches this limitation. *See supra* Section IV.A at limitation 15(a), (d). Specifically, *Skladman* expressly teaches that the protocols used by the user receiving the transcribed voicemail may be distinct from protocols used by the legacy voicemail system on the transmitting side.

B. Ground 2: Skladman in View of Obilisetty in Further View of Schrage Renders Claims 15 and 17 Obvious

For the reasons discussed above, *Skladman* in view of *Obilisetty* renders claims 15 and 17 obvious. However, to the extent it is argued that *Obilisetty* does

not disclose a speech recognition and transcription engine that creates responses to speech information requests because it is not utilizing a process running on a computer to transcribe an audio file to written text, it would have been obvious to include the computer based transcription from *Schrage*.

Schrage was filed October 23, 1998, and claims priority to a provisional application filed on October 28, 1997. *Schrage* (EX1022). *Schrage* qualifies as prior art under 35 U.S.C. § 102(e) (pre-AIA). *Schrage* was not cited or considered during the original prosecution of the '730 Patent. *Supra*, Section II.B. Like the '730 Patent, *Schrage* generally teaches a centralized transcription server that includes an automated speech-to-text system running on a computer that recognizes the speech and provides a transcription of the audio in the form of written text:

In response to the transcription server ready signal, the requester begins providing speech input which is transmitted from communication unit U2 to server line L1 over communication channel C2. The transcript server 12 generates and stores, for each communication channel in use, a separate recording of any received speech, e.g., a digital audio recording.

Id., at 6:50-59.

Audio recordings produced by transcript server 12 are provided, in step 418, to transcription system 13. The transcription system 13 is responsible for performing, in step 420, an **automated speech recognition operation on the recorded speech. The transcription system 13 generates, in step 422, a transcript, e.g., a set of text corresponding to recognized speech in the recording**, for each recording received from the transcript server 12.

Id., at 6:65-7:5 (emphasis added); *see also*, Abstract; 5:5-54; 6:9-7:5.

Similar to the '730 Patent, Figure 2 of *Schrage* further notes that its transcription system would be provided on a centrally located server rather than on a user's individual computer. *Id.*, at Fig. 2; 6:50-7:5. Thus, *Schrage* is from an analogous field to the '730 Patent and reasonably pertinent to the problem to be solved of facilitating communication and transcription of speech between systems.

i. Claim 15

15. A system for facilitating speech recognition and transcription among users, the system comprising:

Petitioner incorporates its discussion of the Preamble of Claim 15 from Ground 1. *See, Supra*, Section IV.A. Additionally, *Skladman*, when combined with *Obilisetty* and *Schrage*, as discussed below, meets the preamble because *Schrage*

discloses a computerized system for automatic speech recognition and transcription. *Schrage* (EX1022), at 6:65-7:5.

[15(a)(1)] a system transaction manager ..., [AND]

[15(a)(2)] the speech information requests...;

Petitioner incorporates its discussion of Limitation 15(a)(1) and 15(a)(2) from Ground 1. *See, supra*, Section IV.A.

[15(a)(3)] the speech information ... responses comprising;

Petitioner incorporates its discussion of Limitation 15(a)(3) from Ground 1. *See, supra*, Section IV.A. In addition, Petitioner notes it would have been obvious to modify the speech recognition and transcription system taught by *Skladman* in view of *Obilisetty* to also include the computerized speech recognition and transcription system of *Schrage*, which would utilize a process running on a computer to generate the response to the speech information request. By the date of the '730 Patent, a PHOSITA would have known that computerized transcription systems were well known and were widely deployed as means to minimize network administration time and cost. *Creusere Dec.* (EX1003) at ¶¶64-65. Additionally, the '730 Patent itself recognizes that computerized transcription systems were well known and commercially available. '730 Patent (EX1001) at 2:18-35. Thus, it would have been obvious to a PHOSITA to utilize the centralized computerized transcription system of *Schrage* instead of a user-based transcription

system of *Obilisetty*. *Creusere Dec.* (EX1003) at ¶¶64-65. A PHOSITA also would have recognized that *Schrage*'s transcription server was located on a central server, which fits with the centrally located system of *Skladman* and that this centralized system would have been preferable for numerous reasons. *Id.* Moreover, a PHOSITA would have been motivated to modify the transcription system to be computerized rather than rely on human translators because a computerized transcription system is faster than human transcription and more efficient to manage and maintain. *Id.* A PHOSITA would have a reasonable expectation of success with such a modification. *Id.* Thus, the modification of *Schrage*'s computerized transcription system to *Obilisetty*'s speech recognition and transcription engine would have provided the responses to the speech information requests.

[15(b)] a first user application service adapter ...;

Petitioner incorporates its discussion of Limitation 15(b) from Ground 1. *See, supra*, Section IV.A.

[15(c)] a speech recognition and transcription engine ...; and

As noted above in Section IV.B. for Ground 2, Limitation 15(a)(3), a PHOSITA would have been motivated to modify *Skladman* in view of *Obilisetty* with the computerized and automated transcription system of *Schrage*. *Creusere Dec.* (EX1003) at ¶¶64-65. Thus, the transcription system of *Obilisetty* as modified

with the computerized transcription system of *Schrage* would have created the claimed speech recognition and transcription engine.

[15(d)] a second user application service adapter...

Petitioner incorporates its discussion of Limitation 15(d) from Ground 1. *See, supra*, Section IV.A.

ii. Claim 17

17. A method of exchanging transcribed spoken text among users, the method comprising:

Petitioner incorporates its discussion of the preamble from Claim 15, ground 2. *See, supra*, Section IV.B.

[17(a)] generating a speech information request...;

Skladman teaches this limitation. *See supra* Section IV.B at limitation 15(a)-15(b). *Id.*

[17(b)] transmitting the speech information request...;

Skladman teaches this limitation. *See supra* Section IV.B at limitation 15(a)-(c).

[17(c)] generating a response to the speech information request ...;

Skladman in view of *Obilisetty* in further view of *Schrage* teaches this limitation. *See supra* Section IV.B at limitations 15(a), (c).

[17(d)] transmitting the response...; and

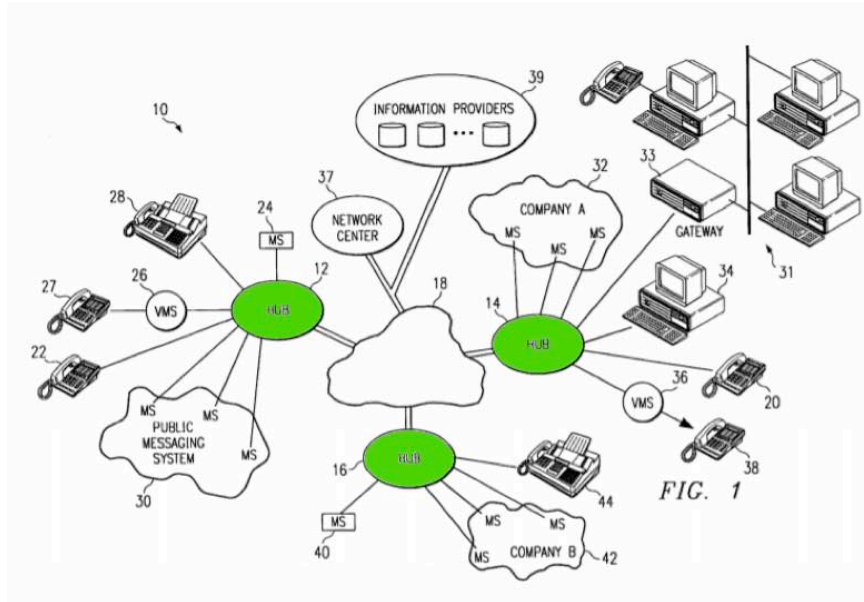
Skladman teaches this limitation. *See supra* Section IV.B at limitation 15(a),
(d).

[17(e)] providing the user with a processed transcription of the spoken text....

Skladman in view of *Obilisetty* in further view of *Schrage* teaches this limitation. *See supra* Section IV.A at limitation 15(a), (d).

C. Ground 3: *Cohn* in view of *Obilisetty* renders claims 15 and 17 obvious

Cohn was issued May 16, 2000 and therefore qualifies as prior art under at least 35 U.S.C. § 102(b) (pre-AIA). *Cohn* (EX1006). Like the '730 Patent, *Cohn* generally teaches a network communications system for enabling communication between messaging systems having disparate capabilities and protocols. *Id.* at Abstract, 1:19-23, 2:54-63. As illustrated and annotated in green in Figure 1 below, *Cohn* provides network hubs 12,14,16 that are operable to “translate messages into various forms of message media” including translation of speech to text.



Id. at 22:66-23:8, Fig. 1 (annotated). *Cohn* is therefore from an analogous field to the '730 Patent and reasonably pertinent to the problem to be solved of facilitating communication between systems employing heterogeneous protocols. *See, e.g.* '730 Patent (EX1001) at 3:56-62 (“The present invention provides a system for facilitating speech recognition and transcription among users employing heterogeneous or disparate entity system protocols...”), 7:51-56; *Cohn* (EX1006) at 35:49-36:2 (“In summary, an integrated communications system is provided that allows for the addressing, transport, and management of multi-media messages between various messaging facilities with disparate capabilities, media and protocols...”); *Creusere Dec.* (EX1003) at ¶53.

As discussed above, *Obilisetty* is an analogous reference that discloses systems and methods for “capturing and transcribing dictated information and for

delivering the transcribed information to an end user” over a network, including the Internet. *Obilissety* (EX1005) at 2:45-47. *Supra*, Section IV.A.

i. Claim 15

15. A system for facilitating speech recognition and transcription among users, the system comprising:

To the extent that the preamble is limiting, *Cohn* teaches a computerized system for facilitating speech recognition and transcription among users. *See, e.g., Cohn* (EX1006) at Abstract, 1:19-23, 2:54-63 (“...a network-based voice and multimedia communications system is provided that comprises a plurality of network hubs operable to communicate with each other and with local messaging systems. The network hubs operate to receive messages in any media using any addressing convention from users either directly or through local messaging systems and to transmit the messages to other users whether located on local messaging systems or otherwise through the plurality of network hubs.”), 22:66-23:8 (“Each network hub also includes the ability to translate messages into various forms of message media.”), 35:49-36:2, Fig. 1.

[15(a)] a system transaction manager, the system transaction manager utilizing a uniform system protocol for handling speech information requests and responses to speech information requests, the speech information requests and responses comprising, respectively, formatted spoken text and formatted transcriptions of the formatted spoken text;

As discussed above in Section III.C, a “system transaction manager” (STM) should be construed to be “[a] server application that provides a central

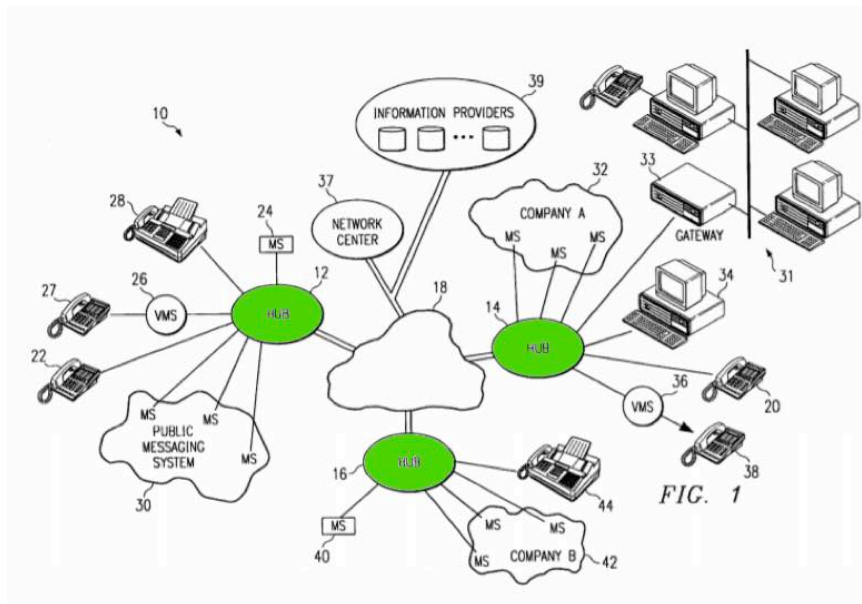
interconnect point (hub) and a communications interface among system components and users having disparate or heterogeneous protocols and an information router (or bridge or switch) within the Speech Recognition and Transcription System.” *Supra*, Section III.C. Further, “uniform system protocol” should be construed as “[a] protocol that processes information expressed in a normalized data format.” Additionally, as mapped below, Claim [15(c)] specifies that the centralized STM communicates with a “speech recognition and transcription engine” (SRTE) that transcribes spoken text—in the patent’s parlance, it “generate(s) responses to the speech information requests.”

Applying the proposed claim constructions, Claim 15 requires a central interconnect point (the STM) that receives formatted spoken text (speech information requests) in a normalized data format and utilizes an SRTE to produce transcriptions of the formatted spoken text. As explained in detail below, *Cohn* discloses a number of interconnecting “network hubs” that receive spoken text (e.g., voice messages), format them pursuant to a data format normalized for communications across the central communications platform, and route those formatted messages to another user’s “network hub.” *Cohn* further discloses that these “network hubs” can translate spoken text into transcriptions that are also formatted pursuant to the normalized data format. Petitioner proposes that, for a number of reasons, it would have been obvious to move the transcription

functionality from the distributed “network hubs” to a centralized processing system such that the network hubs retain their native formatting and routing functionalities, but all transcription is performed centrally. *Creusere Dec.* (EX1003) at ¶¶61, 62.

Cohn’s network hubs format spoken text into transcribed messages and route those messages using a normalized format

As seen below in Figure 1, *Cohn* discloses a number of interconnecting “network hubs” 12,14,16 (annotated in green) that are operable to “translate messages into various forms of message media” including translation of speech to text.



Id. at 22:66-23:8, Fig. 1 (annotated). Each of these hubs individually serve to provide a communications interface among Users utilizing different protocols within the speech recognition and transcription system. For example, *Cohn* teaches

that the hubs “operate to receive messages in any media using any addressing convention from users either directly or through local messaging systems and to transmit the messages to other users whether located on local messaging systems or otherwise through the plurality of network hubs.” *Id.* at 2:54-63. Specifically, *Cohn*’s network hubs utilize a uniform system protocol (referred to by *Cohn* as an “internal protocol”) for routing and processing messages in the centralized portion of the system (i.e., between the hubs):

The communications system of the present invention uses a shared internal protocol for all routing and processing of messages. As such, the communications system of the present invention can easily be adapted as new communications protocols, such as Multipurpose Internet Mail Extension (MIME) or the like, become popular. The network hubs that are required to interface with systems utilizing these new protocols need only translate the new protocol to the internal protocol. As such, the operation of the communications system as a whole can support the addition of unlimited new protocols.

Id. at 10:20-31 (emphasis added), *see also* 22:49-23:21. Additionally, *Cohn* teaches that each of its network hubs include a message router 72 as part of the external interface between the hubs and other network components. *Id.* at 11:1-9, 12:20-44, *See also* Abstract, Figs. 3-4.

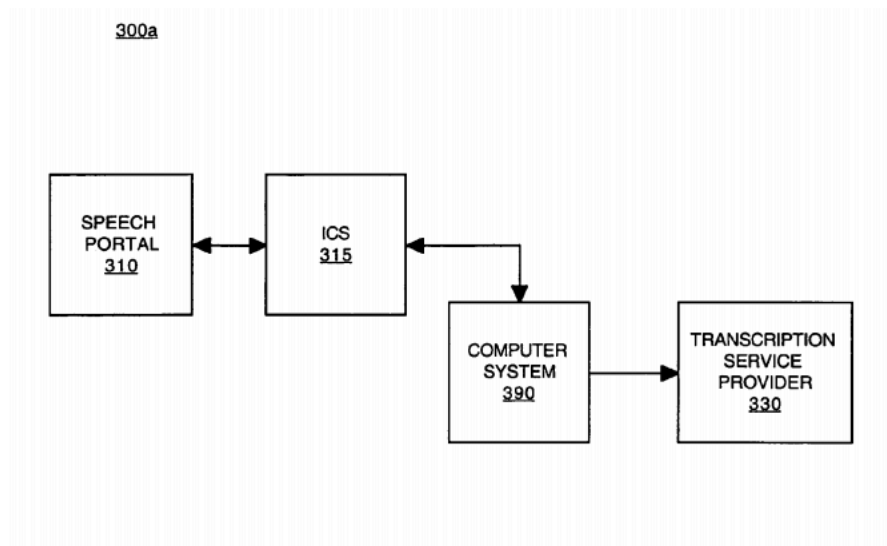
***Cohn*’s network center provides centralized functionality such as message tracking and billing**

Although *Cohn's* network hubs and their respective functionalities are distributed through the system, *Cohn* does teach that certain functionality is centralized in its network center 37. For example, *Cohn's* network center 37 is responsible for receiving and transmitting updates to and from all network hubs in the communications system, storing “all” user information, monitoring operations of the network, and tracking fees to be charged across the system. *Id.* at 8:56-62, 10:33-50, 23:23-40, 33:55-34:6.

It would have been obvious to centralize *Cohn's* transcription functionality—relocating it from the network hubs to the network center

While *Cohn* teaches transcription of a user's speech, *Cohn* teaches that the transcription is done at each individual network hub, using its own system protocol, instead of at a centralized server utilizing a uniform protocol. *See, e.g., Id.* at 9:8-43, Fig. 1. However, centralized voice transcription was well known at the time. For example, *Obilisetty* teaches a configuration in which a central server application transcribes a user's spoken words. Specifically, *Obilisetty* describes “a centrally manageable and accessible system...for capturing, transcribing and delivering information” that utilizes a “WWW” (internet) server 390 to provide remote transcription services. *Obilisetty* (EX1005) at 2:53-3:5, 6:55-57, 8:25-40, 9:13-17, *see also id.* at Abstract, 9:38-10:6. As embodied in Figure 3A below, *Obilisetty* teaches a system in which a user's spoken words are received at a speech portal 310 and provided to computer system 390 (described in *Obilisetty* as a “web

server”) that interacts with a transcription service provider 330 to generate and return a transcription of the speech in response:



Id. at Fig. 3A, 8:41-48 (“...transcription service provider 330 returns to computer system 390 a transcript file providing a transcribed version of the voice file...”); *see also id.* at 7:61-8:8, 6:4-14 (describing the computer system 390 as a “web server”); Figs. 3B, 4.

A PHOSITA would have been motivated to supplant the transcription functions of *Cohn’s* network hubs with *Obilisetty’s* single centralized transcription function located in *Cohn’s* network center 37 for a number of reasons. First, the proposed modification provides the advantage of a single, centralized transcription system that is simpler and more efficient to manage and maintain. *Creusere Dec.* (EX1003) at ¶61. A PHOSITA would have recognized that centralizing this function would have significantly reduced the cost of the overall system, reducing

internal hardware components duplicated within each network hub, such as media translator 69, and reducing the number of components that would need to be updated with new software or functionality. *Cohn* (EX1006) at 11:1-12:19, Fig. 2; *Creusere Dec.* (EX1003) at ¶61. A PHOSITA would have further recognized that centralized server applications were well-known with networked computing systems and were widely deployed as means to minimize network administration time and cost. *Id.*

A PHOSITA would have recognized that centralizing *Cohn's* transcriptions would have created significant efficiencies. For example, *Cohn* explains that “[e]ach network hub . . . includes the ability to translate messages [, including] the ability to translate text to speech . . . and speech to text.” *Cohn* (EX1006) at 22:66-23:3. *Cohn* also teaches that “media translation may be accomplished according to the desires of the message recipient.” *Id.* at 23:9-11. A PHOSITA would have recognized that performing transcriptions at a central location creates an efficient system in which each hub need only be aware of the requirements of its own users. If a recipient desires a specific translation, the centralized system can accommodate the request. But if the message need only accommodate disparate protocols between sender and receiver, rather than additionally transcribing the media format, the latter transcription can be avoided entirely. Accordingly, performing all message transcriptions at one central location as taught by

Obilisetty, saving processing overhead, time, and potential errors. *Creusere Dec.* (EX1003) at ¶61.

Second, *Cohn* expressly contemplates embodiments in which a portion of its network hubs (steps performed by network processor 60, an element within each hub) may be appropriately “implemented as a task on a central server.” *Cohn* (EX1006) at 16:53-17:9 (“...under one embodiment of the present invention, the network processor 60 is implemented as a task on a central server...”), Figs. 1, 7. This express suggestion provided by *Cohn* would have motivated a PHOSITA to consider improvements that would be gained by relocating certain hub-based functionalities centrally, such as centralizing transcriptions pursuant to the teachings of *Obilisetty*. *Creusere Dec.* (EX1003) at ¶61.

An appropriate logical location for placing the central interconnect point system transaction manager taught by the combination of *Cohn* and *Obilisetty* would have been at network center 37, which is already taught by *Cohn* as being a central monitoring system element that communicates with the remainder of the communications system, tracking messages, and updating network hubs 12,14,16 as necessary. *Id.* at 10:33-50, 23:23-40, Figs. 1, 13; *Creusere Dec.* (EX1003) at ¶62. A PHOSITA would have had a reasonable expectation of success for this combination at least because both *Cohn* and *Obilisetty* describe systems that employ TCP to facilitate the process of converting speech to text. *See, e.g., Cohn* at

11:20-26, 22:67-23:3; *Obilisetty* at 6:23-33, 8:41-48; *Creusere Dec.* (EX1003) at ¶¶61-62.

Cohn's network hubs generate the claimed speech information, which is sent to the network center 37 for transcription in the proposed combination

The parties have agreed that a “speech information request” is “[f]ormatted speech, which can be acted upon by system components, including the System Transaction Manager.” *Supra*, Section III.C. *Cohn* teaches user-generated messages that are transmitted in a formatted form, such as in the AMIS (Audio Message Interchange Specification) protocol format of DTMF (Dual Tone Multi-Frequency) signaling:

In operation, a network hub such as network hub 12 will receive a message through, for example, messaging system 24 shown in FIG. 1. By way of example, suppose messaging system 24 utilizes the analog AMIS protocol form of DTMF signaling. Further, assume that the message from messaging system 24 is intended for a party serviced by private system 42 which utilizes a proprietary digital communication protocol. Accordingly, network hub 12 would receive the message from messaging system 24 in the analog AMIS protocol.

Cohn (EX1006) at 10:1-21 (emphasis added), *see also* 8:45-9:4, 9:8-43, 16:9-46. Specifically, such messages may include requests spoken by a user. *See, e.g.*, 5:18-27; 10:1-21; 15:49-52; 16:9-25; 28:1-18.

In the proposed combination, these speech information requests comprising formatted spoken text are sent to *Cohn's* network center 37 for any necessary

speech-to-text transcription. Accordingly, *Cohn's* network center 37 is also the system transaction manager because it is a centralized point that handles these speech information requests.

In the proposed combination, *Cohn's* transcription functionality, centralized pursuant to *Obilissety's* teachings, generates formatted transcriptions of the formatted spoken text

The parties have agreed that “response(s) to [the] speech information request(s)” should be construed as “[a] formatted transcription of formatted speech.” *Supra*, Section III.C. As noted above, *Cohn* teaches that its network hubs are operable to receive speech information requests comprising formatted spoken text. Further, *Cohn* teaches that the messages received by a user may comprise a translation of the formatted speech into text, which a PHOSITA would have understood is a “transcription” of the formatted text. *Id.* at 22:66-23:8 (“Each network hub also includes the ability to translate messages into various forms of message media. These media translation features include the ability to translate text to speech, facsimile images to text and text images, text and text images to facsimile images, and **speech to text.**”); *Creusere Dec.* (EX1003) at ¶61. Further, *Cohn* teaches that the delivered message may need to be converted to a different “format” from which they are received, indicating that the transcription provided to the end user is itself “formatted.” *Id.* at 16:9-27.

In the proposed combination, this transcription functionality would be removed from *Cohn's* network hubs and consolidated in *Cohn's* network center 37. Accordingly, *Cohn's* modified network center 37 is the system transaction manager that “handl[es] speech information requests and responses to speech information requests” as claimed.

[15(b)] a first user application service adapter communicating with at least one user and the system transaction manager, the first user application service adapter configured to generate speech information requests from spoken text produced by the at least one of the users through a first protocol;

As discussed above in Section III.C, an “application service adapter” is “[a]n application layer within the Speech Recognition and Transcription System that provides an interface among Users, Speech Recognition and Transcription Engines, the System Transaction Manager, and other system components by allowing a User’s existing application and/or a system components application to communicate with the Transaction Manager,” while a “user application service adapter” should be construed to be “[a] specific Application Service Adapter that handles formatting and Routing of Speech Information Requests and Responses to elements of a User’s Protocol within the Speech Recognition and Transcription System.” *Supra*, Section III.C.

As the construction makes clear, the primary purpose of the UASA is to handle formatting and routing speech information requests and responses within the overall speech recognition and transcription system. Additionally, limitation

15(b) makes clear that the first user application service adapter must generate the speech information request from spoken text produced by the at least one of the users through a first protocol.

As additionally discussed above in Section IV.C at limitation 15(a), a PHOSITA would have found it obvious and would have been motivated to relocate the media transcription functionalities of the network hubs 12,14,16 to a server application operating centrally at network center 37. *Creusere Dec.* (EX1003) at ¶¶61, 62. The remaining unmodified portions of the network hub are operable to provide an interface among users, the speech recognition and transcription engine (mapped to translator bank 144, as shown below for limitation 15(c)), the system transaction manager (the server application), and other system components by handling formatting and routing of speech information requests, and thus constitute a first UASA pursuant to the proposed construction.

Cohn teaches that its network hubs are responsible for transforming incoming user-spoken messages (Speech Information Requests) into a shared internal protocol format, such as MIME:

The communications system of the present invention uses a shared internal protocol for all routing and processing of messages. As such, the communications system of the present invention can easily be adapted as new communications protocols, such as Multipurpose Internet Mail Extension (MIME) or the like, become popular. The

network hubs that are required to interface with systems utilizing these new protocols need only translate the new protocol to the internal protocol. As such, the operation of the communications system as a whole can support the addition of unlimited new protocols.

Id. at 10:20-31 (emphasis added), *see also* 16:9-44 (explicitly describing the internal protocol as a “format”), 22:49-23:21. Thus, ignoring media transcription that has been centralized in the proposed combination, the network hub are still operable to provide transformation of incoming messages from a proprietary protocol to a uniform system “internal protocol.”

Further, *Cohn* teaches that processors within its network hubs are operable to route speech information requests and manage information related to the sender and recipient:

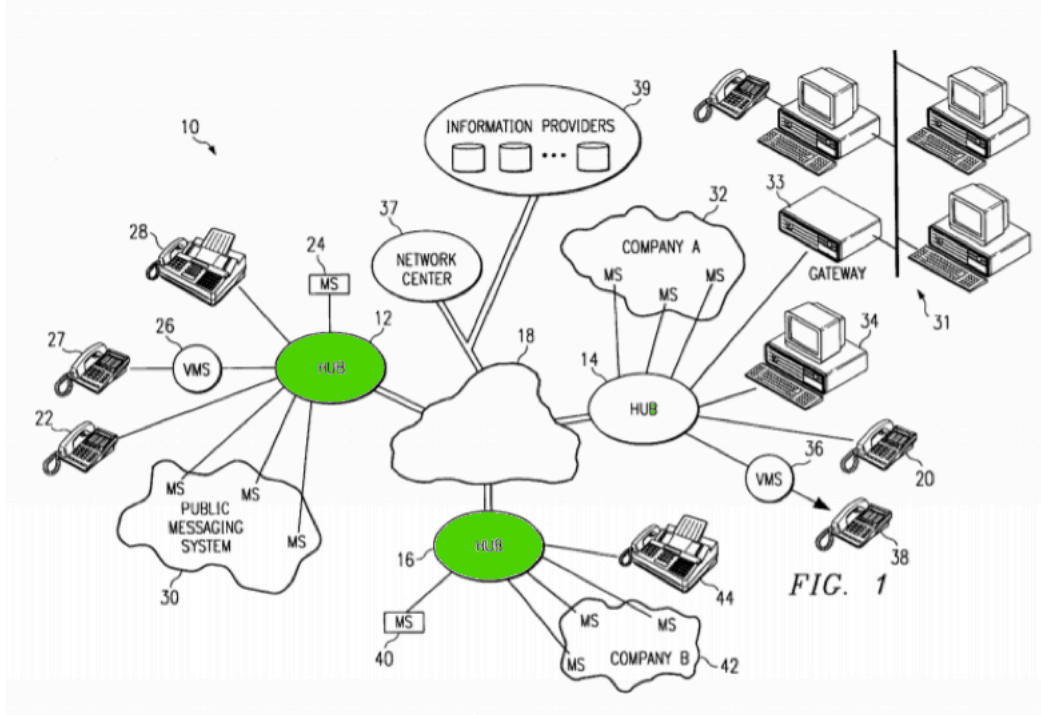
Both the analog connection processor 52 and the digital connection processor 54 are responsible for accepting and validating incoming messages.

In part, these functions are accomplished by encapsulating all received message data with a standard message wrapper to form a message for transport and storage within the communications system 10. This wrapper may be based, for example, on MIME encapsulation protocol. Many messages received by the communications system 10 already include a message wrapper...Each message also contains a “from” field where both the network identification of the message

sender and the message sender's messaging system identifier are combined into an Internet style address. Each message also contains a "to" field where the network identification of the intended recipient and the messaging facility of the intended destination facility are combined into an Internet style address.

Id. at 16:9-44, *see also* 10:20-31 (discussing "routing" messages), 11:1-43 (describing internal components of network hubs), Fig. 2. This functionality would remain in the hubs in Petitioner's proposed combination. Thus, the remaining portions of the message-transmitting user's hub are further operable to route incoming messages toward receiving users.

As an example, consider the disclosed embodiment in which a message is provided by a user of messaging system 24 relying on network hub 12 intended for a party serviced by private system 42 (Company B) relying on network hub 16, as seen annotated Figure 1 below:



Id. at 10:1-32, Fig. 1 (annotated). Network hub 12, in this scenario, is responsible for adapting the message from the analog AMIS protocol used by system 24 into the shared “internal protocol” as well as transmitting (e.g., routing) the message to the appropriate receiving network hub 16. *Id.* In this case, these functions of hub 12 provide the first UASA configured to generate speech information requests from spoken text produced by the at least one of the users through a first protocol.

Cohn also teaches the network hubs satisfy a layer within the speech transcription system that provides an interface between a user who is speaking a message and the user who will receive the transcribed message. *Cohn* (EX1006) at 16:23-46 (describing that the network hubs will format the provided information to “label[] each message media within a message with the addressing information

provided by the sender.”) (emphasis added); *Creusere Dec.* (EX1003) at ¶63. (concluding that *Cohn*’s network hubs constitute a “layer” as that term is used in the proposed construction). Finally, as noted above, *Cohn* specifically teaches that its message spoken by the first user may be transmitted in a first protocol that is different than the protocol used by the receiving user and, therefore, requires conversion to a different format. *Id.*, at 10:1-31. Thus, the system utilizes two different protocols for the two different users in addition to the internal protocol. *Id.*

Therefore, the unmodified portions of network hub serving the transmitting user constitute the first user application service adapter pursuant to the proposed construction of that phrase.

[15(c)] a speech recognition and transcription engine communicating with the system transaction manager, the speech recognition and transcription engine configured to receive speech information requests from the system transaction manager, to generate responses to the speech information requests, and to transmit the responses to the system transaction manager; and

A “speech recognition and transcription engine” (SRTE) is “[a] process running on a computer that recognizes an audio file and transcribes that file to written text to generate a transcription of Speech.” *Supra*, Section III.C.

As discussed above, *Cohn* discloses that formatted spoken messages are transmitted to network hubs 12,14,16 that include system components operable to “translate messages into various forms of message media.” *Cohn* further discloses

that messages are then transmitted to a translator bank 144, where they are processed into text, satisfying speech information “responses” that comprise a formatted transcription of formatted speech:

Referring to FIG. 12, media translator 69 comprises a control module 133 and a translator bank 144. The control module 133 interacts with the event processor 70 using the hub control protocol. The translator bank 144 receives messages from the message store 58 by making requests to file server 59. The translator bank 144 operates to convert the format of data and to translate the media of messages. The messages processed by translator bank 144 are then returned to message store 58 using file server 59. Translator bank 144 comprises a plurality of translators indicated as 144a-n in FIG. 12. Each of the translators 144a-n within translator bank 144 accomplish a stage of the conversion to be performed by translator bank 144. Each translation or conversion is done by first converting the message or data into a high quality internal format that contains sufficient data representation to insure that no data is lost during the conversion process. Translations of messages from one media to another are performed while the message is represented in the internal format. In this manner, the translator bank 144 need only contain a minimum number of media to media translators. In summary, the control module 133 can cause the translator bank 144 to use many different filters in sequence to accomplish any permutation of media-to-media translation, message-to-message transformation or data-to-data conversion.

Id. at 22:24-48. (emphasis added). *Cohn*'s translator bank 144 is a process running on a computer that transcribes audio files and therefore satisfies a "speech recognition and transcription engine." As discussed in Section IV.C at limitation 15(a), the proposed combination consolidates *Cohn*'s media transcription functionalities, including this translator bank 144, in network center 37 pursuant to the teachings of *Obilisetty*. Thus, *Cohn* in view of *Obilisetty* renders obvious the SRTE pursuant to the proposed construction of this phrase and would be located centrally with the system transaction manager in network center 37, as discussed above in Section IV.B at limitation 15(a). A PHOSITA would have understood that *Cohn*'s translator bank 144 would thus be a functional component of the network center 37 in the proposed combination—in communication with the components that satisfy the STM, receiving speech information requests from those components, and transmitting responses to those components, as claimed. *Creusere Dec.* (EX1003) at ¶62.

[15(d)] a second user application service adapter communicating with one or more of the users and with the system transaction manager, the second user application service adapter which is different than the first user application service adapter and configured to provide the one or more users with a transcription of the spoken text that is compatible with a second protocol, the second protocol being the same as or different than the first protocol.

As discussed above in Section IV.C at limitation 15(b), the unmodified portions of *Cohn*'s network hub serving a sending user constitute a first user

application service adapter. Similarly, the unmodified portions of *Cohn*'s network hub serving a receiving user constitute a second user application service adapter.

Returning to the example discussed above in which a message is provided by a user of messaging system 24 relying on network hub 12 and received by a user serviced by private system 42 (Company B) network hub 16, in this scenario, is responsible for using the proprietary communication protocol understood by private system 42 to transmit the message. This message would comprise a formatted transcription of the formatted spoken text (a "Speech Information Response") in the case discussed above in Section IV.B at limitation 15(b). *Id.* Thus, these functions of hub 16 provide the second UASA providing the user with a transcription of the spoken text that is compatible with a second protocol.

Like hub 12 (providing the "first user application service provider"), hub 16 provides an interface between users and components in communication system 10 such as translator bank 144 (mapped to the "speech recognition and transcription engine" above), network center 37 (reconfigured as the location of the System Transaction Manager), and other users (such as those at Company B and on the Public Messaging System). *Id.* at Fig. 1, *see also* 10:1-22. This second hub will utilize its own proprietary protocol that is a different protocol than the first hub. *Id.*, at 10:1-31.

Therefore, the portions of network hub serving the receiving user that have not been reconfigured into a server application as discussed above for Section IV.C at limitations 15(a)-15(b) constitute the second user application service adapter as that phrase is properly construed. *Supra*, Section III.C.

ii. Claim 17

17. A method of exchanging transcribed spoken text among users, the method comprising:

To the extent that the preamble is limiting, *Cohn* teaches a method of exchanging transcribed spoken text among users. *See, e.g., Cohn* at Abstract, 1:19-23, 2:54-63 (“...a network-based voice and multimedia communications system is provided that comprises a plurality of network hubs operable to communicate with each other and with local messaging systems. The network hubs operate to receive messages in any media using any addressing convention from users either directly or through local messaging systems and to transmit the messages to other users whether located on local messaging systems or otherwise through the plurality of network hubs.”).

[17(a)] generating a speech information request from spoken text obtained through a first protocol, the speech information request comprised of formatted spoken text generated using a first user application service adapter;

Cohn in view of *Obilisetty* teaches generating a speech information request from spoken text obtained through a first protocol comprised of spoken text generated using a first user application service adapter. *See supra* Section IV.C at

limitations 15(a)-(b). Specifically, *Cohn* in view of *Obilisetty* also renders obvious a method of exchanging transcribed spoken text among users, such as those at Company B and those on the Public Messaging System.

[17(b)] transmitting the speech information request to a speech recognition and transcription engine via a system transaction manager;

Cohn in view of *Obilisetty* teaches transmitting the speech information request to a speech recognition and transcription engine via a system transaction manager. *See supra* Section IV.C at limitation 15(a)-(c).

[17(c)] generating a response to the speech information request using the speech recognition and transcription engine, the response comprised of a formatted transcription of the formatted spoken text;

Cohn in view of *Obilisetty* teaches generating a response to the speech information request comprised of a formatted transcription of the formatted spoken text using the speech recognition and transcription engine. *See supra* Section IV.C at limitations 15(a), (c).

[17(d)] transmitting the response to a user via the system transaction manager; and

Cohn in view of *Obilisetty* teaches transmitting a response to speech information request to a user via the system transaction manager. *See supra* Section IV.C at limitations 15(a), (d).

[17(e)] providing the user with a processed transcription of the spoken text using a second user application service adapter, the processed transcription being compatible with a second protocol that is different than the first protocol.

Cohn teaches providing the user with a processed transcription of the spoken text using a second user application service adapter, the processed transcription being compatible with a second protocol. *See supra* Section IV.C at limitations 15(a), (d). Specifically, *Cohn* expressly teaches that the protocols used by the networked computer receiving the transcription may be distinct from protocols used by the messaging system of the transmitting user. *Cohn* (EX1006) at 10:1-31, 22:49-65.

V. CONCLUSION

For the forgoing reasons, Petitioner respectfully requests *inter partes* review of claims 15 and 17 of the '730 Patent.

Respectfully submitted,

BY: /s/ Adam P. Seitz
Adam P. Seitz, Reg. No. 52,206
Paul R. Hart, Reg. No. 59,646

COUNSEL FOR PETITIONER

VI. MANDATORY NOTICES UNDER 37 C.F.R. § 42.8(A)(1)

A. Real Party-In-Interest

The Petitioner is the real party-in-interest. 37 C.F.R. § 42.8(b)(1).

B. Related Matters

Pursuant to 37 C.F.R. § 42.8(b)(2), Petitioner is aware of only one patent infringement lawsuit involving the '730 Patent:

- *Advanced Voice Recognition Systems, Inc. v. Apple Inc.*, No. 2:18-cv-02083 (D. Arizona). The complaint alleging infringing of the '730 Patent was served on Petitioner Apple Inc. on July 23, 2018. (EX1020)

C. Lead and Back-Up Counsel

Petitioner provides the following designation and service information for lead and back-up counsel. 37 C.F.R. § 42.8(b)(3) and (b)(4).

Lead Counsel	Back-Up Counsel
Adam P. Seitz (Reg. No. 52,206) Adam.Seitz@eriseip.com	Paul R. Hart (Reg. No. 59,646) Paul.Hart@eriseip.com
<u>Postal and Hand-Delivery Address:</u> ERISE IP, P.A. 7015 College Blvd., Suite 700 Overland Park, Kansas 66211 Telephone: (913) 777-5600 Fax: (913) 777-5601	<u>Postal and Hand-Delivery Address:</u> ERISE IP, P.A. 5600 Greenwood Plaza Blvd. Suite 200 Greenwood Village, CO 80111 Telephone: (913) 777-5600 Fax: (913) 777-5601

APPENDIX OF EXHIBITS

Exhibit 1001	U.S. Patent No. 7,558,730 (“’730 Patent”)
Exhibit 1002	File History of U.S. Patent No. 7,558,730 (“ <i>File History for ’730 Patent</i> ”)
Exhibit 1003	Declaration of Dr. Charles D. Creusere (including Appendix A) (“ <i>Creusere Dec.</i> ”)
Exhibit 1004	U.S. Patent No. 6,487,278 to Skladman et al. (“ <i>Skladman</i> ”)
Exhibit 1005	U.S. Patent No. 7,558,735 to Obilisetty (“ <i>Obilisetty</i> ”)
Exhibit 1006	U.S. Patent No. 6,064,723 to Cohn et al. (“ <i>Cohn</i> ”)
Exhibit 1007	U.S. Patent No. 2,200,351 to Whitehead et al. (“ <i>Whitehead</i> ”)
Exhibit 1008	U.S. Patent No. 4,430,726 to Kasday (“ <i>Kasday</i> ”)
Exhibit 1009	U.S. Patent No. 4,692,042 to Cuff et al. (“ <i>Cuff</i> ”)
Exhibit 1010	U.S. Patent No. 5,754,830 to Butts et al. (“ <i>Butts</i> ”)
Exhibit 1011	U.S. Patent No. 6,122,614 to Kahn et al. (“ <i>Kahn</i> ”)
Exhibit 1012	U.S. Patent No. 6,173,259 to Bijl et al. (“ <i>Bijl</i> ”)
Exhibit 1013	U.S. Patent No. 6,374,223 to Donaldson et al. (“ <i>Donaldson</i> ”)
Exhibit 1014	U.S. Patent No. 6,408,272 to White et al. (“ <i>White</i> ”)
Exhibit 1015	U.S. Patent No. 6,938,087 to Abu-Samaha (“ <i>Abu-Samaha</i> ”)
Exhibit 1016	Assorted Non-Patent Literature re: JFAX Communicator Software (“ <i>JFAX</i> ”)
Exhibit 1017	U.S. Patent No. 6,434,526 to Cilurzo et al. (“ <i>Cilurzo</i> ”)
Exhibit 1018	Andrew S. Tanenbaum, <i>Computer Networks</i> , Prentice-Hall, Inc., 1981, pp.7-21 (“ <i>Tanenbaum</i> ”)
Exhibit 1019	File History of U.S. Patent App. No. 09/996,849 (“ <i>File History for ’849 Application</i> ”)
Exhibit 1020	Proof of Service for Complaint alleging infringing of the ’730 Patent
Exhibit 1021	Claim Constructions Adopted by the District Court for the ’730 Patent
Exhibit 1022	U.S. Patent No. 6,850,609 to Schrage et al. (“ <i>Schrage</i> ”)

CERTIFICATION OF WORD COUNT

The undersigned certifies pursuant to 37 C.F.R. §42.24 that the foregoing Petition for *Inter Partes* Review, excluding any table of contents, mandatory notices under 37 C.F.R. §42.8, certificates of service or word count, or appendix of exhibits, contains 13,874 words according to the word-processing program used to prepare this document (Microsoft Word).

Dated: July 22, 2019

BY: /s/ Adam P. Seitz
Adam P. Seitz, Reg. No. 52,206

COUNSEL FOR PETITIONER

**CERTIFICATE OF SERVICE ON PATENT OWNER
UNDER 37 C.F.R. § 42.105**

Pursuant to 37 C.F.R. §§ 42.6(e) and 42.105, the undersigned certifies that on July 22, 2019, a complete and entire copy of this Petition for *Inter Partes* Review including exhibits was provided via Federal Express to the Patent Owner by serving the correspondence address of record for the '730 Patent as listed on PAIR:

ASCENDA LAW GROUP
333 W. San Carlos Street
Suite 200
San Jose, CA 95110

BY: /s/ Adam P. Seitz
Adam P. Seitz, Reg. No. 52,206

COUNSEL FOR PETITIONER