

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ZOHO CORPORATION and
ZOHO CORPORATION PVT., LTD.
Petitioner,

v.

MEETRIX IP, LLC,
Patent Owner.

U.S. Patent No. 9,253,332

DECLARATION OF HENRY H. HOUH, PH.D., UNDER 37 C.F.R. § 1.68 IN
SUPPORT OF PETITION FOR *INTER PARTES* REVIEW OF
U.S. PATENT NO. 9,253,332

I. INTRODUCTION

1. I have been retained by the petitioner, Zoho Corporation, to provide my expert opinions as to certain issues in connection with a petition for Inter Partes Review (IPR) in the United States Patent and Trademark Office challenging certain claims of U.S. Patent Nos. 9,253,332 (“the ’332 patent”), entitled “Voice Conferencing Call Using PSTN and Internet Networks.”

2. This Declaration is made based on my personal knowledge, expertise, and experience and based on my review of the materials cited herein. If required, I would testify competently and truthfully regarding the contents of this Declaration.

3. I have no personal, financial, or other commercial interest in the outcome of this IPR. Aside from this consulting engagement, I have no relationship with the petitioner.

4. I am being compensated for my work in this matter at an hourly rate. I am also being reimbursed for reasonable and customary expenses associated with my work and testimony in this matter. My compensation is not contingent on the outcome of this matter or the specifics of my testimony. I have previously been retained as an expert witness in over 100 cases.

II. SUMMARY AND MATERIALS CONSIDERED

5. I have been requested to provide my expert opinions regarding the multi-participant conferencing technology that is the subject matter of the ’332

patent, including how a person of ordinary skill in the art (POSITA) would have interpreted the '332 patent and the prior art I discuss herein.

6. I have also been asked to provide my opinions regarding the validity of claims 1-12 of the '332 patent ("Challenged Claims"), including what a POSITA as of the earliest possible priority date of the '332 patent would have found obvious or not obvious in light of the state of the art and the cited prior art references. It is my opinion that the Challenged Claims would have been obvious to a POSITA after reviewing the prior art discussed below.

7. In preparing this Declaration, I have considered my own knowledge and experience, including my work experience in the field of networking systems, including audio/video conferencing systems, and my experience in working with others involved in this field.

8. In preparing this Declaration, I have studied the following references:

- the '332 patent (Ex. 1001)
- U.S. Patent No. 7,180,997 to Knappe ("Knappe") (Ex. 1006);
- U.S. Patent No. 6,690,654 to Elliott, *et al.* ("Elliott") (Ex. 1007);
- U.S. Patent No. 7,089,285 to Drell ("Drell") (Ex. 1008);
- "MPLS and VPN Architectures, A practical guide to understanding, designing and deploying MPLS and MPLS-enabled VPNs," by Guichard et al., Cisco Press, 2001 ("VPN Textbook"), excerpt (Ex. 1010); and
- U.S. Patent No. 6,701,437 to Hoke, *et al.* ("Hoke") (Ex. 1009).

9. In forming the opinions expressed in this Declaration I have also considered the following:

- the prosecution history of the '332 Patent (Ex. 1005);
- U.S. Patent No. 3,060,265 (Ex. 1012);
- U.S. Patent No. 4,139,731 (Ex. 1013);
- U.S. Patent No. 3,601,530 (Ex. 1014);
- U.S. Patent No. 4,796,293 (Ex. 1015);
- RFC 760 (Ex. 1028);
- RFC 791 (Ex. 1029);
- RFC 1112 (Ex. 1030);
- RFC 2401 (Ex. 1031);
- RFC 2406 (Ex. 1032);
- A Primer on the T.120 Series Standard (Ex. 1016);
- T.120 Standard (1996) (Ex. 1017);
- U.S. Patent No. 6,237,025 (Ex. 1018);
- U.S. Patent No. 6,782,413 (Ex. 1019);
- U.S. Patent No. 6,980,526 (Ex. 1020);
- U.S. Patent Pub. No. 2004/0261116 (Ex. 1023);
- "Cisco Packetized Voice & Data Integration," Robert Caputo, McGraw-Hill, 2000 ("Caputo") (Ex. 1021);
- "The Internet Multicast from ITS: How it was Done and Implications for the Future," by de Moraes *et al.*, IEEE Communications Magazine, January 1995 (Ex. 1022);

- Computer Desktop Encyclopedia (Ex. 1027);
- U.S. Patent No. 6,850,496 (Ex. 1024);
- U.S. Patent No. 6,553,515 (Ex. 1025);
- U.S. Patent No. 6,633,848 (Ex. 1026);
- H.323 Standard (Ex. 1033);
- U.S. Patent No. 7,266,091 (Ex. 1034);
- U.S. Patent No. 5,606,550 (Ex. 1035); and
- any additional documents discussed in my Declaration below.

10. I reserve the right to respond to further comments or questions regarding this IPR in order to clarify or supplement this Declaration.

III. QUALIFICATIONS

11. My qualifications and professional experience are described in my *curriculum vitae*, a copy of which is submitted as Ex. 1004. The following is a brief summary of my relevant qualifications.

12. I received a Ph.D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (“MIT”) in 1998. I also received a Masters of Science degree in Electrical Engineering and Computer Science in 1991, a Bachelor of Science degree in Electrical Engineering and Computer Science in 1989, and a Bachelor of Science degree in Physics in 1990, all from MIT.

13. I am currently self-employed as an independent technical consultant. From its founding in 2012 until October 2022, I was also President of Einstein's Workshop, which provides supplemental science, technology, engineering, and mathematics ("STEM") education to children of all ages. I am also a member of the board of a spin-off of Einstein's Workshop, BlocksCAD Inc., which provides an education-focused 3D computer-aided design tool with over 275,000 users worldwide.

14. I first worked in the area of telecommunications in 1987 when I worked as a summer intern at AT&T Bell Laboratories as part of a five-year dual degree program at MIT. I continued to work at AT&T Bell Laboratories as part of this MIT program.

15. Later, as part of my doctoral research at MIT from 1991-1998, I was a research assistant in the Telemedia Network Systems ("TNS") group at the Laboratory for Computer Science. The TNS group built a high-speed gigabit network and created applications that ran over the network. Example applications included remote video capture, video processing, and display of video on computer terminals. In addition to working on the design of core network components, designing and building the high-speed links, and designing and writing the device drivers for the interface cards, I also set up the group's web server.

16. The TNS group was the first group to initiate a remote video display over the web. I helped to build the web pages that initiated these video sessions via a web interface. Vice-President Al Gore visited our group in 1996 and received a demonstration of – and remotely drove – a radio-controlled toy car with a wireless video camera mounted on it; the video was encoded by TNS-designed hardware, streamed over the TNS-designed network and displayed using TNS-designed software.

17. We also configured and maintained all the networking components in our research group, including configuring the host computers, occasionally configuring routers, and at one point setting up a multicast router on the Internet's multicast backbone.

18. I submitted and defended my Ph.D. thesis, titled "Designing Networks for Tomorrow's Traffic," in January 1998. As part of my thesis research, I analyzed local-area and wide-area flows to show a more efficient method for routing packets in a network, based on traffic patterns at the time. My thesis also addressed real-time communications of audio and video streams and proposed more efficient methods to route such data within a network. The flow classification methods I analyzed are similar to modern label switching techniques such as Multiprotocol Label Switching (MPLS).

19. From 1997 to 1999, I was a Senior Scientist and Engineer at NBX Corporation, a start-up that made business telephone systems for streaming packetized audio over data networks instead of using traditional telephone lines. NBX was later acquired by 3Com Corporation, and the phone system is still used today by numerous businesses. As part of my work at NBX, I designed the core audio reconstruction algorithms for the telephones, as well as the packet transmission algorithms. I also designed and validated the core packet transport protocol used by the phone system. The protocol was used for all signaling in the phone system, including for the setup of conference calls. The NBX system also featured a computer interface for initiating phone calls and conference calls. Conference calls in the NBX system utilized conference mixing software which ran on every telephone in the system. Participants in the conference call, which could include telephones in the PSTN participating through a gateway, multicast their audio packets onto the packet network. While the NBX system was initially based on the Ethernet, I implemented the prototype IP stack for the system, and demonstrated our IP telephones connecting through a VPN connecting from various trades shows through the Internet to our office in Andover, MA. I configured all the routing and VPN appliances involved in any such public demonstrations. The NBX system also supported TAPI, the Telephony Application Programming Interface, thus allowing other computer programs to

integrate with the system's telephony features. We obtained U.S. Patent No. 6,697,963, titled "Telecommunication method for ensuring on-time delivery of packets containing time-sensitive data," as a result of part of this work.

20. From 1999-2004, I was employed by Empirix or its predecessor company, Teradyne. Empirix was a leader in test tools for telecommunications protocols and systems, providing functional testing tools as well as load testing tools. From 2000-2001, I conceived of and built a test platform for testing Voice-over-IP (VoIP). The first application on this new test platform was a cloud emulator for simulating the effects of transmitting VoIP over a busy network.

21. From 2004 to 2008, I was employed by BBN Technologies Corp., a technology research and development company located in Cambridge, MA. BBN Technologies is a world-renowned company with expertise in acoustics, speech recognition, and communications technology. BBN Technologies staff have pioneered many internetworking technologies and Internet applications, and built some of the world's largest government and commercial data networks.

22. My duties and responsibilities at BBN Technologies generally included commercialization of the technologies developed by the company, which included spinning off companies and growing commercial businesses in-house. More particularly, I was involved in utilizing the award-winning AVOKE STX speech recognition technology to create the public audio/video search engine

EveryZing (formerly known as PodZinger) which was spun out into a stand-alone company now known as RAMP, Inc. PodZinger won the 2006 MITX Technology Award for best Web 2.0 Application and was also named the 2006 Forbes Favorite Video & Audio Search Engine, beating out Google, Yahoo, and other companies. After managing the creation of the initial prototype system, PodZinger built out a full streaming audio and video search solution when I was the Vice President of Operations and Technology there. I was also involved in the Boomerang Mobile Shooter Detection project as the Vice President of Engineering for the program. The Boomerang system was deployed to Iraq and Afghanistan, and was credited with saving many lives. I also worked closely with Steve Kent, who was a Chief Scientist at BBN and one of the authors of IPsec, on several projects.

23. I have been awarded several United States patents, and I have several patent applications pending including the following examples:

- U.S. Patent No. 7,975,296, “Automated security threat testing of web pages”;
- U.S. Patent No. 7,877,736, “Computer language interpretation and optimization for server testing”;
- U.S. Patent No. 7,801,910, “Method and apparatus for timed tagging of media content”;
- U.S. Patent 7,590,542, “Method of Generating Test Scripts Using a Voice-Capable Markup Language”;
- U.S. Patent No. 6,967,963, “Telecommunication method for ensuring on-time delivery of packets containing time-sensitive

data”;

- U.S. Patent Publication No. 20070106685, “Method and apparatus for updating speech recognition databases and reindexing audio and video content using the same”;
- U.S. Patent Publication No. 20070106693, “Methods and apparatus for providing virtual media channels based on media search”;
- U.S. Patent Publication No. 20070106760, “Methods and apparatus for dynamic presentation of advertising, factual, and informational content using enhanced metadata in search-driven media applications”;
- U.S. Patent Publication No. 20070112837, “Method and apparatus for timed tagging of media content”;
- U.S. Patent Publication No. 20070118873, “Methods and apparatus for merging media content”;
- U.S. Patent Publication No. 20090222442, “User-directed navigation of multimedia search results.”

IV. LEVEL OF ORDINARY SKILL IN THE ART

24. I understand that the level of ordinary skill may be reflected by the prior art of record, and that a POSITA to which the claimed subject matter pertains would have the capability of understanding the scientific and engineering principles applicable to the pertinent art.

25. I understand there are multiple factors relevant to determining the level of ordinary skill in the pertinent art, including (1) the levels of education and experience of persons working in the field at the time of the invention; (2) the

sophistication of the technology; (3) the types of problems encountered in the field; and (4) the prior art solutions to those problems.

26. As discussed above and detailed in my *curriculum vitae*, I am familiar with the conferencing art that is pertinent to the '332 patent. I am also aware of the state of the art as of the earliest claimed priority date. Through my past interaction with work colleagues in industry and in academia, I am also familiar with POSITA skill level as of the earliest claimed priority date.

27. I have been informed by counsel that the earliest claimed priority date for the '332 patent is March 10, 2003. Based on the technologies disclosed in the '332 patent, I believe that a POSITA would include someone who had, at the earliest claimed priority date of the '332 patent, a working knowledge of available audio and video conferencing technologies. A POSITA would also have had a bachelor's degree in electrical engineering, computer science, computer engineering, or an equivalent, and three years of professional experience relating to conferencing systems in packet-based networks. However, I recognize that someone with less technical education but more experience, or more technical education but less experience could also have met this standard. I believe that I possessed at least such experience and knowledge as of the earliest claimed priority date of the '332 patent, and that I am qualified to opine on the subject matter in the '332 patent.

28. For purposes of this Declaration, in general, and unless otherwise noted, my statements and opinions, such as those regarding my experience and the understanding of a POSITA generally (and specifically related to the references I consulted herein), reflect the knowledge that existed in the field before the earliest claimed priority date of the '332 patent.

V. RELEVANT LEGAL STANDARDS

29. I understand that I am obliged to follow existing law as applicable to the opinions set forth herein. I am not an attorney but have been informed of certain legal principles by attorneys in preparation of this Declaration. I have been asked to apply the legal principles set forth below to my analysis. I understand these legal principles and have applied them in my analysis in this Declaration.

A. Prior Art

30. I understand that the applicable patent law provides categories of information that constitute prior art that may be used to anticipate or render obvious patent claims. To qualify as prior art to a particular patent claim under the relevant law, I understand that a reference must have been made, known, used, published, or patented, or be the subject of a patent application by another, before the priority date of the patent, and, in this matter, must satisfy one of the standards of pre-AIA 35 U.S.C. § 102. I also understand that applicable patent law provides

that an admission in a challenged patent also qualifies as prior art that can be used as evidence of obviousness under pre-AIA 35 U.S.C. § 103.

B. Claim Construction

31. In making this Declaration, I have been asked to review the terms in the Challenged Claims to determine what a POSITA would have understood those terms to mean at the time of the alleged invention. I understand that my assessment of the Challenged Claims must be undertaken from the perspective of what would have been known or understood by a POSITA reading the '332 patent as of the time of the filing or earliest claimed priority date (*i.e.*, March 10, 2003), and in light of the specification and file history of the patents. In addition, I understand that, in an *inter partes* review proceeding, the claim terms under consideration are to be construed using their ordinary and customary meaning, as would have been understood by a POSITA at the time of the alleged invention, consistent with the specification and the prosecution history of the patent.

C. Obviousness

32. I also understand that a patent claim is invalid if the differences between the claimed subject matter and the prior art are such that the subject matter as a whole would have been obvious to a POSITA at the time the invention was made.

33. I understand further that the condition “at the time the invention was made” is imposed to rule out impermissible hindsight. Accordingly, permissible reasoning takes into account only knowledge which was within the level of ordinary skill in the art at the time the claimed invention was made and does not include knowledge gleaned only from the applicant’s disclosure. I also understand that an opining expert is obligated to analyze prior art from the perspective of a POSITA, and may not simply provide his or her own personal conclusions.

34. I also understand that an obviousness determination includes several factual inquiries, including: (1) determining the scope and content of the prior art; (2) ascertaining the differences between the claimed invention and the prior art; (3) resolving the level of ordinary skill in the pertinent art; and (4) taking into consideration any secondary indicia (objective evidence) of non-obviousness.

35. I also understand that such objective evidence of non-obviousness may include: (1) a long felt but unsolved need that was satisfied by the claimed invention; (2) commercial success attributable to the claimed invention; (3) unexpected results achieved by the claimed invention; (4) praise by experts of the claimed invention with factual support; (5) taking of licenses under the patent by others for reasons related to the alleged non-obviousness of the claimed invention; (6) the failure of others; (7) the teaching away of others; (8) skepticism by experts;

and (9) evidence that competitors in the marketplace have copied the invention instead of using the prior art.

36. I also understand that, to establish nonobviousness, there must be a demonstrated relationship, or nexus, between any such secondary indicia and the claimed invention, *i.e.*, there must be objective evidence of non-obviousness attributable to the claimed invention.

37. I further understand that near simultaneous invention by two or more equally talented inventors working independently may or may not be an indication of obviousness when considered in light of all the circumstances.

38. I understand that a conclusion of obviousness can be based on a combination of multiple prior art references. I understand that examples of rationales that may support a conclusion of obviousness include:

- a. Combining prior art elements according to known methods to yield predictable results;
- b. Simply substituting one known element for another to obtain predictable results;
- c. Using known techniques to improve similar devices, methods, or products in the same way;
- d. Applying a known technique to a known device, method, or product ready for improvement to yield predictable results;

- e. Using an “obvious to try” solution, that is, choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success
- f. Applying known work in one field of endeavor to yield variations of that work for use in either the same field or a different one based on design incentives or other market forces if the variations are predictable to one of ordinary skill in the art; and
- g. Applying some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference disclosures to arrive at the claimed invention.

39. I also understand that if a technique has been used to improve one device, and if a POSITA would recognize that the technique would improve similar devices in the same way, then using that technique would be obvious to a POSITA unless actual application of that technique is beyond his or her skill.

40. I also understand that practical and common-sense considerations should guide an analysis of obviousness, because familiar items may have obvious uses beyond their primary purposes. I further understand that applying common sense does not require a “specific hint or suggestion in a particular reference.”

Rather, applying common sense requires a reasoned explanation that avoids false deductions.

41. I also understand that, when addressing a problem, a POSITA will often be able to fit together the disclosures of multiple publications. In this regard, I understand that an analysis of obviousness may consider inferences and creative steps that a POSITA would normally employ.

42. I also understand that a particular combination of prior art items may be proven obvious merely by showing that it was obvious for a POSITA to try that combination. For example, when there is a design need or a market pressure to solve a problem, and/or there exists a finite number of identified, predictable solutions with a reasonable expectation of success, a POSITA has good reason to pursue the known options within his or her technical grasp. I understand that if this leads to anticipated success, that success is likely not the result of innovation but of ordinary skill and common sense.

43. I also understand that a combination according to known methods of familiar elements is presumed to be obvious when that combination does no more than yield predictable results. Work known in one field of endeavor may prompt variations of that work for use in either the same field or a different field, based on design incentives and other market forces. If a technique has been used to improve

some device, and a POSITA can implement a predictable variation of that technique, then that variation is likely obvious.

44. I have been informed and understand that obviousness does not require physical combination of components or bodily incorporation of one device into another, but rather consideration of what the combined teachings would have suggested to a POSITA at the time of the alleged invention.

45. I further understand that to be proper for use in an analysis of obviousness, a reference must teach art that is analogous to the claimed invention. I also understand that under a correct analysis, any need or problem known in the field of endeavor at the time of the invention and addressed by the claimed invention can provide a reason for combining the elements in the manner claimed.

46. I also understand further that, in an analysis of obviousness, two items of prior art need not be like two puzzle pieces that must fit together perfectly. For example, a claimed invention may be found obvious if a POSITA would view rearrangement as an obvious design choice.

47. I also understand that an analysis of obviousness requires a comparison between the properly construed claim language and the prior art on a limitation-by-limitation basis, while observing whether the differences between the claimed subject matter and the prior art are such that the subject matter as a whole would have been obvious to a POSITA at the time the invention was made.

VI. OVERVIEW OF THE '332 PATENT

48. The '332 patent issued as a continuation of a chain of applications that claim priority to Provisional Application No. 60/453,307, which was filed on March 10, 2003. Ex. 1001 at 1:6-12.

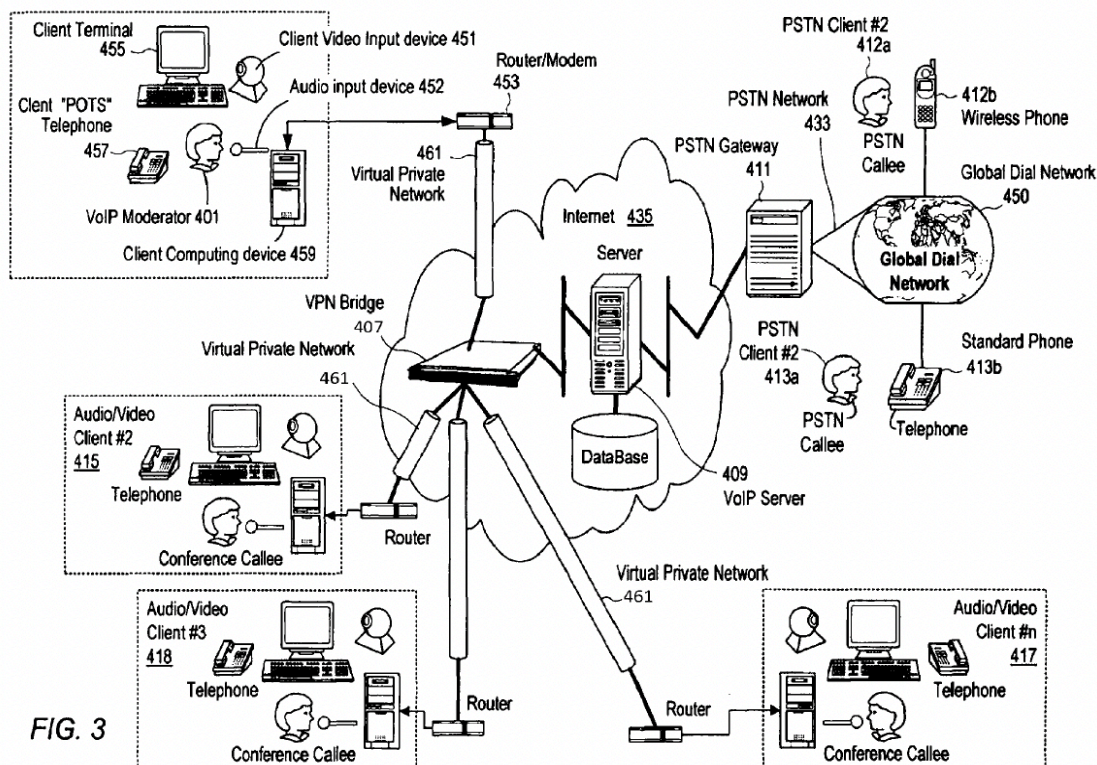
A. The '332 Patent Specification

49. The '332 patent is directed to a “system and method for supporting a multi-participant voice conference call using PSTN and Internet networks.” *Id.* at Abstract. As the Abstract explains, the method includes receiving voice data from a PSTN client, a moderator, and at least one remote client connected to the Internet. *Id.* The method mixes the voice data received from the PSTN client with that received from the moderator, and then transmits the mixed voice data to the remote Internet client. *Id.* It also mixes the voice data received from the moderator with that received from the remote client, and then transmits that mixed voice data to the PSTN client. *Id.*; *see also id.* at 3:20-33.

50. To accomplish this, the specification describes use of a hybrid network connecting a moderator, several remote Internet-based audio video clients, and one or more PSTN clients. *Id.* at FIG. 3, 4:39-5:18 (disclosing VoIP Moderator 401, Audio/Video Clients 415, 417, 418, PSTN Gateway 411, PSTN Clients 412a-413a, and VPN bridge 407 connecting various VPNs 461). Specifically, the patent describes having a moderator that “dial[s] out from a

desktop computer or terminal (using a novel hybrid network structure) connecting an external telephone user's audio into the audio/video conference." *Id.* at 4:43-47.

51. The hybrid network structure is depicted in Figure 3 of the '332 patent:



Id. at FIG. 3.

52. Specifically, Figure 3 depicts Internet-based clients, including moderator 401, and audio video clients 415, 417, and 418 "connected through routers or modems 453 preferably in a virtual private network configuration 461" with "the VPN bridge 407" located in the center of a "star topology," with connections to the VPN bridge serving as a "VPN tunnel." *Id.* at 5:7-32, 6:20-42, FIGS. 4-5. The specification explains that the virtual private network is used to

connect the local and remote Internet-based clients together within a secure private network. *Id.* at 5:10-11.

53. The hybrid network also includes “telephony clients using standard wired 413 or wireless telephone 412 systems,” with these PSTN clients connected to the Internet-based clients via PSTN gateway 411. *Id.* at 4:67-5:1.

54. A connection from the VPN bridge 407 to the a “voice over IP server 409 is used to transfer conference audio from any participant on the IP network to any participant in the PSTN.” *Id.* at 5:10-14. “The voice over IP server 409 is responsible for transcoding audio information from the virtual private network 461 to and from the PSTN gateway 411, thus bridging the PSTN and VPN together. *Id.* at 5:15-18. Specifically, the ’332 patent discloses that “voice over IP server 409 may use standard H.323 or SIP network protocol to establish communications” between PSTN and Internet clients. *Id.* at 7:47-50.

55. The specification also discloses performing encoding and decoding of audio data using codecs of the prior art H.323 standard and the transmission of collaboration data per the prior art T.120 standard. *Id.* at 2:55-61, 3:10-12, FIG. 2. The specification describes application software used to mix and transmit audio in the hybrid network. *See id.* at FIG. 6, 7:20-9:50. Specifically, it describes mixing audio received from the PSTN with audio from the moderator, explaining that audio encoder 520a local to the moderator “combines the PSTN client 412 audio

with the local moderator clients 401 audio then encodes the result for compression of the data in preparation across the VPN networks 461” which then “delivers both the PSTN client’s audio and the local moderator client’s audio to remote audio video clients 418.” *Id.* at 8:60-9:3.

56. The patent specification further describes the process of combining audio from the moderator with audio from remote clients, explaining the “remote audio video clients 418 audio is mixed with the local moderator client audio 401 in preparation for compression by the VoIP encoder 522.” *Id.* at 9:17-20. The mixed audio data is then “placed in virtual private network tunnel for transport to the voice over IP server 409 and onto the gateway for audio presentation to the PSTN[.]” *Id.* at 9:20-23.

57. The specification also explains that “local moderator client 401 receives audio from PSTN client 412 . . . and mixes audio previously decoded by the voice over IP decoder 524 with the audio from the remote client’s 418 for presentation to the local speaker 454.” *Id.* at 9:4-8.

B. The Challenged Claims

58. Claims 1-12 are the Challenged Claims of the ’332 patent. Claims 1 and 8 are the two sole independent claims. Independent claim 1 is directed to a method and recites:

A method for supporting a multi-participant audio/video conference call, the method comprising:

receiving first audio data from a Public Switched Telephone Network (PSTN) client;

receiving second audio data from a moderator;

receiving third audio data, video data, and collaboration data from at least one remote client through a first Virtual Private Network (VPN) tunnel;

mixing the first audio data from the PSTN client with the second audio data from the moderator into a first mixed audio data;

transmitting the first mixed audio data to the remote client through the first VPN tunnel;

mixing the second audio data from the moderator with the third audio data from the remote client into a second mixed audio data; and

transmitting the second mixed audio data to the PSTN client.

Id. at 9:52-10:5.

59. Claim 2 further adds the mixing of “the first audio data from the PSTN client with the third audio data from the remote client into a third mixed audio data.” *Id.* at 10:6-10.

60. Claims 3-7 add further known limitations, such as using IP packets and compressing audio data, to the method steps in claim 1, including “VoIP encoding the second mixed audio data before transmitting the second mixed audio data to the PSTN client.” *Id.* at 10:11-29.

61. Independent claim 8 is directed to a system that includes:

- “a first mixer” configured to mix audio data from a PSTN client with audio data from a moderator and collaboration data from the moderator into a mixed data;
- “a transport output” configured to transmit the first mixed data to at least one remote client through a first VPN tunnel;
- “a second mixer” configured to mix audio data from the moderator with audio data from the remote client commented to the Internet into a first mixed audio data; and
- “a VoIP encoder” to encode the first mixed audio data into a compressed audio data compressed audio data;
- The transport output further configured to transmit the compressed audio data to the PSTN client through the first VPN tunnel.

Id. at 10:30-46.

62. Claims 9-12 add an additional mixer, a speaker, an encoder and decoder, and an audio decompressor to the claimed system. *Id.* at 10:47-63.

VII. SCOPE AND CONTENT OF THE PRIOR ART

A. Audio, Video, and Collaboration Data Conferencing Technology

63. The prior art in the field of multi-party conferencing is very extensive. Prior to March 10, 2003 – the earliest claimed priority date for the ’332 patent – multiparty conferencing, including mixing and transmitting audio, video, and collaboration data over hybrid networks, was already well known.

64. Multiparty audio telephone conferencing is an age-old feature from the days of circuit-switched analog telephony. For example, U.S. Patent No. 3,060,265 (the “’265 patent”), titled “Conference Call Circuit” was issued on

October 23, 1962, and describes an invention that “relates to telephone systems and more particularly to conference call circuits for use in such systems.” ’265 patent (Ex. 1012) at 1:11-13. These conference systems would have conference participants’ calls delivered to a “conference bridge” device that would mix all the audio and then deliver the mixed audio back to all the participants.

65. For example, U.S. Patent No. 4,139,731 (the “’731 patent”), titled “Telephone conference system with active analog conference” was issued on February 13, 1979, and discloses that “[a] telephone conference system includes a conference bridge for coupling multiple conference ports associated with corresponding conference stations. The bridge includes means for summing incoming signals from all the conference ports and supplying the summed signals to all the conference ports.” ’731 patent (Ex. 1013) at Abstract.

66. Video conferencing systems were also well known before the ’332 patent, having existed for decades prior. For example, U.S. Patent No. 3,601,530 (the “’530 patent”), titled “Video conference system using voice- switched cameras,” was issued on August 24, 1971, describes “vi[su]al telephone systems . . . for conference connecting two or more groups of remote conferees in a manner which approaches a true face-to-face conference situation.” ’530 patent (Ex. 1014) at 1:6-10. By using cameras at two locations, “a video conference can be provided for two groups of remotely located conferees” such that the function “approaches a

true face-to-face conference situation,” with the benefit of “greatly reducing travel” cost. *Id.* at 1:6-60.

67. The conferencing art that existed prior to the ’332 patent also recognized the benefit of facilitating exchange of collaboration data such as written or graphic material. ’530 patent (Ex. 1014) at Abstract; *see also* U.S. Patent No. 6,237,025 (the “’025 patent”) (Ex. 1018) at 1:28-47 (“In addition to spoken words, demonstrative gestures and behavioral cues, collaboration often involves the sharing of visual information—e.g., printed material such as articles, drawings, photographs, charts and graphs, as well as video tapes and computer-based animations, visualizations and other displays—in such a way that the participants can collectively and interactively examine, discuss, annotate and revise the information. This combination of spoken words, gestures, visual cues and interactive data sharing significantly enhances the effectiveness of collaboration in a variety of contexts.”); *see also* Elliott (Ex. 1007) at 1:40-44, 5:49-56, 6:53-54.

68. It was also known in the art that participants in a conferencing system can include a leader (which may be referred to as a moderator) that can “a substantial degree of control over the conference...analogous to that appropriated by a leader in a true face-to-face conference situation” so as to more effectively conduct the conference. ’530 patent (Ex. 1014) at 2:34-40, Abstract, 17:20-41 (claims 7-11), 18:30-49 (claims 21-24); *see also* Knappe (Ex. 1006) at 8:52-56.

69. There were also many well-known methods used for initiating a conference call. For example, U.S. Patent No. 4,796,293 (the “’293 patent”), titled “Enhanced dedicated teleconferencing system” was issued on January 3, 1989, and explains that participants may be pre-selected, and “[a]ll conferees can be simultaneously dialed out to by the bridge.” ’293 patent (Ex. 1015) at 5:40-51. In another example, a four-digit code number could be specified, and “[i]n this type of conference, the conferees dial into [] the conference bridge and enter the specified four-digit touch tone code when prompted by a synthesized voice recording which the bridge generates. The touch tone code is used to instruct the bridge as to which conference this heretofore unidentified caller is to be connected.” *Id.* at 5:64-6:4.

70. It was also known that a telephone user may use features built into a PBX system to initiate a conference call. For example, it was known that a telephone user can call someone, place that callee on hold, dial another user, then join the two callees together for a three-way conference. Additional participants may be added in a similar way, up to the limit of the system. I personally, before the earliest claimed priority date of the ’332 patent, worked on the NBX 100 system, which had this method of adding a callee to a conference call.

B. Evolution of the Telephone System from Analog to Digital

71. One important factor in unleashing the expansion of a myriad of features in the telephone network was the migration from analog circuits to digital circuits and the accompanying digitization of the signaling network.

72. Through the conversion of the analog circuit network to a digital circuit network, the conference bridge remained a complex device that remained centrally located in the network. However, the conversion to digital allowed much of the functionality of a conference bridge to be implemented in software, and allowed the ability to develop more advanced features more rapidly than in an analog world. The telephony network further evolved before the earliest claimed priority date of the '332 patent, as the implementation of IP-based telephony to the endpoint began to be commercialized and as network bandwidth to the residence increased dramatically with the rollout and deployment of cable modems.

C. Circuit-Switched Networks

73. In a circuit-switched network, the entire path through a network is determined and reserved prior to any communications occurring on the circuit, and the path remains fixed for the duration of any communications, until the circuit is closed or “torn down.” A circuit can be thought of as a continuous wire between two parties through which data is transmitted. In fact, in the early days of the

telephone system, a series of electrical switches and interconnecting wires did create a continuous wired connection between parties.

74. As the network evolved to a digital network, the concept of a circuit remained, but no longer were entire wires dedicated to a single circuit. Digital representation of information allowed wires to be shared by sending data at regular intervals interleaved with data from other users using the notion of a regular time slot allocated to different “virtual circuits.” For example, 24 voice circuits are bundled into a standard T-1 digital trunk line, where a byte of data from each circuit is bundled together in a frame, and 8,000 frames are transmitted each second. This frame rate and sample size defines the standard unit of carriage in a digital telephone network, a channel with a bit rate of 64 kilobits per second. A specific slot in a frame on a trunk represents a single circuit, for example, the 5th slot in every frame carries data for a single circuit. A route for a virtual circuit still must be determined and reserved ahead of time prior to any communications occurring, and this involved determining free time slots on every trunk between the endpoints. The time slot allocated to the virtual circuit may also change from trunk to trunk, but overall the allocated time slots on the particular trunks remain fixed until the virtual circuit is torn down.

75. Once the route is determined and the virtual circuit is set up, data sent on the virtual circuit is transmitted through the various reserved time slots on the

trunks and arrives at the far end of the virtual circuit. The data need no identifiers as to which virtual circuit they are assigned. The virtual circuit is known by the network based on the appropriate time slot assigned to the virtual circuit for each trunk.

76. The most common and widely deployed example of a circuit-switched network is the public switched telephone network (PSTN).

D. Packet-Switched Networks

77. Packet-switching was independently proposed by Paul Baran and Donald Davies in the 1960's. Packet-switching was a different concept than traditional circuit-switching. In both analog and digital circuit switching, the circuit defined the communications endpoints and the entire path through the network is set up prior to communications occurring. The circuit's path is fixed throughout the duration of the call.

78. In packet switching, the data to be communicated is broken up into chunks, each chunk being a "packet," with each packet bearing a packet header that includes the address of the source and the address of the destination. As the packet traverses the network, each node in the network (which is connected to multiple links to other nodes) examines the destination address and makes a local independent decision on the best next hop in the network. This is known as the "routing" step, and the sending of the packet down the selected link to the next hop

is known as “forwarding.”¹ These nodes are known as “routers.” The routers receive updates from neighbors about the changing conditions and loads within the network, and update the data structures, known as “routing tables” which the router uses to determine the next hop. This allows packets from a particular source to a particular destination to travel different paths through the network, based on changing conditions in the network. While some people take this to mean that every packet travels a different path, this is not true – network paths *may* change, but that does not mean they change continuously. Network routing path updates are typically not a frequent occurrence compared the interval between packets in a data stream.

79. The principles set forth by Baran and Davies were adopted in the ARPANET, a project funded by the U.S. government in the late 1960’s through the Advanced Projects Research Agency (ARPA). The contract to build the ARPANET was awarded to Bolt, Beranek and Newman (BBN) in 1969. I later joined BBN in 2004 and, in my roughly five years working at BBN, discussed project ideas with several of the engineers that worked on the original ARPANET and later phases of the project.

¹ Often the combined act of making the decision on where to forward the packet and forwarding the packet are collectively referred to as “routing.”

80. BBN designed and built the Interface Message Processors (IMP) that performed the similar function in the ARPANET that today's routers in a data network perform.

1. Internet Protocol

81. A specification of the Internet Protocol titled "DoD Standard Internet Protocol," was prepared for the Defense Advanced Research Projects Agency (DARPA) in 1980, long before the patent at issue. RFC 760 (Ex. 1028) at Cover; *see also* Declaration of Ingrid Hsieh-Yee, Ph.D. ("Yee Decl.") (Ex. 1011) ¶¶ 16, 23-28. RFC 760 specified the length of an address to be 32 bits, or four bytes. Initially, the first byte specified the network number and the remaining three bytes were the local address within the network. This provided for only 256 distinct networks, and this method of addressing networks was later modified several times. RFC 760 also specified a 20-byte header for every packet, which included a source address, a destination address, and other fields. Thus, every IP packet contains a destination IP address in the header, which is used by the routers to determine the best next hop as the packet travels through the network.

82. RFC 760 was later obsoleted by RFC 791 in 1981, titled "Internet Protocol – DARPA Internet Program – Protocol Specification" and is one of the few documents made a formal Internet Standard as STD 5. Addresses remained 4 bytes (as IPv4 addresses remain today), but RFC 791 changed the concept of

network number to be more flexible by introducing the notion of Class A, B, and C network numbers. RFC 791 (Ex. 1029) at 7; *see also* Yee Decl. (Ex. 1011) ¶¶ 16, 23, 29-32. Later updates removed this notion to provide classless address space.

2. Unicast and Multicast

83. The addresses in IP initially identified individual host computers. RFC 791 at 1. These are known as “unicast” addresses as they identify a single host.

84. Several later RFCs introduced and refined the notion of a “multicast” address. RFC 966 was published in December 1985, and was followed by RFC 988 in July 1986, RFC 1054 in May 1988, and RFC 1112 in August 1989. RFC 1112 was made part of STD 5. Part of the IP address space was set aside for exclusive use as multicast address space. RFC 1112 (Ex. 1030) at 3; *see also* Yee Decl. (Ex. 1011) ¶¶ 16, 23, 33-36. The IP address range from 224.0.0.0 to 239.255.255.255 is reserved for use as multicast address space.

85. Unlike a unicast address which specifies a single host, a multicast address specifies a “host group,” a set of zero or more hosts identified by a single IP address in the multicast address range. RFC 1112 (Ex. 1030) at 1. The multicast-enabled network makes its best effort to deliver any packet sent to a multicast address to all the hosts in the host group. The membership of a host group is dynamic, with hosts able to join or leave at any time, and hosts may be

anywhere on the network. The routers in the network enabled to handle IP multicast addresses are known as multicast routers. Multicast routers maintain the notion of the group and where members of the group reside with respect to the various output ports of the router. The Internet Group Management Protocol (IGMP) is used by hosts to signal to their membership status (e.g., joining a group or leaving a group) to the hosts' neighboring multicast routers. *Id.* at 11. When there are multicast group members that have joined a particular group that can be reached at a particular output port of the multicast router, the router forwards multicast packets addressed to that group to all ports for which it knows there are group members reachable. When all the hosts reachable through a specific output port leave a multicast group, there is no need for a multicast router to forward packets to that output port any more.

86. Routers in the Internet are not typically enabled to route multicast packets. Packets sent into the Internet to an IP multicast address via the same means as unicast packets are sent will not, typically, be further routed through the Internet, as the Internet is not multicast-enabled. However, it was well known to create a multicast network “overlay” on top of the Internet, through the use of “tunnels,” as discussed below.

87. Multicast is not limited to an IP network. Other layers of the networking stack also incorporate this concept. For example, Ethernet, a link-layer

protocol, also incorporates the notion of multicast, but typically multicast delivery is limited to Ethernet segments as bounded by IP routers.

3. Tunneling

88. Tunneling refers to a networking technique whereby a packet is encapsulated in another network packet, typically so that it can traverse an underlying network that does not support the capability required by the encapsulated packet. IP in IP tunneling is described in RFC 1853 which was published in the mid-1990s. Tunneling can be performed in both directions. At the input end of the tunnel, packets are encapsulated, or wrapped, in a way where the source address of the encapsulating header is that of the encapsulating device and the destination address of the encapsulating header is the far end of the tunnel. At the far end of the tunnel, the packets are stripped of the encapsulating header, or unwrapped, and the original packet emerges for further actions such as routing to its destination using the original packet's header.

89. IP in IP tunneling is used in the MBone, a multicast network overlay on top of the Internet, and IPSec, a method for providing secure encrypted communications over the Internet (or any other unsecured or public network). Both these uses are discussed below.

4. **MBone**

90. To multicast on the Internet, host computers connected to the Internet are configured as multicast routers, and tunnels between the hosts acting as multicast routers are created. The tunnels are unicast connections between the multicast routers. *See de Moraes, Luis F.M., “The Internet Multicast from ITS: How it was Done and Implications for the Future,” IEEE Communications Magazine, January 1995 (Ex. 1022) at 6; see also Yee Decl. (Ex. 1011) ¶¶ 16, 23, 70-78.* Thus, the connected hosts operating as multicast routers are performing the multicast routing. This is known as an “overlay” network. Such an overlay network was implemented over the Internet in the mid 1990’s, and was known as the “MBone,” short for multicast backbone.

91. To participate in multicast using the MBone, a regular host on the Internet sends multicast group join requests to a multicast router, and then the multicast router will forward multicast packets to the host. Similarly, to multicast to a group of hosts, a host sends an IP multicast packet to a multicast router with the destination address being the address of the group of hosts.

92. When I was at MIT in the 1990’s, our research group implemented a multicast router as part of the MBone using a computer in our group. We participated in various events that were multicast on the MBone, including IETF meetings.

5. IPsec

93. RFC 2401, titled “Security Architecture for the Internet Protocol” was published in the late 1990s. *See* RFC 2401 (Ex. 1031); *see also* Yee Decl. (Ex. 1011) ¶¶ 16, 23, 37-40. RFC 2401 specifies the architecture of IPsec compliant systems to provide various security services for traffic at the IP layer, in both the IPv4 and IPv6 environments. RFC 2401 describes the components and how they fit together with each other and into the IP environment. It also describes the security services offered by the IPsec protocols, and how these services can be employed in the IP environment. RFC 2401 notes that “[s]ubsequent documents will address additional architectural details of a more advanced nature, e.g., ... more complete support for IP multicast.” RFC 2401 (Ex. 1031) at 3. The “subsequent documents” were published contemporaneously with RFC 2401. RFC 2401 discloses two security protocols, Authentication Header (AH) and Encapsulating Security Payload (ESP). AH is further described in RFC 2402, titled “IP Authentication Header,” and ESP is further described in RFC 2406, titled “IP Encapsulating Security Payload (ESP).” *See* RFC 2406 (Ex. 1032); *see also* Yee Decl. (Ex. 1011) ¶¶ 16, 23, 41-44.

94. ESP may be operated in a “tunnel mode” in which the original packet is encrypted, and the encrypted packet is encapsulated in a new IP header which specifies the endpoints of the tunnel. While host computers may implement ESP in

tunnel mode, security gateway appliances must use the tunnel mode. RFC 2406 (Ex. 1032) at 9.

95. A “Security Association” (SA) is a “simplex ‘connection’ that affords security services to the traffic carried by it.” RFC 2401 (Ex. 1031) at 8. Security services are afforded to an SA by the use of AH, or ESP, but not both. If both AH and ESP are used, then two (or more) SAs are created. To secure typical, bi-directional communication, two Security Associations (one in each direction) are required. Thus, when IPsec is used in ESP tunnel mode to secure communications between two devices, such as between a pair of security gateway appliances, a pair of security associations is required (one in each direction). Two simplex (unidirectional) tunnels are established, and together the pair of tunnels form a full-duplex bi-directional encrypted tunnel.

6. Virtual Private Network

96. A Virtual Private Network (VPN) is “defined loosely as a network in which customer connectivity amongst multiple sites is deployed on a shared infrastructure with the same access or security policies as a private network.” *See* VPN Textbook (Ex. 1010) at 115; *see also* Yee Decl. (Ex. 1011) ¶¶ 16, 23, 95-111. VPNs were known to be used to securely exchange data between different locations over the Internet. VPN Textbook at 118-119; *see also* U.S. Patent No. 6,980,526 (the “526 patent”) (Ex. 1020) at 2:65-3:3, 5:52-55. It was also known

that VPNs may be implemented over the public Internet using IPsec tunnels between sites.

97. It was well known that for companies with multiple sites, the simpler implementation is to deploy a VPN network by using a central router that acts as a hub for other remote site routers. This implementation is commonly referred to as a “hub-and-spoke” or a “star” topology. *See* VPN Textbook (Ex. 1010) at 129; Computer Desktop Encyclopedia (Ex. 1027) at 438; *see also* Yee Decl. (Ex. 1011) ¶¶ 16, 23, 79-94. In contrast, a dense mesh topology of VPN connections between all sites becomes complex to set up and manage, especially as a company expands to add new satellite offices. Using a hub-and-spoke topology for VPN connections makes securely connecting the network of a new satellite office a much simpler process to manage. *See* VPN Textbook (Ex. 1010) at 129.

7. Voice over IP (VoIP)

98. Voice over IP (VoIP), refers to voice data (e.g., audio) encapsulated in an IP packet for transmission across an IP network, such as the Internet.

99. When a stream of VoIP packets traverses a network, the interaction with network switches, routers, and other cross-traffic causes the packet latency for each packet in the stream to vary slightly, an effect known as “packet jitter.” Because of packet jitter, the destination receiving node must be coded to deal with the irregular and unpredictable arrival times of the

VoIP packets. Typically, packet data are put into the end of a buffer upon arrival, and voice reconstruction occurs with data at the front of a buffer. If the buffer and startup delays are sized appropriately, the buffer is unlikely to be empty when the next byte of voice data is needed to reconstruct the ongoing audio stream. The buffer could be empty because of excessive delays for a particular set of packets in the network (the data didn't reach the destination by the time it was needed), or a packet simply could have become dropped in the network (often due to congestion in a network). Audio reconstruction algorithms can "conceal" such lost data to an extent, but if the losses become high then users will notice.

100. It was well known to utilize VoIP for conference communications. *See, e.g.*, U.S. Patent No. 6,850,496 (Ex. 1024) at 1:23-28 ("Packet-switched networks can also carry real time voice data, and therefore, with proper configuration, conference calls. Voice over IP (VoIP) . . . use a packet-switched network for transport of voice data."); Cisco Packetized Voice & Data Integration, Robert Caputo (Ex. 1021) at 23; *see also* Yee Decl. (Ex. 1011) ¶¶ 16, 23, 53-69. I note that I personally, before the earliest claimed priority date of the '332 patent, designed a packet voice/audio reconstruction algorithm and coded a prototype of the algorithm for the NBX 100 telephones. Also, before the earliest claimed priority date of the '332 patent, when I was employed at Empirix, I conceived of

and designed a new product for testing the quality of VoIP calls. The product was known as a cloud emulator, which varied the jitter of packets traversing the tool as well as the packet loss, allowing testing of the voice reconstruction algorithms in gateways. We sold the product to manufacturers of large VoIP gateways.

101. Also, it was well known that because many of the standard codecs do not specify security features, VoIP calls (including the calls into a conference bridge) are vulnerable to being reconstructed by anyone that has access to the packets carrying the audio data, such as when the VoIP data stream is sent unprotected over the Internet. It was well known for employees at large companies, with multiple sites, to utilize video conferencing as a way to avoid the expense of holding in-person meetings. However, because private leased data lines between a company's sites are much more expensive than direct connections to the Internet at each site, companies frequently utilized VPNs for encrypting site-to-site communications over the Internet, thereby providing the security of a private leased line without the ongoing expense of one. *See, e.g.*, U.S. Patent No. 6,782,413 (Ex. 1019) at 2:66-3:1, 4:23-34, 6:36-54; '526 patent (Ex. 1020) at 2:65- 3:3, 5:52-55. It was also known to utilize VPN tunneling for voice, video, and collaboration data during a conference. *See* U.S. Patent Publication No. 2004/0261116 (Ex. 1023) at ¶¶ [1412]-[1414], [1451]-[1453]; Elliott (Ex. 1007) at 2:66-3:1, 4:29-34, 6:53-54.

8. Audio and Video Compression

102. Audio and video compression were well known concepts, where “codecs” specify algorithms for the coding (the “co” of codec) and decoding (the “dec” of codec) of audio and video data. *See, e.g.,* Cisco Packetized Voice & Data Integration, Robert Caputo (Ex. 1021) at 12-13; U.S. Patent No. 6,850,496 (Ex. 1024) at 3:56-67. Various codecs provide different levels of compression and different levels of perceived quality. G.711 is an early standard for an audio codec, and was available in 1972. G.711 is used in H.323. G.711 specifies that audio be sampled at an 8 kHz rate with 8 bit samples, and thus the bitrate of the digitized encoded audio is 64 kilobits/second. As discussed above, the architecture of the digital PSTN is based around this fundamental base rate of digitized speech. Other voice codecs were known to further reduce the bitrate, and lower bitrate codecs were frequently used in networks to conserve bandwidth.

103. The ’332 patent incorporates by reference the entire H.261 and H.263 specifications which specify video codecs. *See* ’332 patent (Ex. 1001) at 4:15-32. Video codecs specify the algorithms by which video is compressed, and include references to audio encoding standards, since audio is part of the video. Another well-known video codec standard is MPEG-2. Because most video contains moving objects between frames (as opposed to sudden cuts between scenes which occurs less frequently) MPEG uses the concept of motion compensation to increase

the compression. During compression, objects in a current frame can be encoded based on a pointer to the same object in an adjacent frame (either before or after the current frame) without having to re-encode the entire visual representation of that same object in the current frame. Small differences in the object also can be encoded. Only at specified intervals is the entire frame encoded based only on the contents of the current frame. Most video codecs utilize this concept.

9. A Gateway between PSTN and IP Networks

104. As business and consumer VoIP products and services began to be introduced in the late 1990's and early 2000's, it was recognized as important for VoIP devices to be able to communicate with traditional PSTN telephones, not just with other VoIP telephones. For this reason, gateways were widely used to facilitate communications between the VoIP devices and the PSTN telephones. *See, e.g.,* Knappe (Ex. 1006) at FIG. 1. As I mentioned above, the NBX 100 system had a gateway to the PSTN as part of the NBX 100 system itself. Initially, we shipped only a gateway that connected to a conventional analog telephone line, for bridging between our packet voice system and an analog line. In subsequent releases, but prior to the earliest claimed priority date of the '332 patent, we released a line card that interfaced to a digital T1 line.

105. Commercial VoIP providers, such as Vonage, operated larger gateways in the network itself to interconnect its voice customers to the PSTN.

These types of carriers were referred to as “over the top” carriers as their VoIP traffic traversed the customers’ existing Internet connection to reach the gateways of the service provider.

106. Large gateways often were partitioned into parts, with one component for handling the audio conversion between networks, often called the “media gateway,” and another component for handling the signaling conversion between networks, often called the “signaling gateway.”

10. **VoIP for Conferencing**

107. As VoIP’s popularity increased, it was well known to implement VoIP-based conference systems using software-based digital network conference bridges, including replacing the digital circuit interfaces with interfaces to the data network and the software to decode various standard codecs to a standard internal representation

108. It was also known to implement conference bridges towards the edges of a network and even in the endpoints. When conference calls were mixed locally in the endpoint, audio data from the participant was delivered to all the participating endpoints. *See, e.g., Drell (Ex. 1008) at 4:21-31.*

109. From 1997-1999, I worked at a company that designed and shipped a business telephone system private branch exchange (PBX) that worked over a company’s existing data network. The product, called the NBX 100, was a full

featured business PBX, where every device in the system was connected to the data network via an Ethernet connection, which is almost universally the type of connection a computer uses to connect to the network. A gateway card interconnected the system to the PSTN for inbound and outbound calls. Every telephone and gateway in the system utilized a relatively inexpensive digital signal processor (DSP). Every telephone and gateway in the system was capable of performing a four-party conference call through mixing the audio received from each of the other three callers (which could be a mix of internal callers and external callers via a gateway) using software running on each telephone's DSP. There was no centralized conference bridge in the system. For larger conference calls, that is over four parties, an external conference bridge was used (although a six-party conference could be formed when two different telephones which were each part of a three-party conference joined together in a conference).

E. Conferencing Standards

110. The H.323 standard was first published by the International Telecommunications Union (ITU) in 1996, and focused on video conferencing over a packet network. H.323 encompassed a set of standards for providing audio and video conferences on a packet network, including signaling to set up a conference as well as the compression and transport of media including audio and video. *See, e.g., Cisco Packetized Voice & Data Integration, Robert Caputo (Ex. 1021) at 23.*

111. H.323 is regarded by many as the standard for VoIP communication, and also specifies compression protocols. *See, e.g.*, U.S. Patent No. 6,553,515 (Ex. 1025) at 4:35-56 (“H.323 is adapted to support many aspects of multimedia...[and] generally is regarded as the standard for VoIP...This standard also specifies lowest-common denominator compression protocols, such as G.711...G.723 and G.729.”). H.323 also provides for initial call setup, including determining support for full-duplex. *See, e.g., Id.* at 4:61-67 (“H.323 requires that Internet telephones also adhere to H.245. . . . During call initiation, the devices use H.245 to determine proper compression protocols and also whether the phones support full-duplex or half-duplex transmission.”).

112. The H.323 standard recommended the use of collaboration data per the T.120 standard, which was established by the ITU in 1996. *See, e.g.*, A Primer on the T.120 Series Standard (Ex. 1016); *see also* Yee Decl. (Ex. 1011) ¶¶ 16, 23, 45-48; T.120 Standard (Ex. 1017) at 8; *see also* Yee Decl. (Ex. 1011) ¶¶ 16, 23, 49-52; ’332 patent (Ex. 1001) at 3:10-12. The T.120 standard contains a series of application protocols (e.g., T.122, T.123, T.124, and T.125) for collaboration data such as multipoint file transfer, still image viewing and annotation, shared white board, and facsimile. *See* A Primer on the T.120 Series Standard (Ex. 1016).

VIII. INVALIDITY ANALYSIS

113. As noted previously, I have been asked to provide my opinion as to whether the Challenged Claims (claims 1-12) of the '332 patent would have been obvious in view of the prior art. The discussion below provides a detailed analysis of how the prior art references I reviewed teach the limitations of the Challenged Claims of the '332 patent and render the Challenged Claims obvious to a POSITA.

114. As part of my analysis, I have considered the scope and content of the prior art and any potential differences between the claimed subject matter and the prior art. I conducted my analysis as of the earliest claimed priority date of the '332 patent: March 10, 2003. I have also considered the level of ordinary skill in the pertinent art as of that date.

115. I describe in detail below the scope and content of the prior art, as well as any differences between the claimed subject matter and the prior art, on a limitation-by-limitation basis for each Challenged Claim of the '332 patent. This analysis supports my opinion that the differences between the claims of the '332 patent and the prior art discussed herein are such that the subject matter as a whole would have been obvious as of the earliest possible priority date of the '332 patent to a person having ordinary skill in the art to which the subject matter pertains.

116. I will now describe, in the grounds below, on a limitation-by-limitation basis, how the prior art teaches and/or renders obvious all limitations of claims 1-12 of the '332 patent.

A. Claims 1-2 and 4-5 and 7 Are Obvious Over Knappe in Combination with Elliott and the VPN Textbook

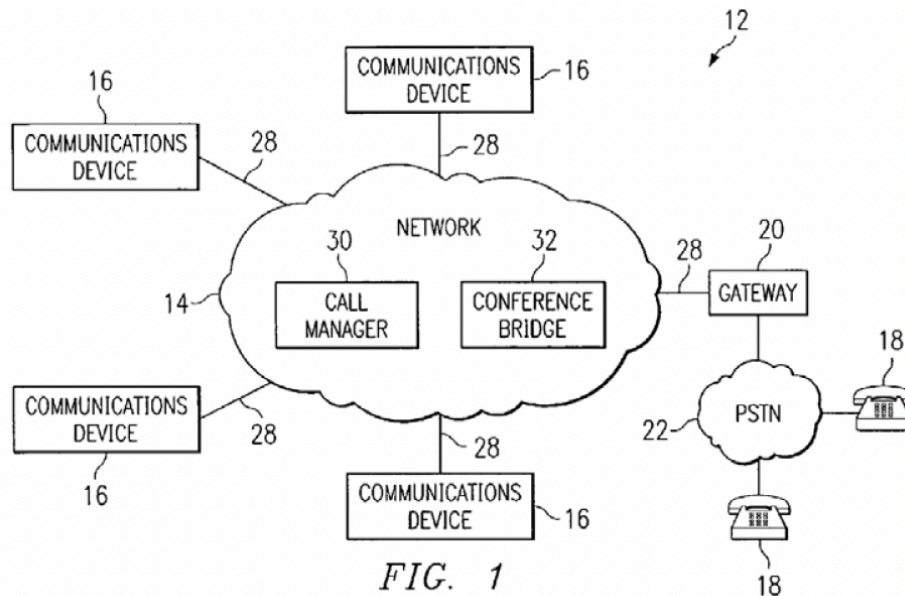
1. The Prior Art

a) Summary of Knappe

117. United States Patent No. 7,180,997 ("Knappe") entitled "Method and System for Improving the Intelligibility of a Moderator During a Multiparty Communication Session" was filed September 6, 2002. *See* Ex. 1006.

118. Like the '332 patent, Knappe discloses a multiparty conferencing system over a hybrid network that includes a network "connecting a plurality of communication devices 16 to each other and to standard analog telephones 18 through a gateway 20 and the public switched telephone network." *Id.* at 3:41-45; *see also id.* at Abstract, 1:8-11, 8:46-10:50, FIGS. 1-8.

119. Figure 1 of Knappe illustrates the conferencing system, showing multiple conference endpoints, including communication devices 16 and standard analog PSTN telephones 18:



Id. at FIG. 1; *see also, id.* at 3:41-4:28. The Knappe system is a “distributed system transmitting audio, video, voice, data, and other suitable types of real-time and/or non real-time traffic between source and destination points.” *Id.* at 3:26-29. As shown in Figure 1, participants in the Knappe conferencing system include “communication devices 16 [that] comprise real-time applications that play traffic as it is received.” *Id.* at 4:5-9.

120. Knappe explains that network 14 can be “the Internet, a wide area network (WAN), a local area network (LAN), or other suitable packet-switched network.” *Id.* at 3:50-52. In the Internet embodiment, Knappe explains that “the network 14 transmits information in Internet Protocol (IP) packets[,]” with telephony voice information “transmitted in the Voice over IP (VoIP) format.” *Id.* at 3:53-56.

121. Knappe further explains:

codec (coder/decoder) converts audio, video or other suitable signals generated by users from analog signals into digital form. The digital encoded data is encapsulated into IP or other suitable packets for transmission over the network 14. IP packets received from the network 14 are converted back into analog signals and played to the user.

Id. at 4:5-14. And “at least one participant may be given the priority of moderator for a particular conference call.” *Id.* at 8:52-56.

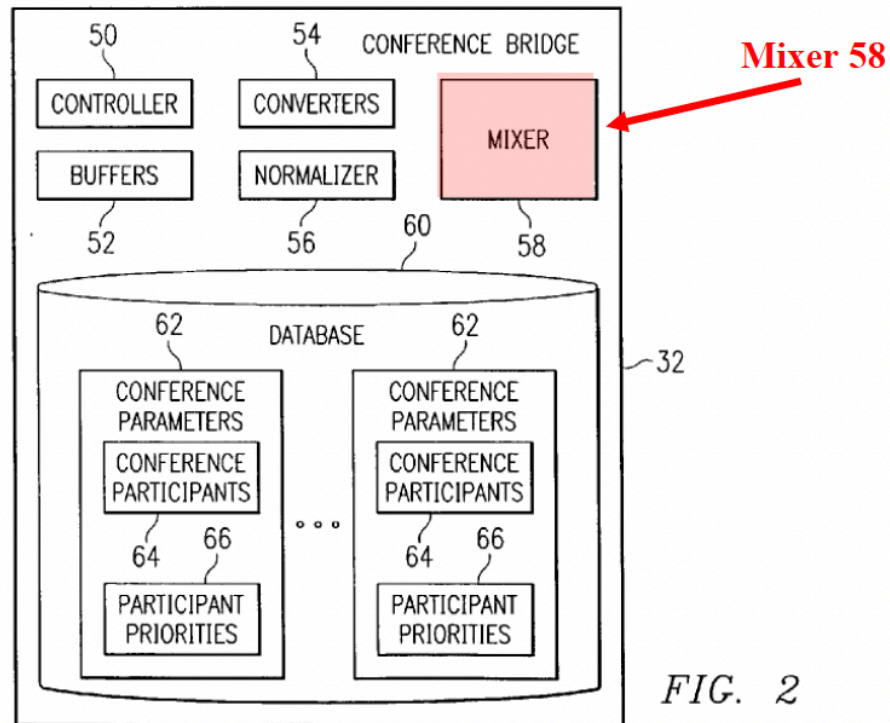
122. The standard analog telephones 18 participating in the conferencing system communicate standard telephony signals through PSTN 22 to the gateway 20.” *Id.* at 4:19-21. Knappe explains that:

[a]t the gateway 20, the signals are converted to IP packets in the VoIP format. Similarly, VoIP packets received from the network 14 are converted into standard telephony signals for delivery to the destination telephone 18 through PSTN 22. The gateway 20 also translates between the network call control system and the Signaling System 7 (SS7) protocol and/or other signaling protocols used in PSTN.

Id. at 4:21-28.

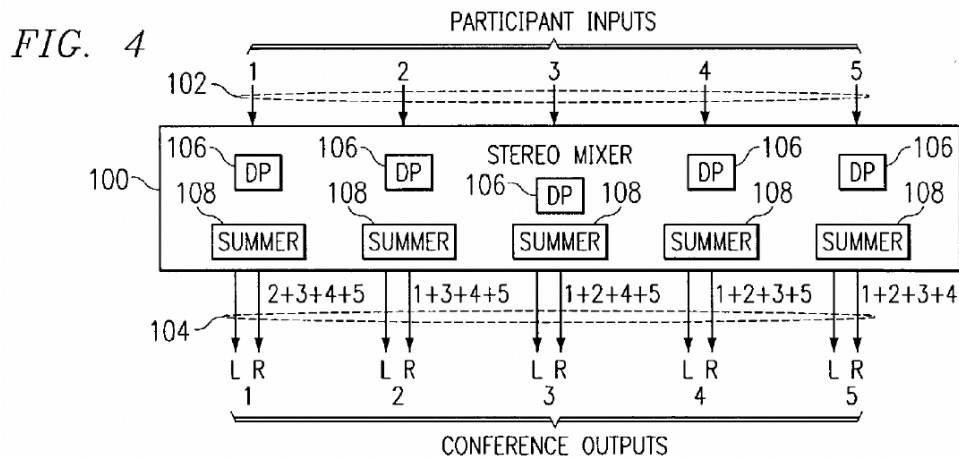
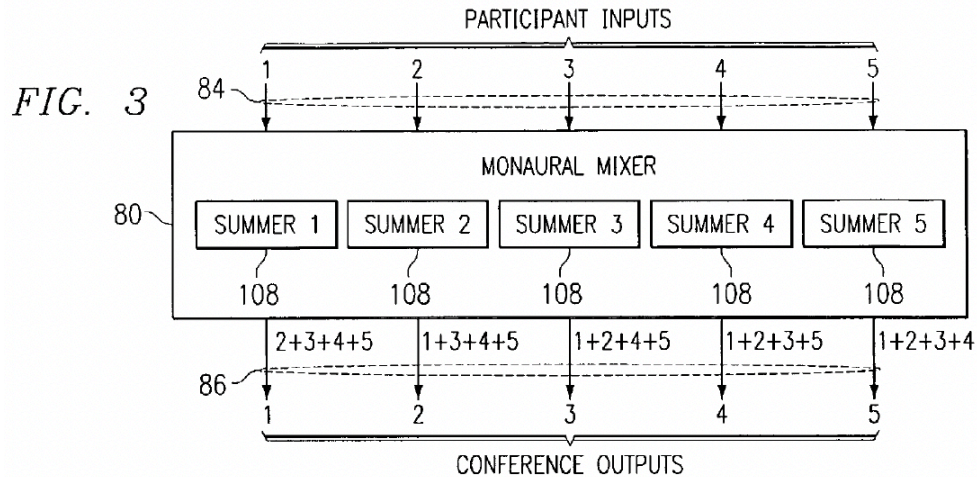
123. The network communication devices 16 and analog telephone 18 “participants are connected and stream media through the conference bridge 32.” *Id.*, 5:32-33. “The conference bridge 32 provides conference call and other suitable audio, video, and/or real-time multi-party communication sessions between communication devices 16.” *Id.* at 4:58-60. As illustrated in Figure 2 below, Knappe’s conference bridge 32 employs mixer 58, which “includes a

plurality of summers or other suitable signal processing resources each operable to sum, add or otherwise combine a plurality of input streams into conference output streams for participants to a conference call.” *Id.* at 6:33-37.



Id. at FIG. 2 (annotated).

124. Figures 3-4 in Knappe show examples of mixers that “combine[] the streams in summers 82 to generate conference output streams 86 for each participant to a conference call.” *Id.* at 6:60-62 (monaural mixer); *see also id.* at 7:36-39 (stereo mixer).



Id. at FIGS. 3-4.

125. In the Knappe system the combined data is then transmitted to each of the other participants, with Knappe explaining that “[d]uring normal operation, each participant receives the audio input of each other participant.” *Id.* at 7:1-2; *see also, id.* at 7:2-15.

b) Summary of Elliott

126. United States Patent No. 6,690,654 (“Elliott”) entitled “Method and System for Multi-Media Collaboration Between Remote Parties” was filed October 1, 1998. *See* Ex. 1007.

127. Like the ’332 patent and Knappe, Elliott also discloses a conferencing system that allows for multi-media collaboration between multiple remote participants, including Internet-based and PSTN clients. *See id.* at Abstract & FIG. 1.

128. Elliott teaches conferencing features such as “collaborative Web browsing, audio conferencing, video conferencing, and application sharing.” *Id.* at 1:40-44.

129. Elliott also teaches the use of “a virtual private network (VPN) ... which uses IP or IP-tunneling” for the transmission of “voice, fax, data, video” and “[o]ther types of collaboration ... data” during the conference. *Id.* at 2:66-3:1, 4:29-34, 6:53-54.

130. As shown in Figure 1, the conferencing system in Elliott shows a “session server 12 [that] is a conference server that follows the ITU T.120 standard” which “contains a series of communication and application protocols and services that provide support for real-time, multi-point data communications.” *Id.* at FIG. 1, 5:49-54.

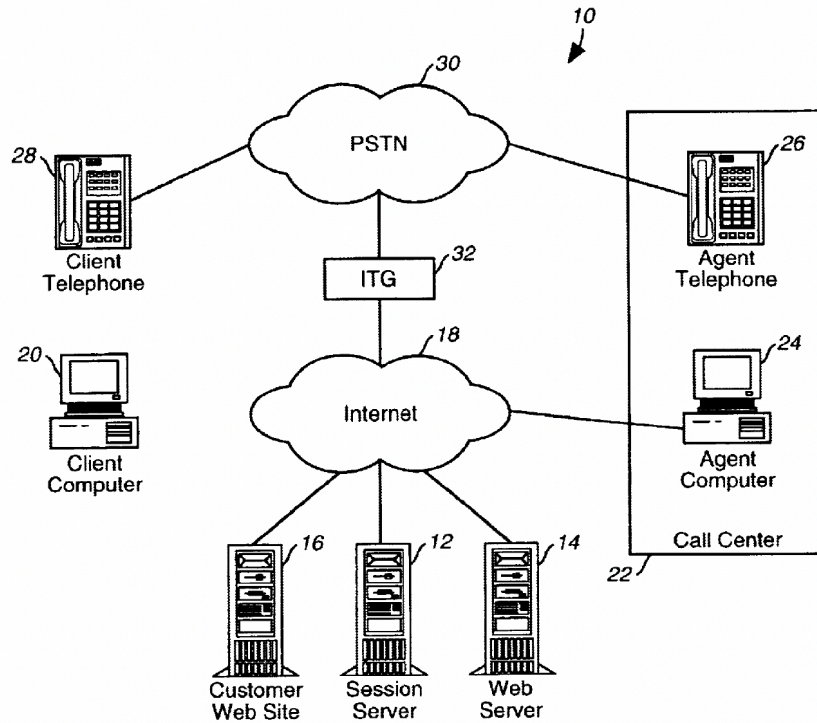


FIG. 1

131. Elliott teaches that “[n]etwork 18 can be either a public or a carrier supported virtual private network (VPN) or local area network (LAN) which uses IP or IP-tunneling. *Id.* at 4:28-32.

c) Summary VPN Textbook

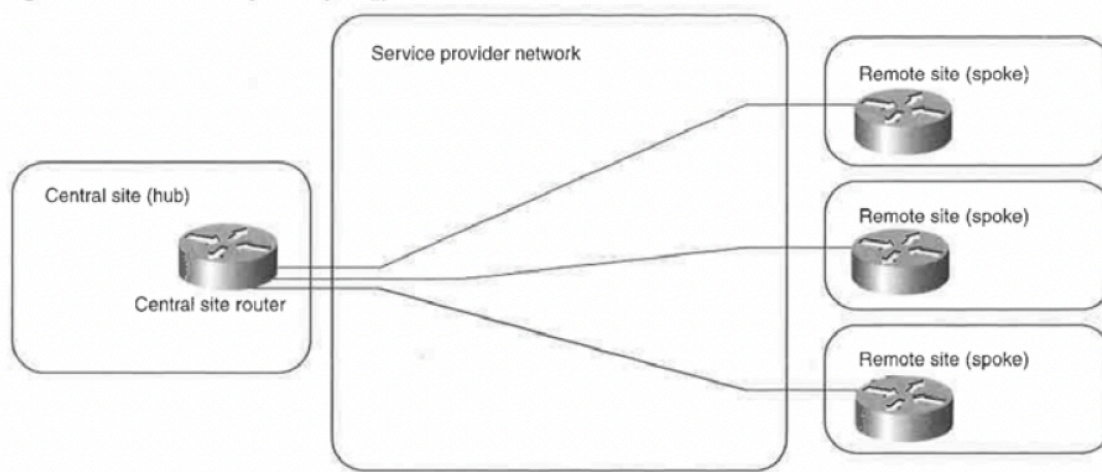
132. Guichard *et al.*, “MPLS and VPN Architectures, A practical guide to understanding, designing and deploying MPLS and MPLS-enabled VPNs,” Cisco Press, 2001 (“VPN Textbook”) is a textbook that describes known implementations of VPNs that was published 2001. *See* Ex. 1010; Yee Decl. (Ex. 1011) ¶¶ 101-105, 111.

133. The VPN Textbook describes well-known topologies for implementing VPNs. The VPN Textbook describes that “[t]he most commonly

encountered topology is a hub-and-spoke topology, where a number of remote offices (spokes) are connected to a central site (hub).” *Id.* at 129.

134. Figure 7-10 of the VPN Textbook depicts the “hub-and-spoke” topology, which is also referred to as a “star” topology:

Figure 7-10 *Hub-and-spoke Topology*



Ex. 1010 at 130; *see also* Computer Desktop Encyclopedia (Ex. 1027) at 438.

135. The VPN Textbook describes this hub-and-spoke (aka star) topology as being the “most commonly encountered topology,” chosen often “for cost or complexity reasons.” Ex. 1010 at 129.

2. Motivation to Combine the Prior Art

a) Reasons to Combine Knappe and Elliott

136. A POSITA would have been motivated to combine the teachings of Knappe and Elliott to produce, among other benefits, the obvious, beneficial, and predictable results of using different types of collaboration data and VPN tunneling within Knappe’s conferencing system.

137. A POSITA, when considering the teachings of Knappe, also would have considered the teachings of Elliott since they are analogous prior art both pertaining to the same field of multiparty conferencing. *See, e.g.*, Knappe (Ex. 1006) at Abstract, 3:26-29; Elliott (Ex. 1007) at Abstract. Also, both Knappe and Elliott support conferencing across a hybrid network, with computer and PSTN telephone conference participants. *See, e.g.*, Knappe (Ex. 1006) at 3:41-49, FIG. 1; Elliott (Ex. 1007) at 5:35-39, 6:1-11, 8:16-39, FIG. 1.

138. A first predictable and beneficial result of combining Knappe and Elliott includes the sharing of collaboration data, as disclosed in Elliott, with the ways of conveying information among participants in a multiparty conference. As discussed above, Knappe teaches a communications system capable of “transmitting audio, video, voice, data and other suitable types of ... traffic between source and destination endpoints.” Knappe (Ex. 1006) at 3:26-29. To a POSITA, Knappe’s disclosure of transmitting “data” or other “traffic” would suggest the transmission of various types of data during the conference call to facilitate collaboration among the participants. A POSITA would have recognized that sharing collaboration data in Knappe is desirable since it beneficially provides additional ways to convey information among participants in a multiparty conference.

139. Notably, Elliott evidences this understanding and expressly teaches the sharing of collaboration data including “collaborative Web browsing, audio conferencing, video conferencing, and application sharing.” Elliott (Ex. 1007) at 1:43-44. And Elliott discloses connecting calls “over the public switched telephone network (PSTN) 30” using “session server 12 ... that follows the ITU T.120 standard” capable of “support[ing] real-time, multi-point data communications,” the server “able to set up to conduct video conferences, audio conferences, and data conferences with application sharing,” as well as “[c]ollaborative Web browsing.” *Id.* at 5:35-61; *see also* T.120 Primer (Ex. 1016); T.120 Standard (Ex. 1017); Ex. 1028 at 1-2, 199-126.

140. Exchanging collaboration data was standardized (e.g., under the T.120 standard) and VPN tunneling was conventional by the time of Knappe and Elliott, and before the earliest claimed priority of the '332 patent. The T.120 standard was established by the International Telecommunications Union (ITU) and contains a series of application protocols (e.g., T.122, T.123, T.124, and T.125), including standardizing the exchange of collaboration data. *See, e.g.*, T.120 Standard (Ex. 1016); '332 patent (Ex. 1001) at 3:10-12. Before the earliest possible priority date of the '332 patent, I personally configured, deployed, and utilized VPN technology for carrying VoIP traffic using the NBX 100 system, which was capable of

performing conference calls. *See also* U.S. Patent No. 6,782,413 (Ex. 1019) at 4:23-27, 6:36-41; Elliott (Ex. 1007) at 5:49-56.

141. It would have been obvious to a POSITA, in view of Elliott's teachings, to implement Knappe's transmission of various types of data to include collaboration data, including the "video conferences, audio conferences," "data conferences with application sharing," and "collaborate Web browsing," (Ex. 1007 at 5:49-57) described in Elliott because this type of collaboration data would allow for a greater variety of information to be conveyed among the participants, thereby improving the conferencing experience and enhancing the effectiveness of collaboration in a variety of contexts.

142. Indeed, other patents and printed publications confirm that it was known in the art at the time of the earliest possible priority date of the '332 patent that additional collaboration data improves the conferencing experience:

collaboration often involves the sharing of visual information—e.g., printed material such as articles, drawings, photographs, charts and graphs, as well as video-tapes and computer-based animations, visualizations and other displays—in such a way that the participants can collectively and interactively examine, discuss, annotate and revise the information. This combination of spoken words, gestures, visual cues and interactive data sharing **significantly enhances the effectiveness of collaboration in a variety of contexts**

U.S. Patent No. 6,237,025 (Ex. 1018) at 1:29-39 (emphasis added).

143. Indeed, sharing collaborative information such as written or graphic material during a video conference had been known in the conferencing art for

over 30 years prior to the priority date of the '332 patent. *See, e.g.*, U.S. Patent No. 3,601,530 (Ex. 1014) at Abstract, 17:20-41 (claims 7-11), 18:30-49 (claims 21-24).

144. Accordingly, a POSITA would have recognized that sharing collaboration data (e.g., web browsing data, application data, and data provided by the T.120 standard, as Elliott teaches) would be beneficial to Knappe's participants. Such a combination would have simply involved combining prior art elements—collaboration data taught by Elliott with the conferencing system of Knappe— according to known elements to yield predictable results, namely allowing participants in Knappe's system to share various types of data during a conference for collaboration purposes, including using standard collaboration data that was specifically designed for a multimedia conference in an environment such as that disclosed in Knappe. Thus, combining Knappe and Elliott involves ordinary use of common techniques to improve the similar conference systems of Knappe and Elliott.

145. A second predictable and beneficial result of combining Knappe and Elliott includes the use of a VPN, as disclosed in Elliott, to protect communications between participants in Knappe's conferencing system.

146. It was known in the art that data transmitted over public IP networks is vulnerable to eavesdropping and other problems. VPN Textbook (Ex. 1010) at

119 (stating that “more and more business-to-business traffic takes place over the Internet,” which is “riddled with security issues”).

147. It was known that a solution for securing communications over public networks was by using a VPN. *Id.* at 115 (“A Virtual Private Network (VPN) ... [provides] connectivity amongst multiple sites ... with the same access or security policies as a private network.”).

148. Thus, it would have been obvious to a POSITA to use VPNs for conferencing systems, as described in Elliott, and to beneficially use said security with Knappe’s multiparty conferencing communications over the Internet.

149. The obviousness of adding a VPN feature to Knappe’s conferencing system is further demonstrated by the work of others in the art making similar feature combinations as evidenced by other prior art patents. *See, e.g.*, U.S. Patent No. 6,980,526 (Ex. 1020) at 2:65-3:3, 5:52-55 (disclosing “need for a videoconferencing system ... for delivering secure ... services over an IP network” by using a “VPN ... over public networks ... which is encrypted to make it secure from outside access and interference”); U.S. Patent No. 6,782,413 (Ex. 1019) at 2:66-3:1, 4:23-27, 6:36-41 (describing a “conference bridge application” that transmits data “over public networks, using technologies such as ... [a] Virtual Private Network (“VPN”)).

150. Therefore, a POSITA would have found it obvious to combine Knappe's conferencing system with the VPN technology described in Elliott and the VPN Textbook (discussed below). *See* U.S. Patent Publication No. 2004/0261116 (Ex. 1023) at ¶¶ [1412], [1452]-[1453] (describing "VPN connectivity," including "hub-and-spoke or mixed topologies" in the context of "services [that] include Internet Access, ... chat and Video-conferencing etc.").

151. Specifically, Elliott teaches using a "virtual private network (VPN) ... which uses IP or IP-tunneling" to transmit "voice, fax, data, video" and "[o]ther types of collaboration ... data." Elliott (Ex. 1007) at 2:66-3:1, 4:29-34, 6:53-54. In view of Elliott's VPN teachings, it would have been obvious to a POSITA to use VPN tunneling when communicating audio, video, and collaboration data across Knappe's Internet network 14 to make the transmitted data secure without the need to lease private data lines and otherwise reduce costs for implanting a private network securely over the Internet. *See* U.S. Patent No. 6,980,526 (Ex. 1020) at 5:52-55 ("VPN refers to a network that is carried over public networks, but which is encrypted to make it secure from outside access and interference."); U.S. Patent No. 6,633,848 (Ex. 1026) at 44:10-12 ("[O]ffices [] create virtual private networks (VPNs) to save money on remote access and interoffice connectivity."); *see also* U.S. Patent No. 6,980,526 (Ex. 1020) at 5:51-52 ("a

Virtual Private Network (VPN) may be created between the [enterprise video] gateway 36 and [videoconferencing services] switch 12.”).

152. Indeed, the combination of Knappe and Elliott is merely the ordinary use of the common technique of VPN tunneling of conferencing data, as disclosed in Elliott, in the similar conferencing system of Knappe to yield predictable results, including securing transmitted data and reducing costs associated with leasing private data lines.

153. The combination of Knappe with the teachings of Elliott permits, but does not require, physical incorporation of elements from the conferencing systems of Elliott and Knappe.

b) Reasons to Combine Knappe and the VPN Textbook

154. A POSITA would have been motivated to combine the teachings of Knappe and the VPN Textbook to produce the obvious, beneficial, and predictable results of providing hub-and-spoke VPN topology when implementing VPN tunneling in the Knappe and Elliott combination, and to provide a cost effective and streamlined VPN implementation.

155. As discussed above, Knappe teaches a conferencing system capable of “transmitting audio, video, voice, data and other suitable types of ... traffic between source and destination endpoints,” and Elliott teaches VPN tunneling of conferencing data for a conferencing system like Knappe’s. Knappe (Ex. 1006) at

3:26-29; *see* Elliott (Ex. 1007) at 2:66-3:1, 4:29-34, 6:53-54. A POSITA would have been familiar with conventional VPN tunneling implementations well documented in readily available textbooks, such as the VPN Textbook.

156. The VPN Textbook describes various VPN topologies. VPN Textbook (Ex. 1010) at 116-133. A POSITA looking to implement Knappe's system with VPN tunneling would have looked to the VPN Textbook to consider various topologies, including the "most commonly encountered topology is a hub-and-spoke topology, where a number of remote offices (spokes) are connected to a central site (hub)." *Id.* at 129; Computer Desktop Encyclopedia (Ex. 1027) at 438.

157. Combining Knappe with the VPN Textbook would merely be a combination of prior art elements (e.g., VPN topology teachings of the VPN Textbook and Knappe's system) according to known methods to yield predictable results—securing traffic exchanged in Knappe's conferencing system. I personally researched and purchased VPN hardware routers for use with the NBX 100 system in the 1998-1999 time frame.

158. Thus, a POSITA would have been able to predict the result of combining a hub-and-spoke VPN topology, as described in the VPN Textbook, in Knappe's system, for example, as modified by Elliott (as described above), to communicate data between conferencing participants (e.g., remote clients) over secure VPN tunnels.

159. Furthermore, a POSITA would have been motivated to combine the teachings of Knappe, as modified by Elliott, with the teachings of the VPN Textbook for cost and complexity savings. VPN Textbook (Ex. 1010) at 129 (“there are many examples where the customer could benefit from a different topology but has nonetheless chosen the hub-and-spoke topology for cost or complexity reasons”).

160. A POSITA would have had a reasonable expectation of success in combining Knappe with the VPN Textbook’s hub-and-spoke topology because the hub-and-spoke topology was conventional and known to be applied in the conferencing context. For example, U.S. Patent No. 6,782,413, filed February 11, 2000, describes a “conference bridge application” that can transmit data “over public networks, using technologies such as ... Virtual Private Network (VPN) or any other conventional communication method.” Ex. 1019 at 4:23-27, 6:36-41; *see also* U.S. Patent Publication No. 2004/0261116 (Ex. 1023) at ¶¶ [1412], [1451]-[1453] (disclosing VPN connectivity, including hub-and-spoke or mixed topologies, for services including video-conferencing).

161. Modifying Knappe to accommodate teachings of the VPN Textbook would have been well within the level of ordinary skill in the art, as demonstrated by the numerous examples of systems, patents, and printed publications combining conferencing and VPN technologies. Indeed, I personally configured, deployed,

and used VPN technology for carrying VoIP traffic using the NBX 100 system, capable of performing conference calls, before the earliest claimed priority date of the '332 patent.

162. This combination permits, but does not require, physical incorporation of elements from the VPN Textbook into Knappe's system.

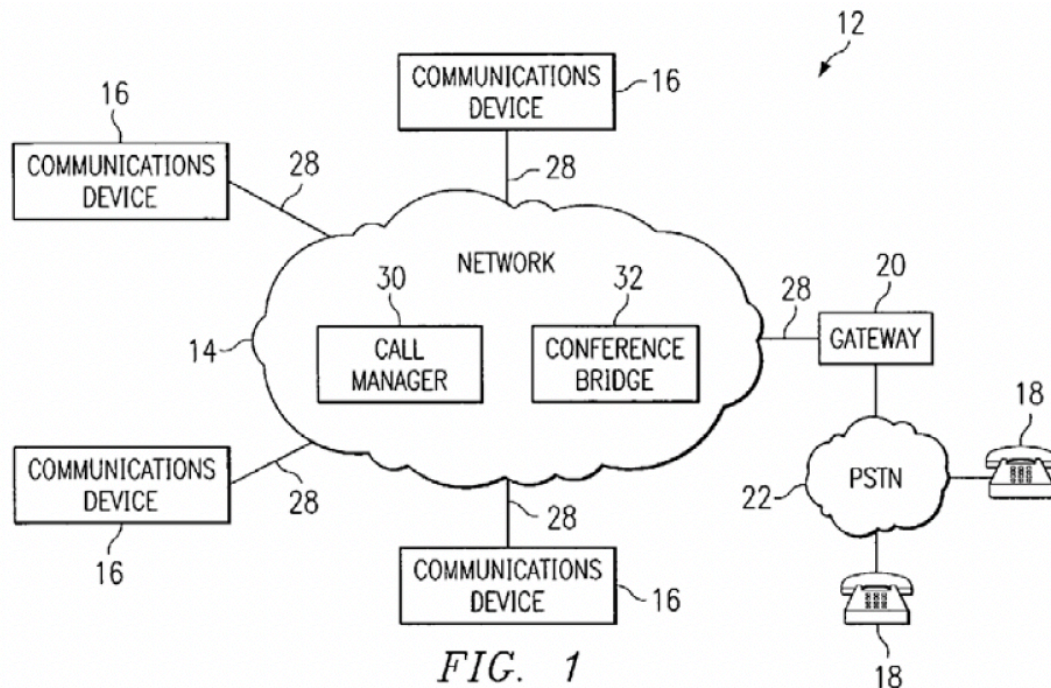
3. Claim 1

[1.P] A method for supporting a multi-participant audio/video conference call, the method comprising:

163. Knappe discloses "a method for supporting a multi-participant audio/video conference call" as required by the preamble of claim 1.

164. Specifically, Knappe discloses "a method and system for improving ... a multiparty communication session ... [that] allow[s] multiple people to participate in ... a 'conference call.'" Knappe (Ex. 1006) at 1:8-27, 8:46-10:50, FIG. 8. Knappe's system supports "transmitting audio, video, voice, data ... between source and destination endpoints ... to conduct multiple party telephone conference communication sessions." *Id.* at 3:24-31; *see also, id.* at 4:58-63.

165. As shown in Figure 1 of Knappe, Knappe's system can support multiple participants in an audio/video conference call:



Id. at FIG. 1. Specifically, Knappe explains that “conference bridge 32 provides conference call and other suitable audio, video, and/or real-time multi-party communication sessions between communication devices 16.” *Id.* at 4:58-61. Thus, a POSITA would understand Knappe to support a multi-participate audio/video conference call.

[1.1] receiving first audio data from a Public Switched Telephone Network (PSTN) client;

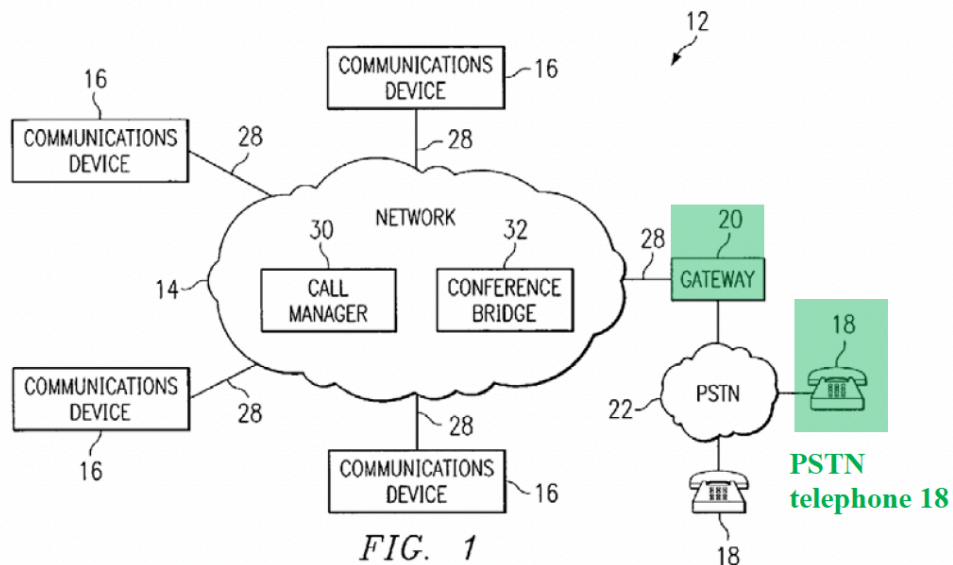
166. Knappe discloses “receiving first audio data from a PSTN client” as required by this limitation.

167. Specifically, Knappe teaches “[c]ommunication networks,” such as one or more telephones 18 and a gateway 20 connected to a PSTN network 22, for

“transporting electrical representations of audible sounds.” Knappe (Ex. 1006) at 1:16-17, FIG. 1.

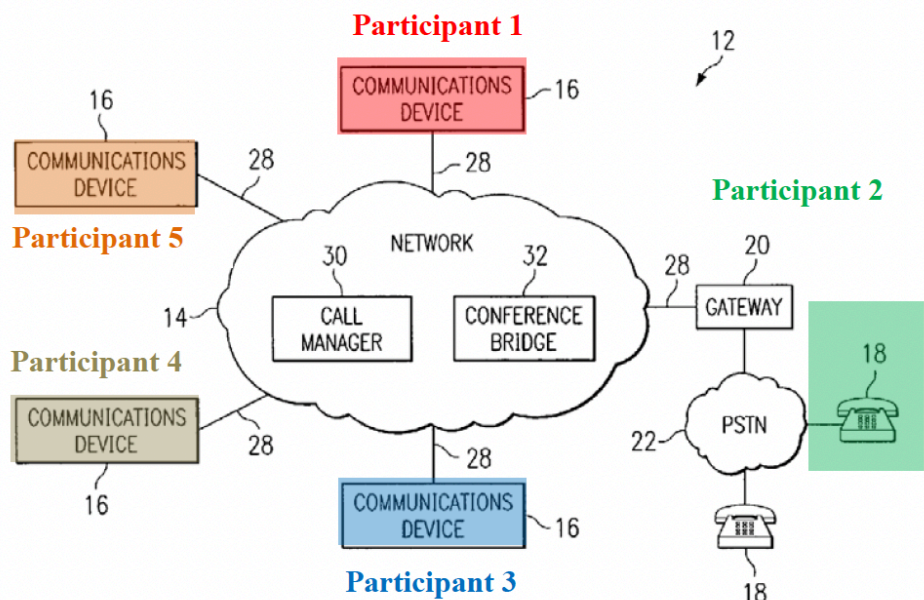
168. According to Knappe, “telephones 18 communicate standard telephony signals through PSTN 22 to the gateway 20.” *Id.* at 4:18-20; *see also, id.* at 5:23-26. The signals from PSTN telephone 18 are “converted to IP packets in the VoIP format” by gateway 20, the VoIP packets “encapsulated in real-time transport protocol (RTP) packets ... or other protocols and formats” for transmission over a network. *Id.* at 3:56-61, 4:20-22; *see also, id.* at 5:23-33 (“If the originating device is a PSTN telephone 18, the media streaming occurs between the gateway 20 and the destination device ... participants are connected and stream media through the conference bridge 32”).

169. Thus, a POSITA reviewing Knappe would have understood that audio signals from PSTN telephone 18 are sent to gateway 20, which communicates corresponding audio data to conference bridge 32, which facilitates the multiparty communications, as shown in the annotated version of the '332 patent's Figure 1, below:



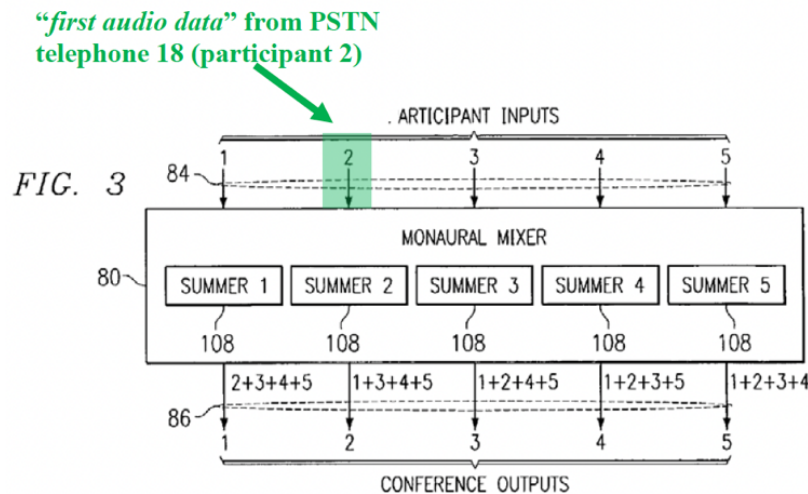
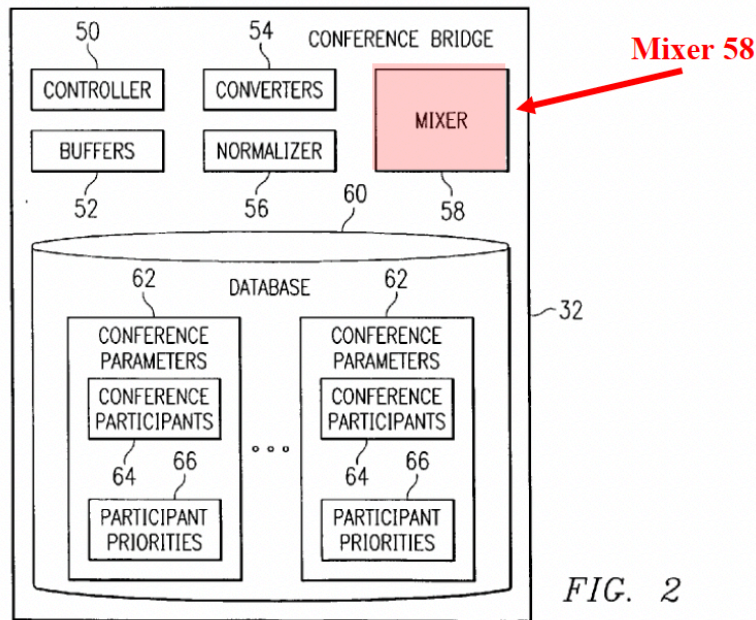
170. Thus, a POSITA understands that in Knappe, audio data from PSTN clients, such as telephone(s) 18, are communicated over link 28 and network 14 to conference bridge 32.

171. For explanatory purposes only, without limiting the Knappe disclosure, the below Figure 1 has been annotated to show exemplary participant labels to correspond to participants shown in Figure 3:



Knappe explains that “conference bridge 32” includes buffers 52, including “input buffers [that] receive and buffer packets of input audio streams from participants for processing by conference bridge 32.” *Id.* at 6:3-6. Thus, Knappe discloses receiving audio data from a PSTN client.

172. In addition, as shown in annotated versions of Figures 2-3 of Knappe below corresponding to the annotated version of Figure 1 above, conference bridge 32 includes at least mixer 58, which may correspond to monaural mixer 80 that “receives participant input streams 84 and combines the streams in summers 82 to generate conference output streams 86 for each participant to a conference call”:



Id. at 6:59-7:11, FIGS. 2-3.

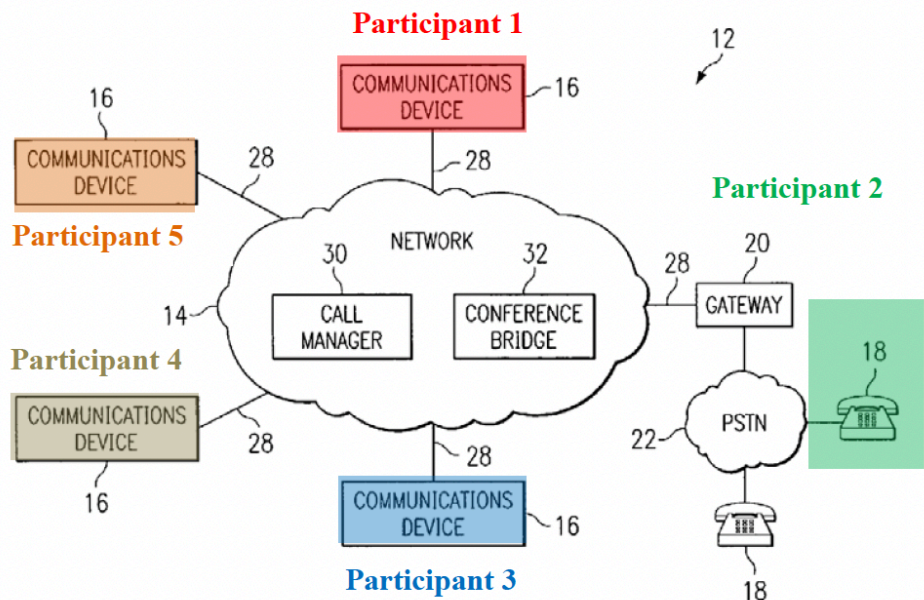
173. Thus, Knappe discloses conference bridge 32, which receives audio data from a PSTN telephone 18 through gateway 20, thereby disclosing this limitation.

[1.2] receiving second audio data from a moderator;

174. Knappe discloses “receiving second audio from a moderator” as required by this limitation.

175. Specifically, Knappe teaches “communication devices 16 comprise IP or other digital telephones, personal and other suitable computers or computing devices ... capable of communicating real-time audio, video and/or other information over the network 14.” Knappe (Ex. 1006) at 3:62-4:1. Knappe further teaches that “at least one participant may be given the priority of moderator for a particular conference call.” *Id.* at 8:52-56; *see also id.* at 7:12-15 (describing the receiving and processing of “audio input 84 of the conference moderator”). Thus, a POSITA reading Knappe would understand that one of Knappe’s communications devices 16 may be designated a moderator.

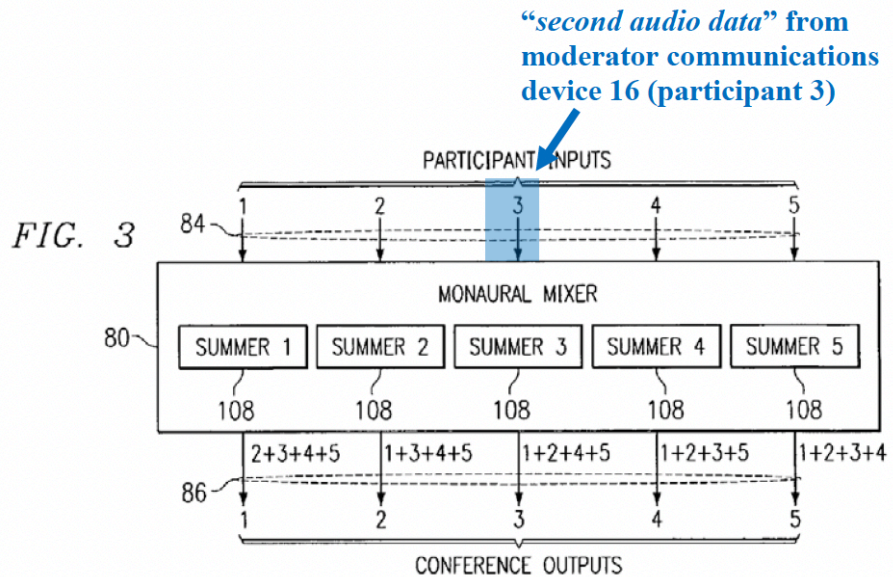
176. Thus, in one example of the Knappe conferencing system, Participant 3 below could be a moderator:



Id. at FIG. 1 (annotated).

177. And as Knappe teaches that the “input buffers receive and buffer packets of input audio streams from participants for processing by conference bridge 32,” Knappe teaches receiving audio streams from participant 3 (the moderator). *Id.* at 6:3-6.

178. Additionally, Knappe teaches that mixer 80 “receives participant input streams 84,” which may include second audio data from a moderator, “and combines the streams in summers 82 to generate conference output streams 86 for each participant to a conference call.” *Id.* at 6:59-7:11, FIGS. 2-3.



Id., FIG. 3 (annotated).

179. Thus, Knappe discloses that conference bridge 32 receives audio data from a moderator, thereby disclosing this limitation.

[1.3.1] receiving third audio data, video data, and collaboration data from at least one remote client;

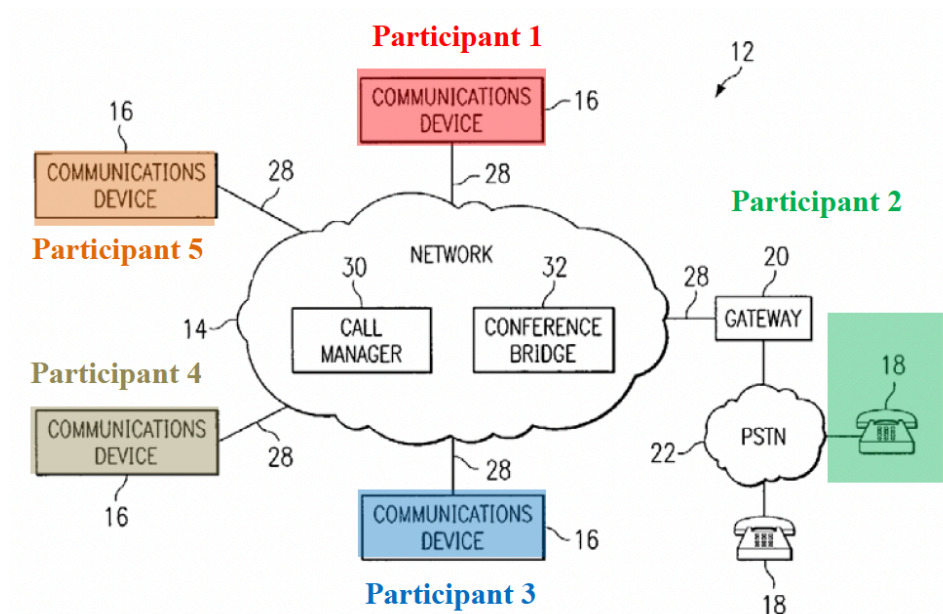
180. The prior art renders obvious “receiving third audio data, video data, and collaboration data from at least one remote client through a first Virtual Private Network (VPN) tunnel” as required by this limitation.

181. Specifically, Knappe teaches receiving “third audio data, video data, and collaboration data from at least one remote client.” Knappe teaches “communication devices 16 comprise IP or other digital telephones, personal and other suitable computers or computing devices ... capable of communicating real-time audio, video and/or other information over the network 14” and “transmitting

audio, video, voice, data and other suitable types of real-time ... traffic.” *Id.* at 3:62-4:1, 3:24-29.

182. Knappe also teaches communication between remote clients over a network and PSTN where “each participant in the conference call may share information with all of the other participants,” contrasting the disclosed system with “a meeting with all persons in the same physical location” and noting the benefit of “minimizing travel expenses.” *Id.* at 1:29-47.

183. A POSITA would have understood that the participants and participants’ communication devices and telephones in the Knappe system were physically remote from one another. As shown in this annotated Figure 1, Knappe discloses at least a third remote client connected to the network over link(s) 28 exchanging “audio, video, and/or ... other suitable information”:



Id. at FIG. 1 (annotated); *see also id.* at 4:58-5:1 (“conference bridge 32 provides conference call and other suitable audio, video, and/or real-time multiparty communication sessions between communication devices 16. A Multiparty communication session includes three or more parties exchanging audio and/or other suitable information.”).

184. As discussed above, Knappe teaches that these conference “participants are connected and stream media through the conference bridge 32.” *Id.* at 5:31-33; *see also, id.*, 4:63-64 (“conference bridge 32 receives media from participating devices 16”), 6:59-67 (“... each participant is assigned a summer 82 that receives audio input streams from each other participant to the conference call”).

185. In Knappe, examples of media that is streamed over network 14 during a conference include “audio, video, voice, data and other suitable types of real-time ... traffic.” *Id.* at 3:24-29. A POSITA would have understood Knappe’s description of “data and other suitable types of real-time ... traffic” to include data to be used for collaboration, including well-known collaboration data such as T.120 standard data. *Id.* at 3:27-28.

186. In addition, Elliott further teaches “collaboration services [between] two or more parties ... can be readily provided over the Internet and include collaborative Web browsing, audio conferencing, video conferencing, and

application sharing,” along with “data conferencing.” Elliott (Ex. 1007) at 1:40-44, 6:46-54; *see also id.* at 5:49-61 (“Session server 12 is preferably able to set up to conduct video conferences, audio conferences, and data conferences with application sharing.”).

187. A POSITA would have understood that Knappe’s conferencing system could be implemented with Elliott’s teachings to include application sharing data, web browsing data, and other collaboration data. *See supra* ¶¶138-144.

188. Therefore, A POSITA would have understood Knappe, individually, and in combination with Elliott, renders obvious this limitation.

[1.3.2] through a first Virtual Private Network (VPN) tunnel

189. The prior art renders obvious that the audio, video, and collaboration data is received through a first Virtual Private Network (VPN) tunnel.

Specifically, Elliott teaches that such “collaborative Web browsing, audio conferencing, video conferencing, and application sharing” may be implemented over a “virtual private network (VPN)” using “IP or IP-tunneling.” Elliott (Ex.

1007) at 1:43-44, 4:31-32; *see also id.* at 2:66-3:1, 6:53-54, 4:6-12

(“communications between parties can be made over a variety of networks 18 . . .

the public Internet is shown as an exemplary Internet Protocol-based (IP-

based) network 18 for multi-media collaboration”), 4:31-34 (“**network 18 can be**

either a public or carrier supported **virtual private network (VPN) . . . which uses IP or IP-tunneling**")

190. Based on the above disclosures, a POSITA would have understood that Elliott uses VPN tunneling for communicating conference voice, video, and collaboration data over a public network 18, such as the Internet. In view of Elliott's teachings, it would have been obvious to a POSITA to use VPN tunneling when communicating audio, video, and collaboration data to and from Knappe's remote communication devices 16 and conference bridge 32 to provide the Knappe system with secure communications over Internet network 14. *See also* Reasons to Combine Knappe and Elliott, *supra* ¶¶ 136-153.

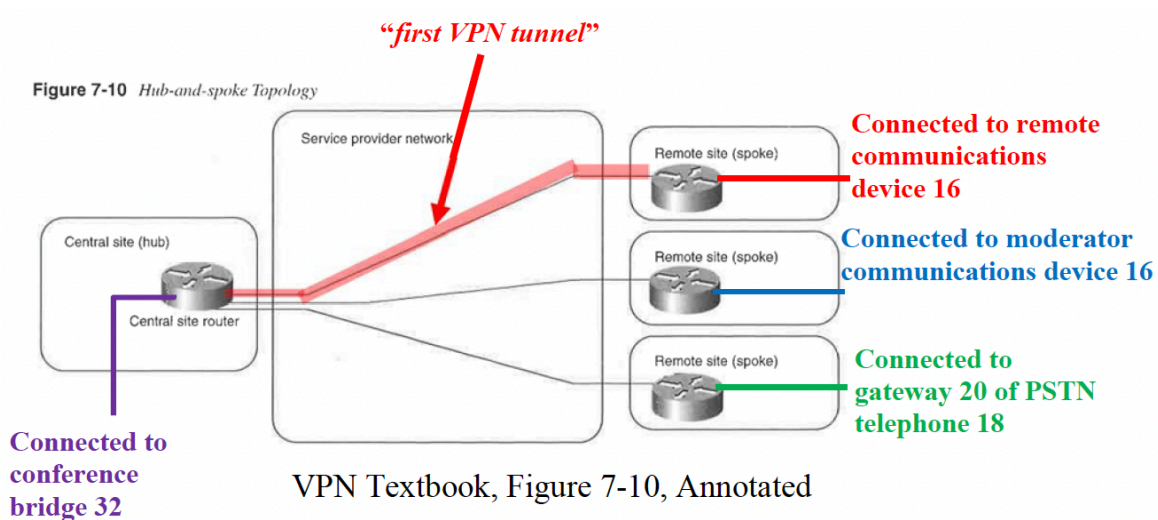
191. Further, the VPN Textbook describes and shows "a hub-and-spoke topology, where a number of remote offices (spokes) are connected to a central site (hub) through a VPN tunnel." VPN Textbook (Ex. 1010) at 129. A POSITA would have understood that each remote site (spoke) would have one or more network devices connected to it.

192. As explained above, Knappe's devices may be at different remote locations on the network. *See, e.g.*, discussion at [1.3.1]; Knappe (Ex. 1006) at 4:30-33, FIG. 1. When implementing VPN tunneling in Knappe, per Elliott, it would have been obvious to a POSITA to connect each of Knappe's devices (which may be located at different locations) to a separate remote router (spoke)

that is connected to the central router (hub), as the VPN Textbook teaches, because this common topology provides a cost effective and less complex implementation. *See, e.g.*, VPN Textbook (Ex. 1010) at 129 (“This is based purely on business needs due to higher costs or increased routing complexity associated with other topologies that use these types of technologies. In other words, there are many examples where the customer could benefit from a different topology but has nonetheless chosen the hub-and-spoke topology for **cost or complexity reasons.**”); *see also* Reasons to Combine Knappe and the VPN Textbook, *supra* ¶¶ 154-162.

193. A POSITA would have recognized that the call manager 30 and conference bridge 32 would be located at a central site because they are depicted in a topologically central location in Knappe’s Figure 1. *See also* Knappe (Ex. 1006) at 3:30-31 (“The call manager 30 and the conference bridge 32 may be located in a central facility.”). Similarly, a POSITA would have recognized that each communications device 16 would be located at a remote site because they are depicted at the topological periphery in Knappe’s Figure 1. *See id.* at 3:62-4:1, 6:17-21. It also would have been obvious to a POSITA for each communications device 16 to be located at a remote site because they are endpoint devices used by remotely-located conference participants. *See, e.g., id.* at 1:42-47 (describing how conferencing allows people to meet without all being the same physical location). Accordingly, it would have been obvious to a POSITA that when

implementing the commonly used hub-and-spoke VPN topology, per the VPN Textbook, Knappe's remote communications device 16, moderator communications device 16, and gateway 20 connected to the PSTN telephone 18 would be connected to different remote site (spoke) routers, while the conference bridge 32 would be connected to the central site (hub) router, as illustrated below in annotated Figure 7-10:



VPN Textbook (Ex. 1010) Figure 7-10 (annotated).

194. As shown, the remote site router (spoke) of the remote communications device 16 and the central router (hub) of the conference bridge 32 are connected through a VPN tunnel. When the conference bridge 32 receives data from or transmits data to the remote communications device 16, the data would be received/sent through the same VPN tunnel, which corresponds to “a first VPN

tunnel.” Thus, a POSITA would have understood Knappe in combination with Elliott and/or the VPN Textbook renders obvious this limitation.

[1.4] mixing the first audio data from the PSTN client with the second audio data from the moderator into a first mixed audio data;

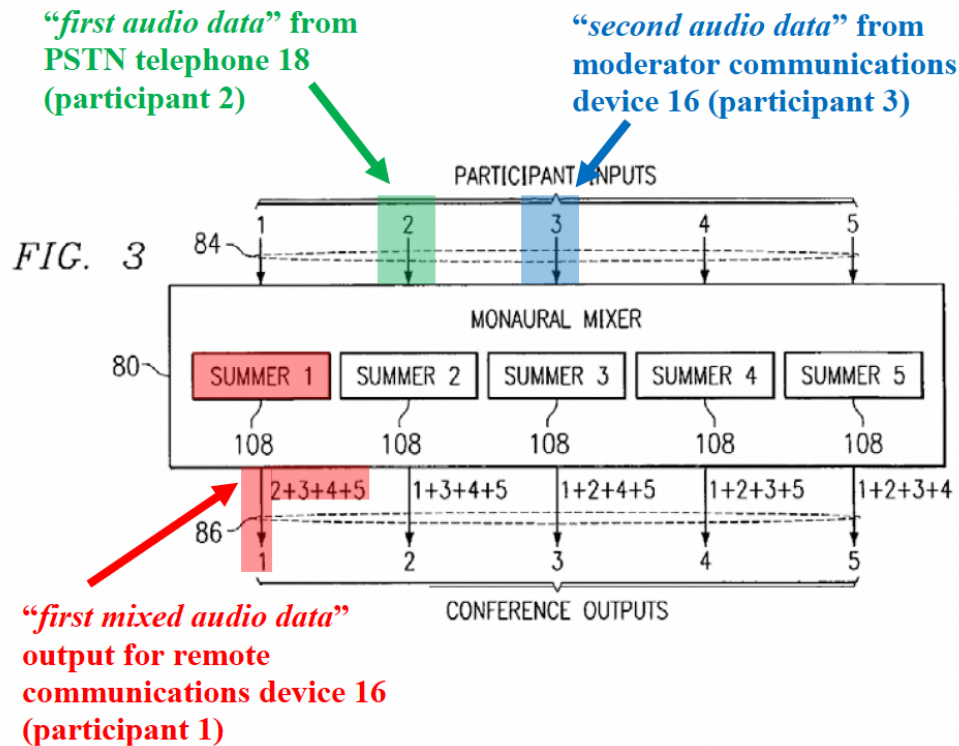
195. Knappe discloses “mixing the first audio data form the PSTN client with the second audio data from the moderator into a first mixed audio data.”

196. As discussed above in connection with limitations [1.1] and [1.2], Knappe teaches receiving “*the first audio data from the PSTN client*” and “*the second audio data from the moderator.*” Knappe also teaches that mixers 58, 80, and 100, of conference bridge 32 mixes the audio data received from PSTN telephone 18 with audio data received from a moderator’s communications device 16 into “*a first mixed audio data.*”

197. For example, Knappe explains that “monaural mixer 80 receives participant input streams 84 and combines the streams in summers 82 to generate conference output streams 86 for each participant.” Knappe (Ex. 1006) at 6:59-67. As a result, “the conference output stream of participant 1 includes the audio inputs of participants 2-5 ... the conference output stream of participant 2 includes the audio inputs of participants 1 and 3-5,” and so on. *Id.* at 7:1-11.

198. A POSITA would have understood Knappe to disclose the “*mixing the first audio data from the PSTN client with the second audio data from the*

moderator into a first mixed audio data,” as shown in an exemplary implementation in the below annotated Figure 3:



Id. at FIG. 3 (annotated).

199. Therefore, Knappe teaches mixing audio data from a PSTN client with audio data from the moderator.

[1.5] transmitting the first mixed audio data to the remote client through the first VPN tunnel;

200. The prior art renders obvious “transmitting the first mixed audio data to the remote client through the first VPN tunnel.” As discussed above in connection with limitations [1.3.1]-[1.4], Knappe teaches generating “*the first mixed audio data*” (e.g. output streams 86) as well as “*remote client[s]*” (e.g.,

remote communications devices 16). Knappe also discloses “transmitting audio ... between source and destination endpoints,” including “a conference output stream for delivery to the participant.” Knappe (Ex. 1006) at 3:24-29; 6:65-67; *see also id.* at 5:35-39 (“The conference output stream for a participant includes the media of all other participants, a subset of other participants or other suitable mix”). Thus, a POSITA would have understood Knappe to disclose “*transmitting the first mixed audio data to the remote client.*”

201. As discussed above in connection with limitation [1.3.2], Knappe in combination with Elliott and the VPN Textbook teaches using a VPN tunnel between a central site (e.g., Knappe’s conference bridge 32) and a remote client (e.g., remote communications devices 16). It would have been obvious to a POSITA to transmit “the first mixed audio data to the remote client” using a VPN tunnel set up between conference bridge 32 and remote communications device 16.

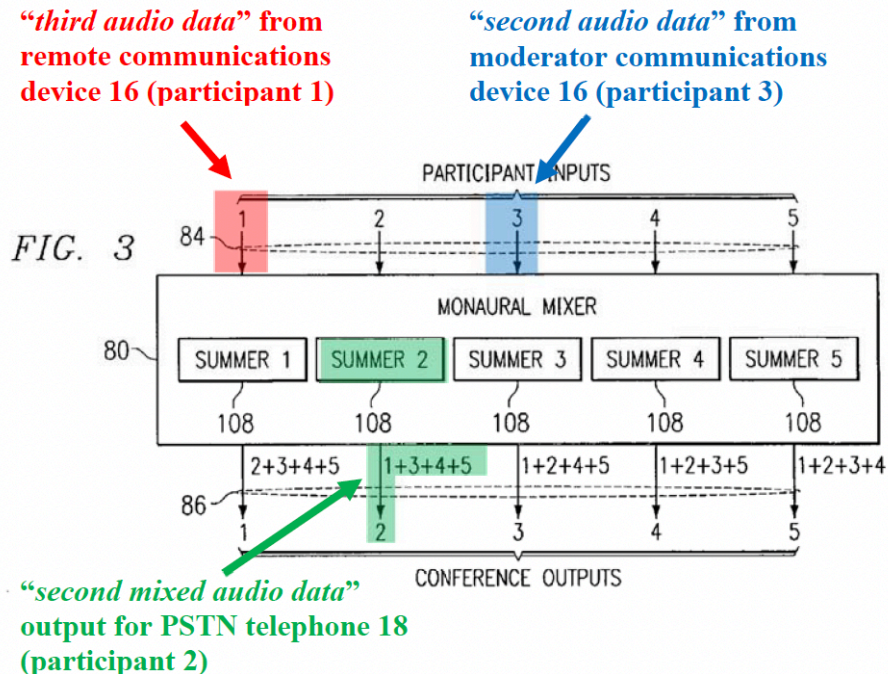
202. Thus, Knappe, in combination with Elliott and the VPN Textbook renders obvious this limitation.

[1.6] mixing the second audio data from the moderator with the third audio data from the remote client into a second mixed audio data; and

203. Knappe discloses “mixing the second audio data from the moderator with the third audio data from the remote client into a second mixed audio data.”

204. As discussed above in connection with limitations [1.2] and [1.3.1], Knappe teaches receiving “*the second audio data from the moderator,*” and “*the third audio data from the remote client.*” Also, as discussed above in connection with limitation [1.4], Knappe teaches that mixers 58, 80, and 100, of conference bridge 32, mixes the audio data from PSTN telephone 18 with audio data from a moderator’s communications device 16 into “*a first mixed audio data.*”

205. As shown in annotated Figure 3 of Knappe below, “monaural mixer 80 receives participant input streams 84 and combines the streams in summers 82 to generate conference output streams 86 for each participant,” which also may include mixing audio data from a moderator’s communications device 16 with another participant’s communications device 16. Knappe (Ex. 1006) at 6:59-67. As a result, Knappe’s “conference output stream of participant 2” can comprise “*a second mixed audio data*” resulting from a mix of “the audio inputs of participants 1 and 3-5,” which may include “*the second audio data from the moderator with the third audio data from the remote client*”:



Id. at FIG. 3 (annotated), 7:1-11.

206. Thus, a POSITA would have understood this description in Knappe to disclose this limitation.

[1.7] transmitting the second mixed audio data to the PSTN client.

207. Knappe discloses “transmitting the second mixed audio data to the PSTN client.”

208. As discussed above in connection with limitation [1.5], Knappe discloses “transmitting audio ... between source and destination endpoints,” including “a conference output stream for delivery to the participant.” Knappe (Ex. 1006) at 3:24-29; 6:65-67; *see also id.* at 5:35-39 (“The conference output stream for a participant includes the media of all other participants, a subset of other participants or other suitable mix”).

209. Further, Knappe teaches the transmitted audio data is received at gateway 20 and converted to standard telephony signals for delivery to destination PSTN telephones 18, which receives the audio. *Id.* at 4:21-25 (“At the gateway 20 ... VoIP packets received from the network 14 are converted into standard telephony signals for delivery to the destination telephone 18 through PSTN 22.”), 7:1-7 (“... each participant receives the audio input of each other participant.

210. Thus, a POSITA would have understood Knappe to disclose “*transmitting the second mixed audio data to the PSTN client.*”

4. Claim 2

[2.P] The method of claim 1 further comprising:

211. See analysis for [1.P]-[1.7].

[2.1] mixing the first audio data from the PSTN client with the third audio data from the remote client into a third mixed audio data;

212. Knappe further discloses “mixing the first audio data from the PSTN client with the third audio data from the remote client into a third mixed audio data” as required by this limitation [2.1].

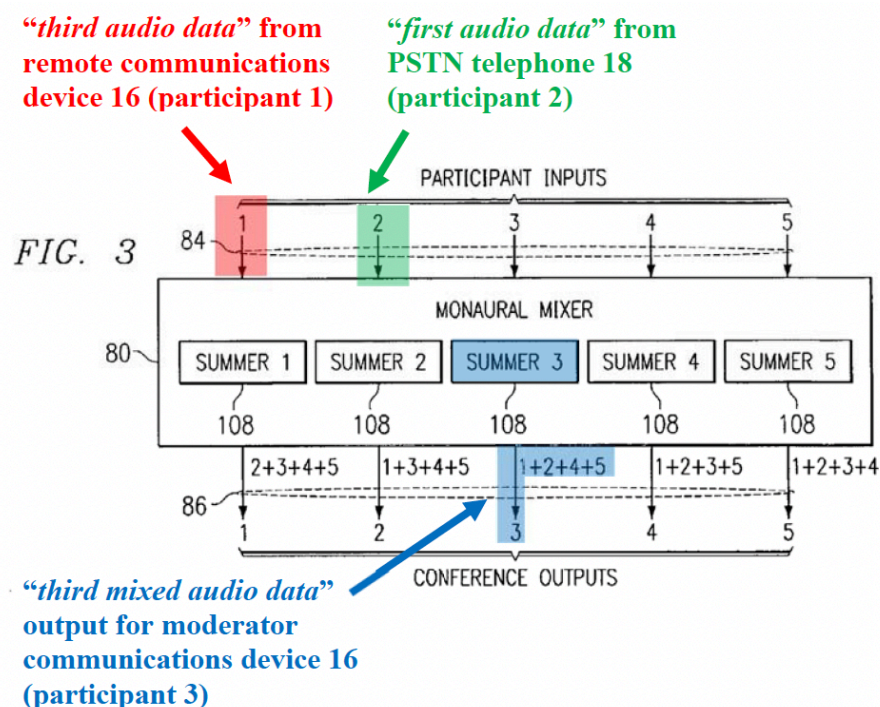
213. As discussed above in connection with limitations [1.1] and [1.3.1], Knappe teaches receiving “*the first audio data from the PSTN client*” and “*the third audio data from the remote client*” by a mixer (e.g., mixers 58, 80 or 100).

214. Also, as discussed above in connection with limitations [1.4] and [1.6], Knappe teaches that mixers 58, 80, and 100, of conference bridge 32, can

mix the audio data from PSTN telephone 18 with audio data from a moderator's communications device 16 into "*a first mixed audio data.*"

215. As shown in annotated Figure 3 of Knappe below, "monaural mixer 80 receives participant input streams 84 and combines the streams in summers 82 to generate conference output streams 86 for each participant," which also may include mixing audio data from PSTN telephone 18 with another participant's communications device 16. Knappe (Ex. 1006) at 6:59-67.

216. As a result, Knappe's "conference output stream of participant 3" can comprise "*a third mixed audio data*" resulting from a mix of "the audio inputs of participants 1-2 and 4-5," which may include "*the first audio data from the PSTN client with the third audio data from the remote client*":



Id. at FIG. 3 (annotated), 7:6-7.

217. Thus, a POSITA would have understood this description in Knappe to disclose the “*mixing the first audio data from the PSTN client with the third audio data from the remote client into a third mixed audio data.*”

[2.2] playing the third mixed audio data to the moderator.

218. Knappe discloses “playing the third mixed audio data to the moderator.”

219. As discussed above in connection with limitation [2.1], Knappe teaches that mixers 58, 80, and 100, of conference bridge 32, can generate “*the third mixed audio data,*” and that “each device 16 receives a conference signal that includes contributions from all other participating devices.” *Id.* at 4:58-5:1.

220. Also, as discussed above in connection with limitation [1.2], one of Knappe’s communications devices 16 may be designated as a moderator, which can receive the mixed audio data from said mixers. *Id.* at 8:52-53 (“... at least one participant may be given the priority of moderator for a particular conference call.”).

221. Further, Knappe teaches that “the communication devices 16 comprise real-time applications that play traffic as it is received” with “IP packets received from the network 14 [being] converted back into analog signals and played to the user.” *Id.* at 4:5-17. Thus, as shown above in [2.1], “*the third mixed*

audio data” generated by summer 3 is output to the moderator communication device 16, where it is played to the moderator communication device 16.

222. Therefore, Knappe discloses communications device 16 receiving “*the third mixed audio data*” and “*playing*” it “*to the moderator.*”

5. Claim 4

[4.P] The method of claim 1

223. See analysis at [1.P]-[1.7].

[4.1] wherein the receiving third audio data from the remote client includes receiving IP packets containing the third audio data corresponding to the remote client and decoding the IP packets containing the third audio data from the remote client.

224. The prior art renders obvious “wherein the receiving third audio data from the remote client includes receiving IP packets containing the third audio data corresponding to the remote client and decoding the IP packets containing the third audio data from the remote client.”

225. As discussed above in connection with limitation [1.3.1], Knappe teaches conference bridge 32 “receiving third audio data from the remote client” (e.g., remote communications device 16). Knappe also teaches that network 14 uses IP packets for communication. Knappe (Ex. 1006) at 3:50-54 (“network 14 is the Internet ... In the Internet embodiment, the network 14 transmits information in Internet Protocol (IP) packets.”).

226. Knappe further discloses “decoding the IP packets” as it teaches that conference bridge 32 includes input converters 54:

The converters 54 include input and output converters. The input converters receive input packets of a participating from a corresponding buffer and **convert the packet** from the native format of the participant device 16 to a standard format of the conference bridge 32 . . .

Id. at 6:16-21 (*see also id.* at 6:3-15).

227. A POSITA would understand that Knappe’s conversion of the received IP packets from the native format of the participant device to a standard format would include the claimed “decoding.” Indeed, Knappe discloses the use of

[a] codec (coder/decoder) [that] converts audio, video, or other suitable signals generated by users from analog signals into digital format. The digital encoded data is encapsulated into IP or other suitable packets from transmission over the network 14. IP packets received from network 14 are converted back into analog signals and played to the users. It will be understood that the communication devices 16 may otherwise suitably encode and decode signals transmitted over or received form the network 14.

Id. at 4:4-17 (*see also id.* at 4:18-28), 6:3-21; *see also* Caputo (Ex. 1021) at 11.

228. Furthermore, as disclosed above for limitation [1.3.2], the data sent from remote client is sent over a “virtual private network (VPN)” using “IP or IP-tunneling.” Elliott (Ex. 1007) at 1:43-44, 4:31-32. Thus, a POSITA would have understood Knappe to have disclosed “receiving third audio data from the remote client includes receiving IP packets containing the third audio data corresponding

to the remote client and decoding the IP packets containing the third audio data from the remote client.”

6. Claim 5

[5.P] The method of claim 1

229. See analysis at [1.P]-[1.7].

[5.1] wherein the data communicated between the moderator and the PSTN client is transported through a second VPN tunnel.

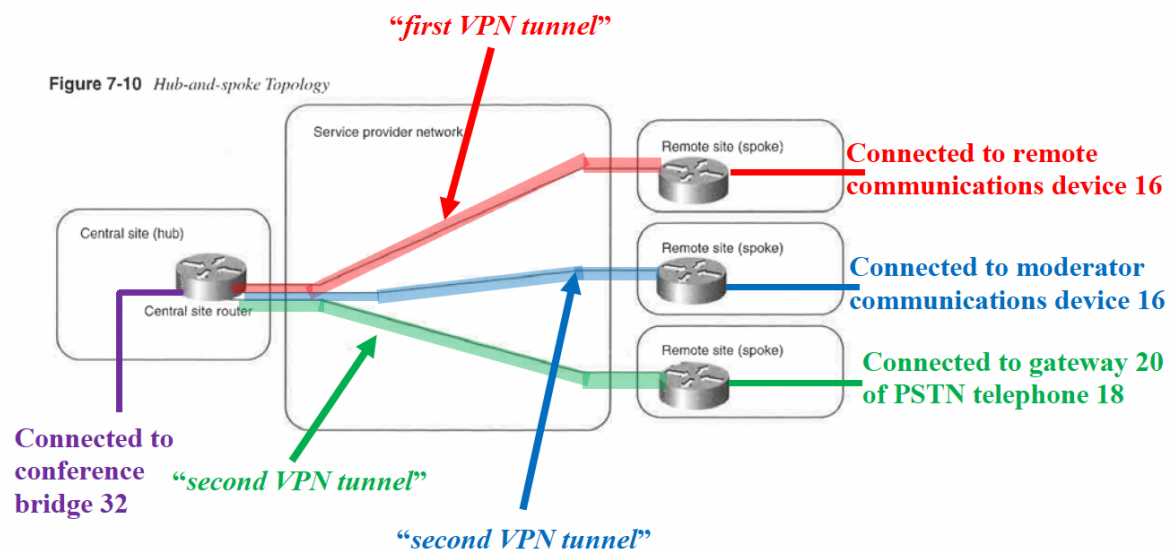
230. The prior art renders obvious “the data communicated between the moderator and the PSTN client is transported through a second VPN tunnel.”

231. As a threshold matter, claim 1, from which claim 5 depends, uses the phrases “first audio data from a Public Switched Telephone Network (PSTN) client,” “second audio data from a moderator,” “first mixed audio data,” and “second mixed audio data,” but does not recite the phrase “data communicated between the moderator and the PSTN client.” For the purposes of this IPR only, I will assume that “*the data communicated between the moderator and the PSTN client*” corresponds to one of the data recited in claim 1, whereby this claim would have been obvious to a POSITA.

232. As discussed above in connection with limitations [1.1]-[1.2] and [1.6]-[1.7], Knappe teaches communicating data between “*the moderator*” (e.g., a communications device 16 assigned to be moderator) and “*the PSTN client*” (e.g., a PSTN telephone 18) via gateway 20.

233. And as discussed above in connection with limitations [1.5], a POSITA would have been motivated to combine Knappe with Elliott and the VPN Textbook to understand that Knappe's "data" could be "transported through ... a VPN tunnel."

234. Thus, it would have been obvious to a POSITA to implement "a second VPN tunnel" for transporting "the data communicated between the moderator and the PSTN client" as shown in this annotated Figure 7-10 from the VPN Textbook:



VPN Textbook (Ex. 1010) at FIG. 7-10 (annotated).

235. The VPN tunnel between the central (hub) router (e.g., conference bridge 32) and the remote (spoke) routers (e.g., moderator communications device 16 and PSTN telephones 18) may correspond to "a second VPN tunnel."

236. Thus, during a conference, data communications from the moderator communications device 16 may be transmitted to conference bridge 32 through a second VPN tunnel, as labeled in blue in the above figure. After mixing by conference bridge 32, mixed audio data may be transmitted to the PSTN telephone 18, also through a second VPN tunnel, as labeled in green in the above figure.

237. Thus, Knappe in combination with Elliott and the VPN Textbook discloses “*data communicated between the moderator and the PSTN client is transported through a second VPN tunnel,*” thereby rendering this limitation of claim 5 obvious.

7. Claim 7

[7] The method of claim 1 further comprising

238. See analysis at [1.P]-[1.7].

[7.1] VoIP encoding the second mixed audio data before transmitting the second mixed audio data to the PSTN client.

239. The prior art renders obvious the “VoIP encoding the second mixed audio data before transmitting the second mixed audio data to the PSTN client” limitation of claim 7.

240. As discussed above in connection with limitations [1.6]-[1.7], Knappe discloses “*transmitting the second mixed audio data to the PSTN client.*”

241. Knappe further teaches that such data may be transmitted as VoIP packets. Knappe (Ex. 1006) at 3:54-56 (“Telephony voice information is

transmitted in the Voice over IP (VoIP) format”). Knappe explains that during a conference, “VoIP packets received” by gateway 20 may be “converted into standard telephony signals for delivery to destination telephone 18 through PSTN 22.” *Id.* at 4:17-28.

242. Since Knappe’s system communicates using VoIP packets and converts the VoIP packets to standard telephony signals for the PSTN telephone 18, and Knappe discloses “a codec (coder/decoder) converts audio,” a POSITA would understand Knappe’s conference bridge 32 to perform VoIP encoding of the second mixed audio data before transmission to PSTN telephone 18 via gateway 20. *Id.* at 4:9.

243. Thus, Knappe discloses “*VoIP encoding the second mixed audio data before transmitting the second mixed audio data to the PSTN client,*” rendering claim 7 obvious.

B. Claims 3, 6, and 8-12 Are Obvious Under 35 U.S.C. § 103(a) Over Knappe in View of Elliott, the VPN Textbook, and Drell

1. Summary of Drell

244. United States Patent No. 7,089,285 (“Drell”) entitled “Videoconferencing Apparatus Having Integrated Multi-Point Conference Capabilities” was filed October 5, 2000, claiming priority to a provisional application filed on October 5, 1999. Drell (Ex. 1008).

245. Like Knappe, Drell also describes multi-point (MP) conferencing, including videoconferencing, among a plurality of endpoints, as shown in Figure 1:

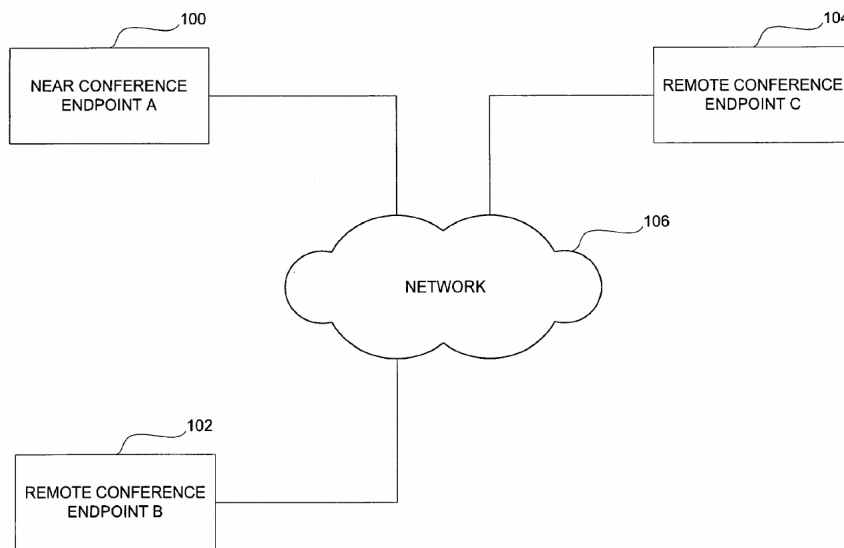


FIG. 1

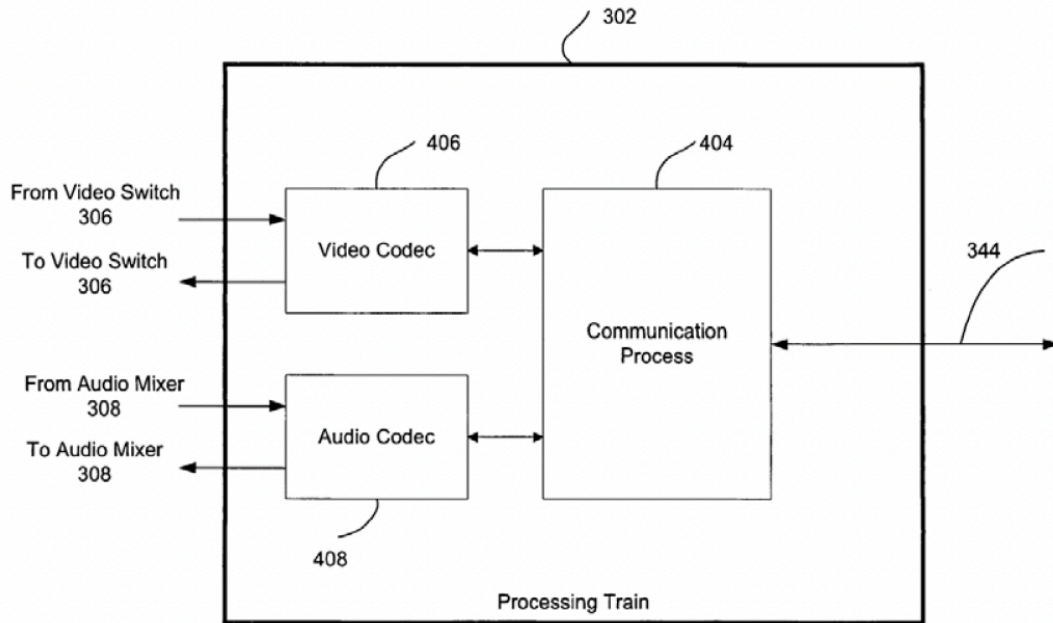
Id. at FIG. 1.

246. Specifically, Drell discloses a videoconferencing apparatus “to combine and distribute audio and video signals received from a plurality of remote conference endpoints[.]” *Id.* at Abstract. Drell explains that “network 106” transmits “audio and video data between and among near conference endpoint 100 and remote conference endpoints 102 and 104.” *Id.* at 3:20-23. “Although a single network 106 is shown,” Drell “contemplates the use of two or more networks (for example, the PSTN and the Internet) to connect conference endpoints utilizing different communication protocols.” *Id.* at 3:33-38. Thus, Drell also discloses conferencing in a hybrid Internet and PSTN network.

247. Drell explains that centrally located conference bridges can have significant cost disadvantages and thus teaches use of “MP conferencing application 234” that is “software-based” and located at a “near conference point 100.” *Id.* at 1:5-2:8, 4:6-20.

248. Drell explains that “MP conferencing application 234 is configured to instantiate a processing train for each remote conference endpoint” where the “processed audio and video data streams are combined with each other and with locally generated audio and video streams.” *Id.* at 4:21-29. The “audio and video streams are thereafter distributed to remote conferencing endpoints 102 and 104.” *Id.* at 4:29-31.

249. Drell further teaches that MP conferencing application 234 uses “a set of codecs” and may “comprise an H.320 process (ISDN-based) or H.323 (packet-based) process.” *Id.* at 2:17-33, FIG. 4.



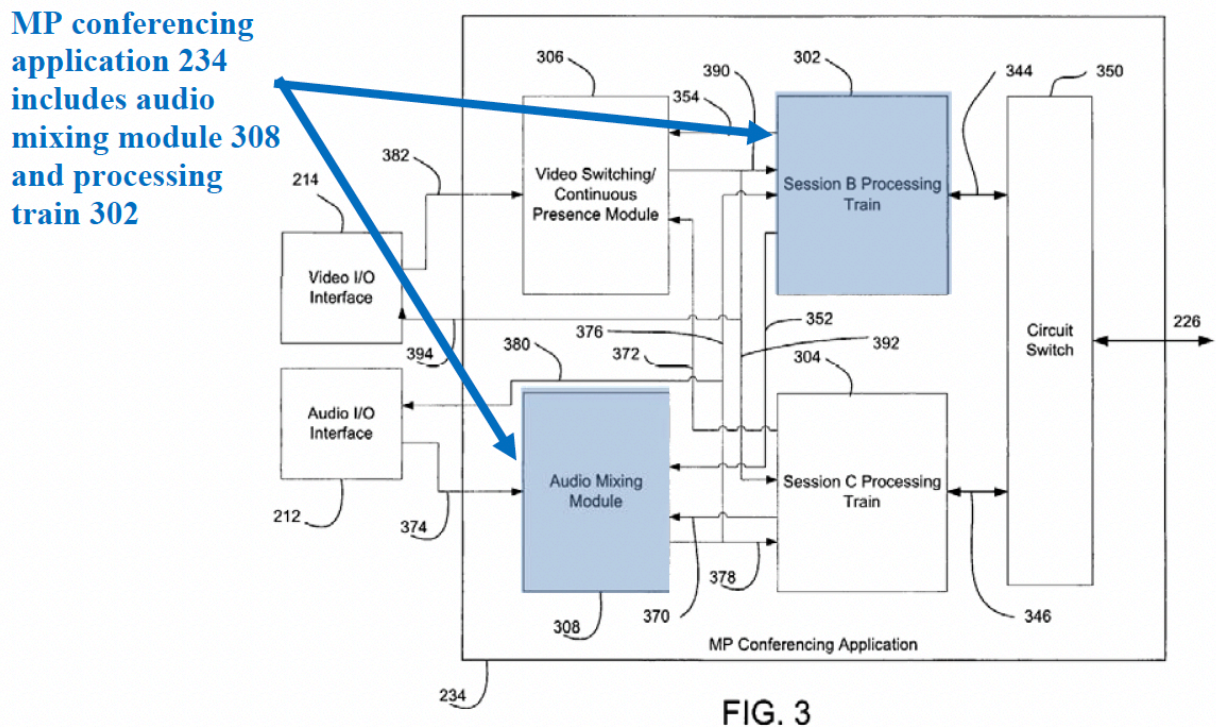
Id. at FIG. 4.

250. “Audio codec 408 receives the audio data stream ... and applies ... decoding in accordance with a standard ... or proprietary audio compression algorithm” and “video codec 406 receives the video data stream and applies ... decoding in accordance with a standard ... or proprietary video compression algorithm.” *Id.* at 6:18-27.

251. Drell’s MP conferencing application 234 further includes audio mixing module 308 and a “processing train 302 [that] may include a data codec for coding and encoding still images and the like received from or transmitted to remote conference endpoints 102 and 104.” *Id.* at 6:53-56. “In the transmission mode, processing train 302 is configured to receive the audio and video data streams output by audio mixing module 308 and video switching/continuous

presence module 306,” along with conferencing data such as the aforementioned “still images,” and “encode[] and combine[] to form a mixed encoded audio/video data stream.” *Id.* at 5:47-54, 6:55, FIG. 3.

252. Annotated Figure 3 of Drell, below, highlights the audio mixing module 308 and the processing train 302, which a POSITA would have understood to be able to encode and mix “the audio and video data streams” with collaboration data, such as “still images,” to generate mixed data:



Id. at FIG. 3 (annotated); *see also id.* at 5:47-54, 6:55.

253. Finally, Drell teaches “[a] video monitor 208 and loudspeaker 210 present images and speech of the remote images and speech.” *Id.* at 3:45-48, FIG.

2.

2. Reasons to Combine Knappe and Drell

254. A POSITA would have been motivated to combine the teachings of Knappe and Drell to produce the obvious, beneficial, and predictable results of implementing various aspects of Knappe's conferencing system, including use of a conference moderator, as software and with an audio codec using a standard or proprietary compression algorithm to reduce the amount of data being transmitted.

255. Knappe and Drell are analogous prior art both pertaining to the same field of endeavor of multiparty conferencing, including mixing audio from Internet and PSTN endpoints. A POSITA understood that it was routine and would have recognized it is desirable to have a person (e.g. a leader or a moderator) to control the conference. *See e.g.*, U.S. Patent No. 3,601,530 (Ex. 1014) at 2:34-40 (“[A] conference leader mode of operation is provided. In this mode . . . the conference leader . . . [has] a substantial degree of control over the conference.”). This would be desirable because frequently in business conferencing, at the time and even now, one person is in charge of the conference. Consistent with this understanding, Knappe teaches that at least “one participant may be given the priority of moderator for a particular conference call” to “control the flow and/or organization of the conference.” Knappe (Ex. 1006) at 8:52-56; 1:53-55.

256. Knappe also teaches that its “call manager 30 and conference bridge 32” may “have their have their functionality distributed across and/or at the

periphery of the network 14.” Knappe (Ex. 1006) at 4:31-33. Drell further teaches a “MP conferencing application” that performs functions similar to Knappe’s call manager 30 and conference bridge 32 implemented as software at a conference endpoint. Drell (Ex. 1008) at 1:65-2:6, 6:63-66 (“The present invention advantageously utilizes software-based processing of video and audio data streams to implement a multi-point conferencing capability in a conference endpoint.”). Thus, it would have been obvious to a POSITA that the moderator communications device 16 of Knappe could implement the call manager and conference bridge 32 functionality (like Drell’s MP conferencing application) as software at the moderator.

257. The combination of Knappe and Drell is merely the ordinary use of a common technique (implementing conference bridging and moderator functionality in software) in the system of Knappe where a participant is designated as moderator to yield predictable results of initiating, organizing, controlling the conference through software at a conference endpoint and reducing the cost of having a separate conference bridge. Each of these reasons separately and together would have motivated a POSITA.

258. In addition, a POSITA also would have been motivated to implement Drell’s “[a]udio codec” capable of encoding/decoding “in accordance with a standard ... or proprietary audio compression algorithm” and “video codec 406”

capable of encoding/decoding “in accordance with a standard ... or proprietary video compression algorithm” with Knappe’s “communication devices 16 compris[ing] ... [a] codec (coder/decoder).” Drell (Ex. 1008) at 6:18-45; Knappe (Ex. 1006) at 4:5-11. This would have the benefit of reducing the amount of data transmitted over Knappe’s network 14.

259. Knappe teaches that “[i]n the Internet embodiment, the network . . . transmits information in Internet Protocol (IP) packets.” Knappe (Ex. 1006) at 3:50-55. Drell teaches that its conference system uses codecs to encode mixed voice data before it is transmitted over the Internet network. *See e.g.* Drell (Ex. 1008) at 6:33-45 (“In the transmit mode . . . [a]udio codec 408 encodes the audio data . . . using a standard or proprietary audio compression algorithm (e.g., G711_ and delivers the encoded audio data to communication process 404.”), 3:33-34 (“network 106 may comprise a packet switched network, such as the Internet). Thus, a POSITA would have understood that Drell is transmitting its audio data over the Internet using IP packets, like in Knappe and thus it would have been obvious to use in Knappe the disclosed codecs for encoding/decoding data using audio compression algorithms as taught in Drell.

260. The beneficial results would have been predictable and a POSITA would have had a reasonable expectation of success given Knappe’s mention of a

codec and Drell's discussion of codecs for encoding, decoding, and compressing in the audio and video conferencing context.

261. The combination permits, but does not require, physical incorporation of elements from Drell into Knappe's system.

3. Claim 3

[3.P] The method of claim 1

262. See analysis at [1.P]-[1.7]

[3.1] wherein the receiving first audio data from the PSTN client includes receiving IP packets and decoding the IP packets containing the first audio data from the PSTN client at a moderator computer system.

263. The prior art renders obvious this limitation of claim 3.

264. As discussed above in connection with limitation [1.1], Knappe teaches that conference bridge 32 receives audio data from the PSTN telephone 18 through gateway 20, disclosing "*receiving first audio data from the PSTN client.*" Further, as discussed at [1.2], Knappe teaches a moderator communications device 16, which may be implemented as a computer system. See, e.g., Knappe (Ex. 1006) at 3:62-64 ("The communication devices 16 comprise IP or other digital telephones, personal and other suitable computers or computing devices.").

265. Knappe further teaches that the audio data from the PSTN telephone is transmitted as VoIP packets by gateway 20. See *id.* at 3:54-58 (Telephony voice information is **transmitted in the Voice over IP (VoIP) format**"), 4:21-22

(“At the gateway 20, the signals are **converted to IP packets in the VoIP format**”).

266. Knappe also teaches that the conference bridge 32 receives and converts packets containing audio data from the PSTN telephone via gateway 20. *Id.* at 6:16-21 (“The input **converters receive input packets . . . and convert the packet** from the native format of the participant device 16 **to a standard format of the conference bridge 32**”), 3:62-44 (“The communication devices 16 comprise IP or other digital telephones . . . **such as the telephone 18 and gateway 20 combination capable of communicating real-time audio**”).

267. A POSITA would have understood that as part of converting the received input packets, Knappe’s converters 54 decode the IP packets so that they are in a standard format which the conference bridge 32 can utilize in the mixing process. A POSITA would understand that Knappe’s step of converting the received packetized audio data would include “*decoding*” as claimed since in Knappe “converting” of audio is performed with a codec (coder/decoder). *See id.* at 4:9 (“[A] codec (coder/decoder) converts audio”).

268. Therefore, Knappe discloses “*receiving IP packets and decoding the IP packets containing the first audio data from the PSTN client.*”

269. Knappe contains further disclosures recognizing that the implementation of the conference bridge 32 is not limited to a central facility. *See,*

e.g., id. at 4:31-33 (“The call manager 30 and the conference bridge 32 may be located in a central facility or have their functionality distributed across and/or at the periphery of the network 14.”). Drell further supplements this idea, explaining that an external conference bridge (which it refers to as a “multi-point control unit (MCU)”) has a “significant disadvantage... [— its] expense” and further notes a “high cost of purchasing or renting.” Drell (Ex. 1008) at 1:45-54. To mitigate cost, Drell teaches a “relatively low-cost videoconferencing apparatus” that is software-based and located at a “near conference endpoint.” *Id.* at 1:65-66.

270. Specifically, Drell teaches that a conference bridge such as that taught in Knappe could be implemented at a conference endpoint, like the moderator disclosed in Knappe, using software applications including a “call manager application 232 and the MP conferencing application 234.” *Id.* at 2:3-8 (“The IMUX functions are implemented in a software-based circuit switch operatable to aggregate a plurality of processing trains”), 4:6-20. Drell describes near conference point 100 as having “a call manager application 232, and the MP conference application 234.” *Id.* at 3:39-41, 4:11-13, FIG. 2.

271. In Drell, the “[c]all manager application 232 controls the establishment and termination of connections between near conferencing endpoint 100 and remote conference endpoints 102 and 104, and may also furnish information characterizing the nature of individual connections to MP

conferencing application 234.” *Id.* at 4:15-20. The “MP conferencing application 234 is configured to instantiate a processing train for each remote conference endpoint...The processed audio and video data streams are combined with each other and with locally generated audio and video streams.” *Id.* at 4:21-29. Once combined, the “audio and video streams are thereafter distributed to remote conferencing endpoints 102 and 104.” *Id.* at 4:29-31.

272. It would have been obvious to a POSITA to distribute Knappe’s conference bridge 32 functionality (as Knappe suggests and Drell describes) as software located at moderator communications device 16 of Knappe because doing so would reduce the cost of facilitating conference communications. Further, it would have been obvious that in the combination, Knappe’s communications devices 16 (which can be any suitable computer, Knappe (Ex. 1006) at 3:62-64), including the device designated as the moderator (as discussed at [1.2]), would include a conference bridge implemented as software similar to Drell’s conference endpoint (which is also a computer with CPU, memory, and network interfaces, Drell (Ex. 1008) at 3:54-64).

273. Further, Drell explains that the conferencing application 234 of the near conference endpoint 100 receives packets containing audio and video data from remote endpoints 102 and 104, including from PSTN telephones. *Id.* at 4:21-31 (“The processing trains [of MP conference application 234] process **audio and**

video streams received from remote conference endpoints 102 and 104”), 3:33-38.

274. Drell also explains that conferencing application 234 decodes packets. *Id.* at 6:18-27 (“[a]udio codec 408 [that] receives the audio data stream ... and applies ... decoding in accordance with a standard ... or proprietary audio compression algorithm. The decoded audio data stream is then sent to audio mixing module 308”). Drell explains that the communication process of the conferencing application may comprise an “H.323 (packet-based) process” where “audio and video signals are thereafter respectively delivered to audio and video codecs for decoding.” *Id.* at 2:26-33. A POSITA would have understood that Drell’s disclosure of decoding received packetized audio corresponds to converting received audio as Knappe discloses.

275. In view of Drell’s teaching, it would have been obvious to a POSITA for Knappe’s conference bridge 32 (implemented as software in the moderator communications device 16) to decode received IP packet data, because it would allow for mixing of the decoded audio data for each participant (as discussed at [1.4], [1.6], and [2.1]) and for playing mixed audio data to the moderator (as discussed at [2.2]).

276. It would have been obvious to a POSITA that in the combination, Knappe’s conference bridge 32 would be implemented at least in part as software

within communications device 16, including the computing device of the moderator. During operation, the conference bridge 32, implemented as software within the moderator communications device 16, would receive IP packets and convert the audio signals from the remote participants, including audio from the PSTN telephone 18.

277. Thus, Knappe in combination with Drell teaches that receiving audio data from the PSTN telephone 18 includes receiving IP packets and converting the IP packets containing the audio data from the PSTN telephone 18 at the moderator communications device 16, which renders obvious this limitation.

4. Claim 6

[6.P] The method of claim 1 further comprising

278. See analysis at [1.P]-[1.7].

[6.1] encoding the first mixed audio data into a first compressed audio data before transmitting the first mixed audio data to the remote client.

279. The prior art renders obvious “encoding the first mixed audio data into a first compressed audio data before transmitting the first mixed audio data to the remote client” as required by claim 6.

280. As discussed above in connection with limitations [1.4]-[1.5], Knappe teaches mixing “*the first mixed audio data*” and “*transmitting the first mixed audio data to the remote client.*”

281. Drell further teaches using a compression algorithm to compress mixed audio data before transmission to a remote endpoint. Drell (Ex. 1008) at 6:33-45 (“In the transit mode ... **[a]udio codec 408 encodes the audio data stream output by audio mixing module 308** ... using a standard or proprietary audio **compression algorithm** ... and delivers the encoded audio data”).

282. Drell also teaches that the communication process may include a standard process such as H.323. *See, e.g. id* at 2:26-28 (“The communication process, which may for example comprise an...**H.323 (packet-based) process.**”).

It would have been obvious to a POSITA that performing H.323 processing includes utilizing the compression protocols and codecs described in the H.323 standard because they were commonly used (and some were required). *See, e.g.,* U.S. Patent No. 6,553,515 (Ex. 1028) at 4:35-56 (“**H.323** is adapted to support many aspects of multimedia...[and] generally is regarded as the **standard for VoIP**...This standard also specifies lowest-common denominator **compression protocols, such as G.711...G.723 and G.729.**”); U.S. 6,850,496 (Ex. 1027) at 3:56-67 (“[A] codec (compressor/ decompressor) algorithm [] will be used to transmit capture channel data from encoder 24 to decoder 30. The codec may use a technique as simple as Pulse-Code Modulation (PCM), or a very complex technique, e.g., one that uses subband coding and/or predictive coding to decrease bandwidth requirements... Many codecs have been standardized and are well

known to those skilled in the art, and the particular codec selected is not critical to the operation of the invention.”). In fact, the ’332 patent’s encoder is simply a well-known prior art H.323 codec. *See, e.g.*, ’332 patent (Ex. 1001) at 2:56-62, FIG. 2 (“The H.323 protocol stack consists of an audio codec 214, and a video CoDec 213 connected to the audio/video input and output blocks 452 and 451, respectively. The audio and video CoDecs are responsible for compression and decompression of the audio and video sources.”).

283. Since Knappe also discloses “a codec (coder/decoder) converts audio,” it would have been obvious to a POSITA to use in Knappe a codec that compresses the mixed audio data before transmission to a remote communications device 16 because this would result in the known benefit of reducing the amount of data transmitted over the network.

284. Thus, the prior art teaches “*encoding the first mixed audio data into a first compressed audio data before transmitting the first mixed audio data to the remote client*,” rendering claim 6 obvious.

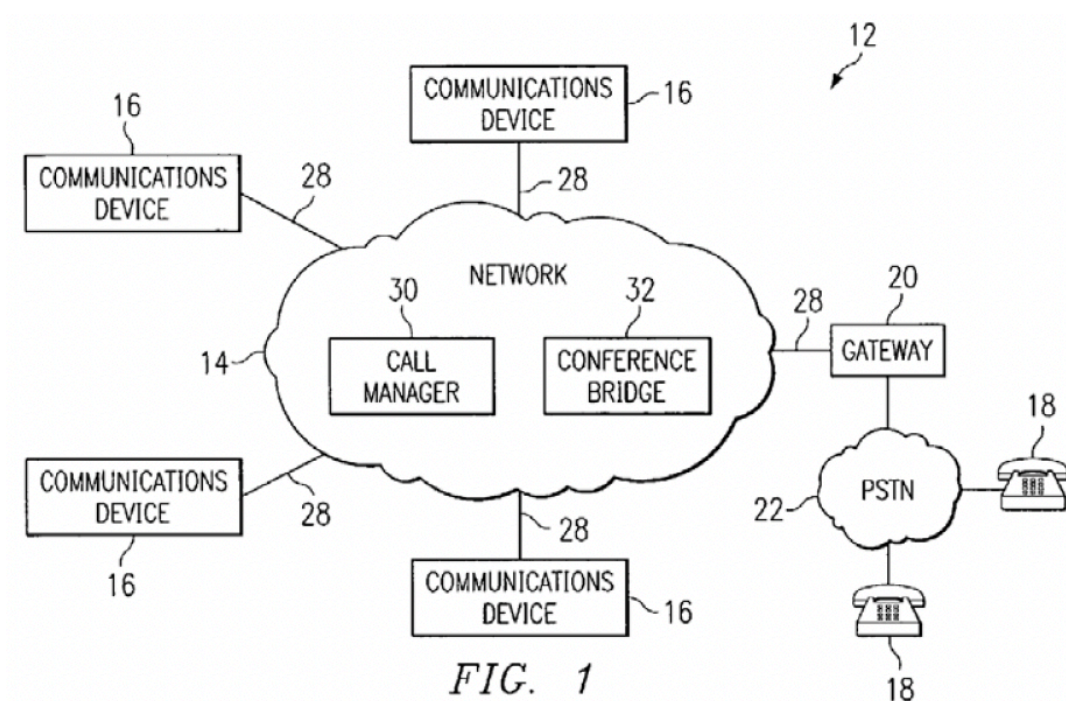
5. Claim 8

[8.P] A system for supporting a multi-participant audio/video conference call comprising:

285. Knappe discloses “a system for supporting a multi-participant audio/video conference call” and thus discloses the preamble of claim 8.

286. As discussed above in connection with the preamble [1.P] of claim 1, Knappe discloses “a method and system for improving ... a multiparty communication session ... [that] allow[s] multiple people to participate in ... a ‘conference call.’” Knappe (Ex. 1006) at 1:8-27, 8:46-10:50, FIG. 8.

287. As shown in Figure 1 of Knappe, Knappe’s system can support multiple participants in an audio/video conference call:



Specifically, Knappe explains that “conference bridge 32 provides conference call and other suitable audio, video, and/or real-time multi-party communication sessions between communication devices 16.” *Id.* at 4:58-61. Thus, a POSITA would understand Knappe to support a multi-participate audio/video conference call.

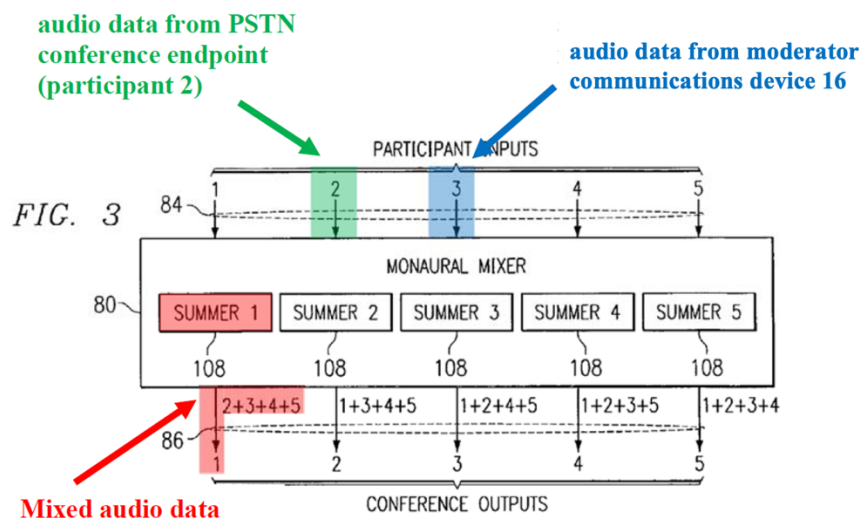
288. Thereby, Knappe teaches the preamble of claim 8.

[8.1.1] a first mixer configured to mix audio data from a Public Switched Telephone Network (PSTN) client with audio data from a moderator;

289. The prior art renders obvious “a first mixer configured to mix audio data from a Public Switched Telephone Network (PSTN) client with audio data from a moderator and collaboration data from the moderator into a mixed data.”

290. As discussed at [1.1], Knappe discloses receiving audio data from a PSTN client. *See supra* ¶¶ 166-173. As discussed at [1.2], Knappe discloses receiving audio from participant communications devices 16, one of which can be designated as a moderator. *See supra* ¶¶ 174-179.

291. As discussed above in connection with limitation [1.4], Knappe discloses “*mixing the first audio data from the PSTN client with the second audio data from the moderator into a first mixed audio data.*” Specifically, as discussed in [1.4] and shown in annotated Figure 1 below, mixers 58, 80, and 100 of Knappe’s conference bridge 32 mixes audio data from PSTN telephone 18 with audio data from the moderator communication device 16:



Knappe (Ex. 1006) at FIG. 3 (annotated).

292. In addition, as discussed at [3.1], Drell teaches a conferencing system that includes an MP conferencing application 234 (which, as discussed at [3.1], can be implemented at Knappe's moderator communications device 16). Drell's conferencing application 234 includes an audio mixing module 308:

MP conferencing application 234 includes audio mixing module 308

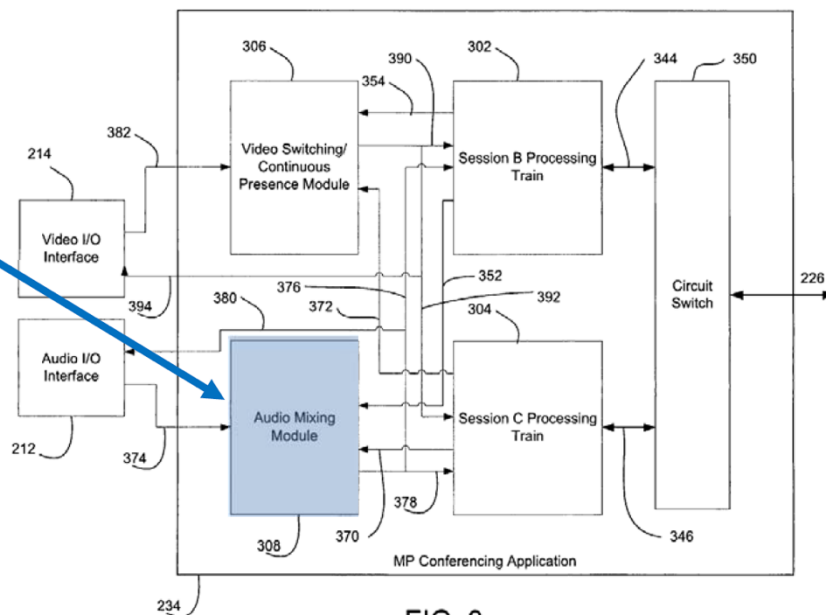


FIG. 3

Drell (Ex. 1008) at FIG. 3 (annotated). The “[a]udio mixing module 308 is configured to combine audio data received from remote conference endpoints . . . with locally generated audio data . . . [where] audio mixing module 308 may simply mix the received audio input data streams.” *Id.* at 5:8-23; *see also id.* at Abstract, 2:34-52, 3:45-48, 4:25-31. Thus, a POSITA would understand that the audio mixing module 308 mixes the locally generated audio (which as discussed at [3.1] can be Knappe’s moderator communications device 16) with audio data received from a remote PSTN conference endpoint.

293. Thus, a POSITA would have understood the prior art to teach “*a first mixer configured to mix audio data from a Public Switched Telephone Network (PSTN) client with audio data from a moderator*” thereby rendering this limitation obvious.

[8.1.2] and collaboration data from the moderator into a mixed data

294. The prior art renders obvious that the mixed audio data is further mixed with collaboration data from the moderator into a mixed data.

295. As discussed at [1.3.1], a POSITA would understand that the “audio, video, voice, data and other suitable types of real-time . . . traffic” communicated between conference participants in Knappe would include data to be used for collaboration, including well-known collaboration data such as T.120 standard data. As further discussed at [1.3.1], Elliott teaches that collaboration services, such as collaborative audio and video conferencing, can be provided over the Internet, such that a POSITA would understand that Knappe’s conferencing system could be implemented with Elliott’s teachings to include sharing of collaboration data.

296. Drell also teaches that the conferencing application 234 has a processing train 302 that utilizes a data codec to encode “still images” for transmission to conference endpoints. Ex._1008, 6:53-56 (“processing train 302 may include a **data codec for coding and encoding still images** and the like received from and **transmitted to remote conference endpoints**”). A POSITA would understand Drell’s “still images” which are encoded with a “data codec” to be “*collaboration data*.” See, e.g., T.120 Primer (Ex. 1016) at 10 (“Still Image Exchange and Annotation (SI) – T.126”); T.120 Standard (Ex. 1017) at 2 (“This

Recommendation makes provisions to ensure interoperability of commonly required functionality such as ... still image exchange ... through the definition of standardized application protocols”). It would have been obvious to a POSITA for Knappe’s moderator communications device 16 (which as discussed at [3.1] may implement Drell’s conferencing application 234) to also be the source of still images, since this would allow for greater variety of collaboration among the participants.

297. Drell also teaches that its communication process may include a standard process such as H.323. *See, e.g. id* at 2:26-28 (“The communication process, which may for example comprise an...**H.323 (packet-based) process.**”). It would have been obvious to a POSITA that employing an H.323 process also discloses “*collaboration data*,” because H.323 was a well-known standard that recommended T.120 collaboration data. *See, e.g.,* H.323 Standard (Ex. 1033) at 15 (“The standardized data application for real-time audiographics conferencing is ITU-T T.120.”) & 18 (“ITU-T Rec. T.120 is the default basis of data interoperability between an H.323 terminal and other H.323, H.324, H.320, or H.310 terminals.”); T.120 Primer (Ex. 1016) at 13-15; T.120 Standard (Ex. 1017). Thus, a POSITA would have understood that Knappe’s conferencing system could be implemented with Drell’s and Elliott’s teachings to include “*collaboration data from the moderator.*”

298. Drell further explains the that encoded data is multiplexed and output as a single stream by the communication process:

The combined audio and video data streams are conveyed as input to each processing train for encoding and transmission to the corresponding conference endpoints. **In the send mode, the audio and video signals are encoded by the audio/video codecs and multiplexed into a single data stream by the communication process.**

Drell (Ex. 1008) at 2:43-48; *see also id.* at 6:46-48, FIG. 4.

299. Since the communication process outputs a “single stream,” it would have been obvious to a POSITA that Drell’s collaboration data (such as the encoded still images) are combined with the mixed audio data (from the audio mixing module, discussed at [8.1.1]) for transmission to remote conference endpoints. Annotated Figure 3 of Drell, below, highlights the audio mixing module 308 and the processing train 302, which a POSITA would have understood to encode and mix “the audio and video data streams” with collaboration data, such as “still images,” to generate mixed data:

MP conferencing application 234 includes audio mixing module 308 and processing train 302

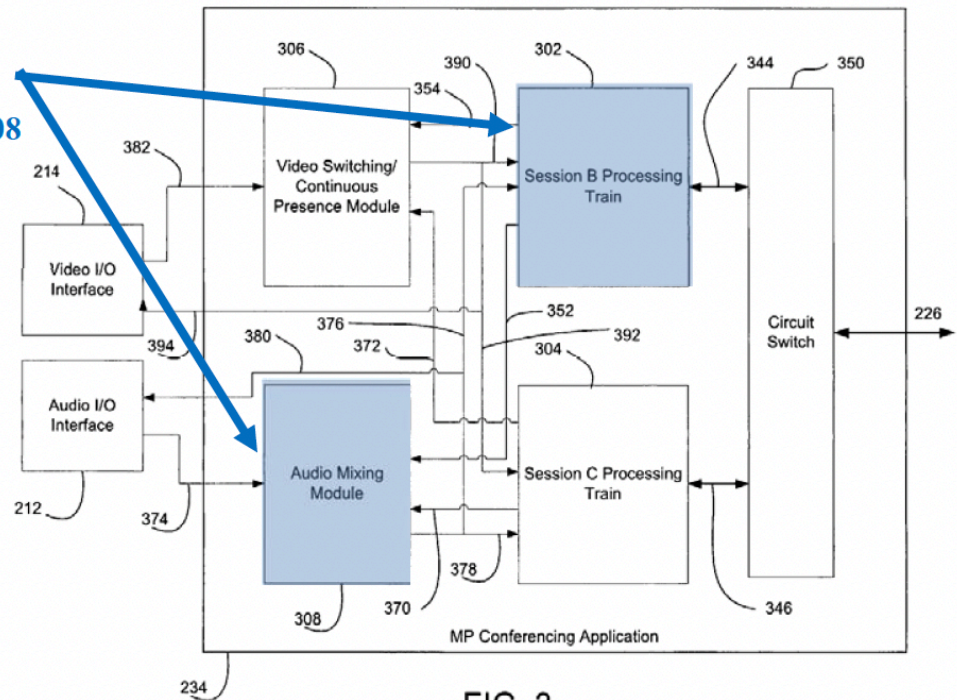


FIG. 3

Id. at FIG. 3 (annotated).

300. Thus, the prior art teaches using a processing train that further mixes the mixed audio data with collaboration data (e.g., still image data, application data, web browsing data, and other collaboration data per the T.120 standard) from the moderator into a mixed data, which renders obvious this limitation.

[8.2] a transport output configured to transmit the first mixed data to at least one remote client through a first Virtual Private Network (VPN) tunnel;

301. The prior art renders obvious the “a transport output configured to transmit the first mixed data to at least one remote client through a first Virtual Private Network (VPN) tunnel.”

302. As discussed above in connection with limitation [8.1.1]-[8.1.2], Knappe, in view of Elliott and Drell, teaches generating mixed data.

303. As a threshold matter, the phrase “*first mixed data*” is not previously recited in the claim, which previously only recited “*mixed data*” and later in limitation [8.3] recites “*first mixed audio data*.” For the purposes of analysis for this IPR only, I assume that “*first mixed data*” corresponds to the “*mixed data*” discussed in connection with limitation [8.1.1] and [8.1.2] above.

304. As discussed above in connection with limitation [1.5], Knappe discloses “transmitting audio ... between source and destination endpoints,” including “a conference output stream for delivery to the participant.” Knappe (Ex. 1006) at 3:24-29; 6:65-67; *see also id.* at 5:35-39 (“The conference output stream for a participant includes the media of all other participants, a subset of other participants or other suitable mix”).

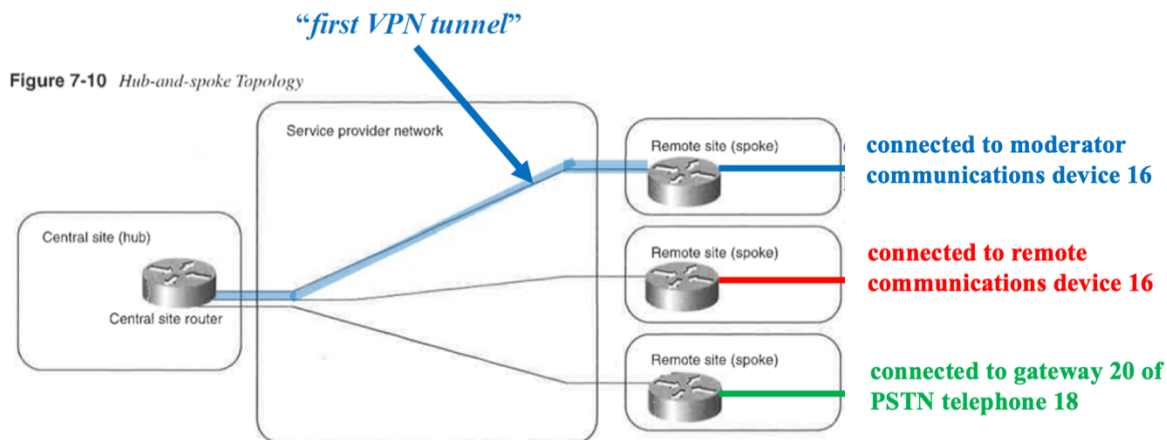
305. Drell also discloses that its “processing trains ... process the combined signals according to a transmit mode for distribution to the remote endpoints over the network.” Drell (Ex. 1008) at Abstract, 5:48-54, 6:53-56. Therefore, Drell’s processing trains are able to transmit “*the first mixed data*” of limitation [8.1.2] by the combination of Knappe, Drell, and Elliott.

306. Thus, a POSITA would understand that when the collaboration data is mixed with the mixed audio data, as discussed above in connection with limitation

[8.1.2], the mixed data could be transmitted via a transport output as taught in Knappe and Drell. *Id.* at FIGS. 1 & 3; Knappe (Ex. 1006) at FIG. 1.

307. As discussed above in connection with limitations [1.3.2] and [1.5], it would have been obvious to a POSITA to use VPN tunneling when communicating audio, video, and collaboration data to provide the Knappe system with secure communications over Internet network 14.

308. As noted at [1.3.2] when implementing VPN tunneling in Knappe, per Elliott, it would have been obvious to a POSITA to connect each of Knappe's devices located at remote sites to a separate remote router (spoke) that is connected to a router (hub), as the VPN Textbook teaches, because this common topology provides a cost effective and less complex implementation. Accordingly, it would have been obvious to a POSITA that when implementing the commonly used hub-and-spoke VPN topology, per the VPN Textbook, Knappe's remote communications device 16, moderator communications device 16, and gateway 20 connected to the PSTN telephone 18 would be connected to different remote site (spoke) routers, with a separate central site (hub) router, as illustrated below in annotated Figure 7-10:



VPN Textbook (Ex. 1010) Figure 7-10 (annotated).

309. As shown, the remote site router (spoke) of the moderator communications device 16, which as discussed at [3.1] can implement the conference bridge mixing functionality as software per Drell, can transmit data to the remote communications device 16 through a central hub using a “first VPN tunnel.” Thus, a POSITA would have understood Knappe (in combination with the teachings in Drell) in view of Elliott and/or the VPN Textbook to also teach the further limitation of transmitting the first mixed data *“through a first Virtual Private Network (VPN) tunnel.”*

[8.3] a second mixer configured to mix audio data from the moderator with audio data from the remote client connected to the Internet into a first mixed audio data; and

310. The prior art renders obvious “a second mixer configured to mix audio data from the moderator with audio data from the remote client connected to the Internet into a first mixed audio data.”

311. As discussed above in connection with [1.3.1], Knappe discloses receiving audio data from a remote client. As discussed at [1.6], Knappe discloses “*mixing ... audio data from the moderator with ... audio data from the remote client into a ... mixed audio data.*” As shown and discussed in connection with limitation [1.6], this mixing of audio data may be accomplished by one or more of Knappe’s mixers 58, 80, and 100, thereby rendering this limitation obvious.

[8.4] a VoIP encoder to encode the first mixed audio data into a compressed audio data;

312. The prior art renders obvious the “a VoIP encoder to encode the first mixed audio data into a compressed audio data” limitation.

313. As discussed above in connection with limitation [7.1], Knappe discloses “*VoIP encoding the second mixed audio data,*” which is audio data from the moderator mixed with audio data from the remote client. As noted, a POSITA would have understood Knappe’s disclosure of “[t]elephony voice information [being] transmitted in the Voice over IP (VoIP) format,” the “VoIP packets [being] encapsulated in real-time transport protocol (RTP) packets ... or other suitable packets for transmission over the network 14” to indicate “*a VoIP encoder*” encoding said VoIP packets. Knappe (Ex. 1006) at 3:54-58.

314. As discussed above in connection with limitation [6.1], Drell teaches using a compression algorithm to compress mixed audio data before transmission to a remote endpoint. Drell (Ex. 1008) at 6:18-45 (disclosing “[a]udio codec” capable of encoding/decoding “in accordance with a standard ... or proprietary audio compression algorithm”).

315. A POSITA would have recognized the benefit of compressing the mixed audio data generated using Knappe’s mixers (e.g., mixers 58, 80, and 100) for bandwidth savings in transmission.

316. Therefore, a POSITA would have understood the combination of Knappe and Drell to disclose “*a VoIP encoder to encode the first mixed audio data into a compressed audio data,*” as claimed, thereby rendering this limitation obvious.

[8.5] the transport output further configured to transmit the compressed audio data to the PSTN client through the first VPN tunnel.

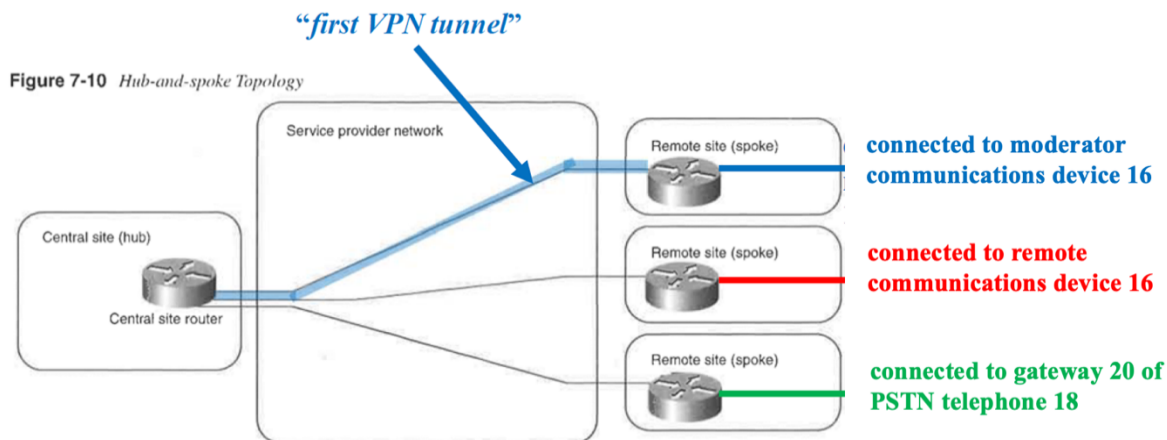
317. The prior art renders obvious the “the transport output further configured to transmit the compressed audio data to the PSTN client through the first VPN tunnel.”

318. As discussed above in connection with limitations [7.1] and [8.4], Knappe and Drell in combination disclose a VoIP encoder that encodes the first mixed audio data (audio data from the moderator mixed with audio data from the remote client) into “*the compressed audio data.*” As discussed above in connection

with limitation [1.7], Knappe in view of Elliott and the VPN Textbook discloses transmitting the mixed audio data (the audio data from the moderator mixed with audio data from the remote client) to the PSTN.

319. Specifically, Drell teaches that “[a]udio code 408 encodes the audio data stream output by audio mixing module 308 . . . using a standard or proprietary compression algorithm . . . and delivers the encoded audio data to communication process 404” for transmission. Drell (Ex. 1008) at 6:39-45; *see also id.* at 6:46-62, FIG. 5 (describing process for transmission of data following communication process 404).

320. Further, as discussed at [1.3.2] and [8.2], it would have been obvious to a POSITA to implement VPN tunneling when transmitting data in the Knappe system. For the reasons shown and discussed at [1.3.2] and [8.2], the remote site router (spoke) of the moderator communications device 16, which as discussed at [3.1] can implement the conference bridge mixing functionality as software per Drell, can transmit data to the PSTN client through a central hub using the “first VPN tunnel”:



VPN Textbook (Ex. 1010) Figure 7-10 (annotated).

321. Thus, a POSITA would have understood Knappe (in combination with the teachings in Drell) in view of Elliott and/or the VPN Textbook to also teach the further limitation of transmitting the compressed audio data to the PSTN client “through a first Virtual Private Network (VPN) tunnel.”

6. Claim 9

[9.P] The system of claim 8 further comprising:

322. See analysis at [8.P]-[8.5].

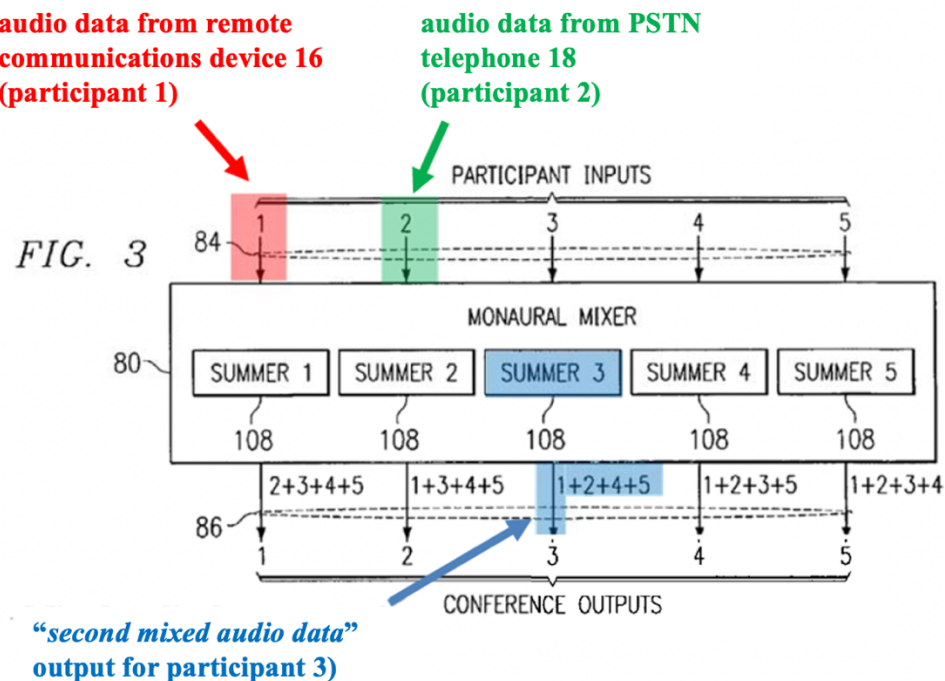
[9.1] a third mixer configured to mix the audio data from the PSTN client with the audio data from the remote client into a second mixed audio data; and

323. The prior art renders obvious the “a third mixer configured to mix the audio data from the PSTN client with the audio data from the remote client into a second mixed audio data” limitation.

324. As discussed above in connection with limitation [2.1], Knappe discloses “mixing the ... audio data from the PSTN client with the ... audio data from the remote client into a ... mixed audio data.”

325. As shown and discussed above in connection with limitations [1.2], [1.4], and [2.1], this mixing of audio data may be accomplished by one or more of Knappe’s mixers 58, 80, and 100.

326. For example, as shown in annotated Figure 3 below, Knappe discloses mixing “audio data from the PSTN client” (e.g. participant 2) with “the audio data from the remote client” (e.g., participant 1), and providing that “mixed audio data” to “the moderator” (e.g., participant 3):



Knappe (Ex. 1006) at FIG. 3 (annotated).

327. Thus, the prior art renders obvious “a third mixer configured to mix the audio data from the PSTN client with the audio data from the remote client into a second mixed audio data.”

[9.2] a speaker configured to play the second mixed audio data to the moderator.

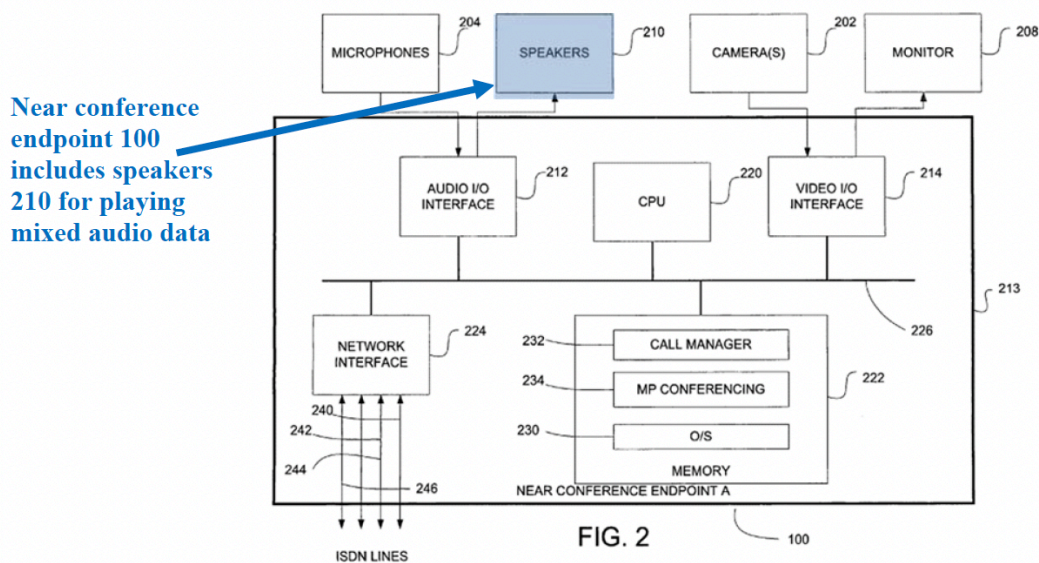
328. The prior art teaches the “a speaker configured to play the second mixed audio data to the moderator” limitation of claim 9.

329. As discussed above in connection with limitation [9.1], Knappe discloses generating “*the second mixed audio data.*”

330. Drell further discloses a “loudspeaker 210.” Drell (Ex. 1008) at 3:45-46, FIGS. 2-3.

331. As shown and discussed above in connection with limitation [3.1], Drell’s near conference endpoint A, including its MP conferencing application 234, audio I/O interface 212, and other components, may be implemented with Knappe’s conferencing system.

332. Drell’s Figure 2 shows audio I/O interface 212 coupled to loudspeaker 210 and MP conferencing application 234, which has been shown above, as well as in the below Figure 3, to have audio mixing capabilities:



Drell (Ex. 1008) at FIG. 2 (annotated).

333. Drell's Figure 3 shows that same audio I/O interface 212 able to receive mixed audio from audio mixing module 308 in MP conferencing application 234:

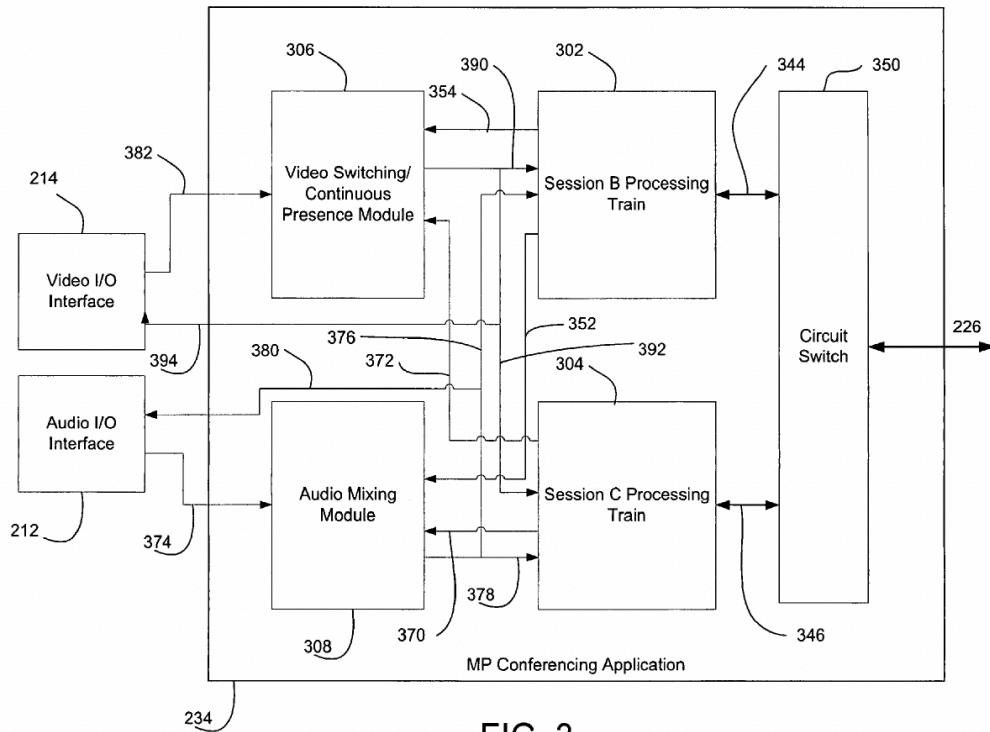


FIG. 3

Id. at FIG. 3.

334. Thus, a POSITA would have understood Knappe in combination with Drell to teach “a speaker configured to play the second mixed audio data to the moderator,” as claimed, thereby rendering this claim obvious.

7. Claim 10 Drell

[10.P] The system of claim 8 further comprising

335. See analysis at [8.P]-[8.5].

[10.1] a VoIP decoder configured to receive IP packets containing the audio data from the PSTN client, and further configured to decode the IP packets containing the audio data from the PSTN client.

336. The prior art renders obvious the “a VoIP decoder configured to receive IP packets containing the audio data from the PSTN client, and further

configured to decode the IP packets containing the audio data from the PSTN client” limitation of claim 10.

337. As discussed above in connection with limitation [3.1], Knappe discloses a “codec (coder/decoder)” for “*receiving first audio data from the PSTN client include[ing] receiving IP packets and decoding the IP packets containing the first audio data from the PSTN client.*” See Knappe (Ex. 1006) at 4:5-27, 6:3-21 (describing a “codec (coder/decoder)” to encode data that “is encapsulated into IP ... packets for transmission over the network 14” and “converters 54” to “receive input packets ... and convert the packet from the native format of the participants device 16 to a standard format of the conference bridge 32 for cross linking and processing”); *see also* Caputo (Ex. 1021) at 11.

338. Also, as discussed above in connection with limitation [7.1], Knappe discloses “*VoIP encoding the second mixed audio data before transmitting the ... mixed audio data.*” See Knappe (Ex. 1006) at 3:54-56 (“Telephony voice information is transmitted in the Voice over IP (VoIP) format.”).

339. Therefore, a POSITA would have understood Knappe to disclose “*a VoIP decoder configured to receive IP packets containing the audio data from the PSTN client, and further configured to decode the IP packets containing the audio data from the PSTN client,*” thereby rendering this limitation obvious.

8. Claim 11

[11.P] The system of claim 8 further comprising

340. See analysis at [8.P]-[8.5].

[11.1] at least one audio decompressor configured to decode IP packets containing the audio data from the remote client.

341. The prior art renders obvious the “at least one audio decompressor configured to decode IP packets containing the audio data from the remote client” limitation of claim 11.

342. As discussed at [1.3.1], Knappe discloses receiving audio data from a remote client. As discussed at [1.6], Knappe discloses “*mixing the second audio data from the moderator with the third audio data from the remote client into a ... mixed audio data.*”

343. Knappe also discloses that data, including audio from remote communications device 16, is transmitted as IP packets. *E.g.* Knappe (Es. 1006) at 4:3:53-54 (“In the Internet embodiment, the network 14 transmits information in Internet Protocol (IP) packets”); 4:9-17 (“digital encoded data is encapsulated into IP or other suitable packets for transmission over network 14”).

344. Knappe also discloses converting audio data into and back from IP packets. *Id.* at 4:13-14 (“IP packets received from the network 14 are converted back into analog signals and played to the user”); 6:16-21 (“The input converters receive input packets of a participant . . . and convert the packet from the native

format of the participant's device 16 to a standard format[.]. Conversely, the output converters . . . convert the conference output streams to the native format of participant's devices.”).

345. Also, as discussed above in connection with limitations [3.1] and [4.1], it would have been obvious to a POSITA that Knappe's conversion of IP packets would include “*decoding the IP packets containing the first audio data from the PSTN client.*” See Caputo (Ex. 1021) at 11 (disclosing in well-known art “the transport of voice traffic over an IP network requires a conversion of the signal from analog to digital, packetization of the digital voice data, transport of the packetized information through the network, de-packetization of the voice data, and conversion of the digital voice data back to an analog signal.”).

346. In addition, as discussed at [3.1], Drell also explains that conferencing application 234 (which can be implemented at Knappe's moderator communications device 16) uses audio codec 408 to decode packets. Drell (Ex. 1008) at 6:18-22 (“[a]udio codec 408 [that] receives the audio data stream ... and applies ... decoding in accordance with a standard ... or proprietary audio compression algorithm. ”). A POSITA would have understood that decoding audio using an audio compression algorithm is performing decompression and thus discloses an “*audio decompressor*” as claimed.

347. Drell further explains that the communication process of the conferencing application may comprise an “H.323 (packet-based) process” where “audio and video signals are thereafter respectively delivered to audio and video codecs for decoding.” *Id.* at 2:26-33. A POSITA would understand Drell’s H.323 process to include an “*audio decompressor*” because H.323 was a well-known standard that specified certain compression protocols. *See, e.g.* U.S. Patent No. 6,553,515 (Ex. 1025) at 4:35 56 (“H.323 is adapted to support many aspects of multimedia...[and] generally is regarded as the standard for VoIP...This standard also specifies lowest common denominator compression protocols, such as G.711...G.723 and G.729.”).

348. Thus, a POSITA would have understood the prior art to render obvious “*at least one audio decompressor configured to decode IP packets containing the audio data from the remote client.*”

C. Claim 12 Is Obvious Under 35 U.S.C. § 103(a) Over Knappe in View of Elliott, the VPN Textbook, Drell and Hoke

1. Summary of Hoke

349. United States Patent No. 6,701,437 (“Hoke”) entitled “Method and Apparatus for Processing Communications in a Virtual Private Network” was filed November 9, 1998. *See* Hoke (Ex. 1009).

350. Hoke describes “computer systems for processing communications in a virtual private network.” *Id.* at Abstract. Hoke further describes the compression

and decompression of data communicated over VPN. *See, e.g. id.* at FIGS. 3-4 & 7, 2:50, 5:40-42 (“Many of the elements may be configured using well-known structures, particularly those designated as relating to various compression or encryption techniques.”).

351. Hoke explains that “[w]hen VPN traffic is sent between source and destination endstations ..., the VPN unit serving the source endstation processes the data packet, encrypts it, compresses it (if necessary), and ... [t]he receiving VPN unit authenticates, decrypts and decompresses the packet (as necessary) before forwarding it toward the destination endstation.” *Id.* at 8:38-50.

2. Reasons to Combine Knappe and Hoke

352. A POSITA would have been motivated to combine the teachings of Knappe and Hoke to produce the obvious, beneficial, and predictable results of compressing data before it is transmitted over a VPN tunnel.

353. As discussed above, Knappe teaches a communications system capable of transmitting conferencing data—“audio, video, voice, data and other suitable types of ... traffic”—between source and destination endpoints. Knappe (Ex. 1006) at 3:26-29.

354. Also in the context of conferencing, Elliott teaches the sharing of collaboration data, including “collaborative Web browsing, audio conferencing, video conferencing, and application sharing” and the use of “a virtual private

network (VPN) ... which uses IP or IP-tunneling” for the transmission of “voice, fax, data, video” and “[o]ther types of collaboration ... data” during a conference. Elliott (Ex. 1007) at 1:43-44 2:66-3:1, 4:29-34, 6:53-54.

355. As I discussed above, a POSITA would have been motivated to combine the conferencing system of Knappe to include sharing of collaboration data through a VPN as taught in Elliot and the VPN Textbook. Thus, in considering Knappe, in view of Elliott and the VPN Textbook, a POSITA would have considered the teachings of Hoke, as well, for teachings relevant to VPN tunneling implementations.

356. In particular, a POSITA would have recognized the desirability of compressing data before it is transmitted over a network, as it was well known in the art that compressing data would reduce bandwidth usage and improve performance. See Caputo (Ex. 1021) at 12-13 (“compression algorithms can be used to reduce the amount of voice data that is transmitted over the packet switched network. ... This reduction helps compensate for ... limited bandwidth.”); T.120 Primer (Ex. 1016) at 10 (“compressed to improve performance over lower-speed communication links.”).

357. As Hoke describes the compression and decompression of data communicated over VPN, it would have been obvious to a POSITA to implement the compression techniques in Hoke with an implementation of Knappe having the

VPN tunneling of Elliott and the VPN Textbook. *See* Hoke (Ex. 1009) at FIGS. 3-4 & 7, 2:50, 5:40-42.

358. For example, Hoke's description of a VPN unit that performs both encryption and compression would have suggested to a POSITA to include a compression feature in the VPN routers described in the VPN Textbook.

359. It would have been obvious to a POSITA to compress data transmitted over VPN tunnels implemented in Knappe, in view of Elliott and the VPN Textbook, to minimize bandwidth usage and improve performance, particularly for lower speed links. In addition, bandwidth carries a monthly cost that increases with higher bandwidth, but compression is a one-time equipment cost, so a POSITA would be motivated to trade off more capable VPN routers with compression to save money over the long term.

360. A POSITA would expect this application of known compression algorithms with known VPN techniques to successfully yield predictable and beneficial results. *See* Hoke (Ex. 1009) at 10:16-18 ("Many compression algorithms are known"), 5:41-48 ("Many of the elements may be configured using well-known structures, particularly those designated as relating to various compression or encryption techniques. ... those of ordinary skill in the art will be able to implement the particular methods without undue experimentation.").

361. Any modifications needed to be made to Knappe's system to accommodate the teachings of Hoke would have been within the level of ordinary skill in the art. The proposed combination permits, but does not require, physical incorporation of elements from Hoke into Knappe.

3. Claim 12

[12.P] The system of claim 8 further comprising

362. See analysis at [8.P]-[8.5].

[12.1] an encoder configured to encode the mixed data into a compressed data before transmitting the compressed data to the remote client.

363. The prior art renders obvious the “an encoder configured to encode the mixed data into a compressed data before transmitting the compressed data to the remote client” limitation of claim 12.

364. As discussed above in connection with limitations [6.1] and [8.4], Knappe, in combination with Drell, discloses a VoIP encoder configured to encode mixed audio data into a compressed audio data before transmitting the compressed data to the remote client.

365. As discussed above in connection with limitations [8.1.1]-[8.1.2] above, Knappe and Drell in combination with Elliott also disclose the generation of “*the mixed data.*”

366. To the extent determined that the claim requires applying a single compression process to a mixture of audio and collaboration data, Hoke teaches

compressing data that is transmitted over VPN tunnels. Specifically, Hoke teaches that “VPN units may transmit **VPN traffic in... tunnel...** format.” Hoke (Ex. 1009) at 9:6-10. In tunnel mode, “a data packet is routed to/from the VPN units serving the communicating endstations or remote clients.” *Id.* at 9:10-12. In this context, Hoke teaches that a “VPN unit serving the source endstation processes the data packet, encrypts it, **compresses it** (if necessary)[.]” *Id.* at 8:39-44; *see also id.* at 10:13-16 (“[T]he lookup table associated with a particular VPN will indicate whether or not data packets transferred between members of the VPN are to be compressed and, if so, the specific compression algorithm to be used.”); *see also id.* at 11:6-14, FIG. 3.

367. Hoke further teaches that “[m]any compression algorithms are known, but in one embodiment of the invention LZW compression is used.” *Id.* at 10:16-18. The compression algorithm, as taught by Hoke, corresponds to the claimed “*encoder*.”

368. It would have been obvious to a POSITA to compress the mixed data before it is transmitted over the first VPN tunnel (as discussed at [8.2]), because compressing the mixed data would minimize the utilized bandwidth and improve the perceived performance of lower-speed links. *See, e.g.,* Caputo (Ex. 1021) at 12-13; T.120 Primer (Ex. 1016) at 10; *see also* Reasons to Combine Drell and Hoke, *supra* ¶¶ 352-361.

369. In combination, packetized mixed data from Knappe's moderator communications device 16 would be sent to a router (as taught by the VPN Textbook), as discussed at [8.2], and as described in this section would apply a compression algorithm (as taught by Hoke) before transmission over the first VPN tunnel.

370. Thus, the prior art teaches that the mixed audio data from the moderator communications device 16 is provided to the VPN router, which compressed the mixed data before transmitting the compressed data over the VPN tunnel to the remote communications device 16, which renders obvious this limitation.

IX. SECONDARY CONSIDERATIONS

371. Based on my experience and knowledge of this field, I am not aware of any secondary considerations that would provide any evidence of non-obviousness as to any of the claims in the '332 patent.

372. I am not aware of any allegations made by the named inventor of the '332 patent or any assignee of the '332 patent that any secondary considerations tend to rebut the obviousness of any of the claims of the '332 patent.

373. I reserve the right to respond to any evidence or arguments regarding secondary considerations of non-obviousness the patent owner may raise in its response.

X. DECLARATION

374. I hereby declare that, to the best of my knowledge and understanding, and based on the materials available to me at the time of signing this Declaration, all statements and opinions in this Declaration are true or believed to be true. I further declare that all statements made in this Declaration of my own knowledge are true, and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001.

Executed this 14th day of December 2022.

By: Henry H. Houh.
Henry H. Houh, Ph.D.