

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. 9,843,612

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. 9,843,612

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. 9,843,612 (“the ’612 patent”), entitled “Voice Conference 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 ’612

patent, including how a person of ordinary skill in the art (POSITA) would have interpreted the '612 patent and the prior art I discuss herein.

6. I have also been asked to provide my opinions regarding the validity of claims 10-29 of the '612 patent ("Challenged Claims"), including what a POSITA as of the earliest possible priority date of the '612 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 '612 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. 6,798,753 to Doganata, *et al.* ("Doganata") (Ex. 1039)

- WO2001018665A1 to Hendricks, *et al.* (“Hendricks”) (Ex. 1040); and

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

- the prosecution history of the '612 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 (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. 6,275,588 (Ex. 1027);
- U.S. Patent No. 7,882,247 (Ex. 1028);
- U.S. Patent No. 6,055,575 (Ex. 1029);
- U.S. Patent No. 5,606,550 (Ex. 1035);
- U.S. Patent No. 6,768,914 (Ex. 1041);
- U.S. Patent No. 6,501,740 (Ex. 1042);
- U.S. Patent No. 6,424,647 (Ex. 1043);
- U.S. Patent No. 6,961,416 (Ex. 1044);
- U.S. Patent No. 6,011,851 (Ex. 1045); 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 person of ordinary skill in the art (“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 ’612 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 a POSITA’s skill level as of the earliest claimed priority date.

27. I have been informed by counsel that the earliest claimed priority date for the ’612 patent is March 10, 2003. Based on the technologies disclosed in the

'612 patent, I believe that a POSITA would include someone who had, at the earliest claimed priority date of the '612 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 '612 patent, and that I am qualified to opine on the subject matter in the '612 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 '612 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 '612 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, in order 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 '612 PATENT

48. The '612 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 1:7-16.

A. The '612 Patent Specification

49. The '612 patent relates “to computer system architectures and more particularly to audio and video telecommunications for collaboration over hybrid networks.” *Id.* at 1:21-23.

50. With reference to Figure 3 shown below, the '612 patent describes a hybrid system that includes an Internet network 435 and a Public Switched Telephone Network (PSTN) network 433. *Id.* at 5:10-21, FIG. 3.

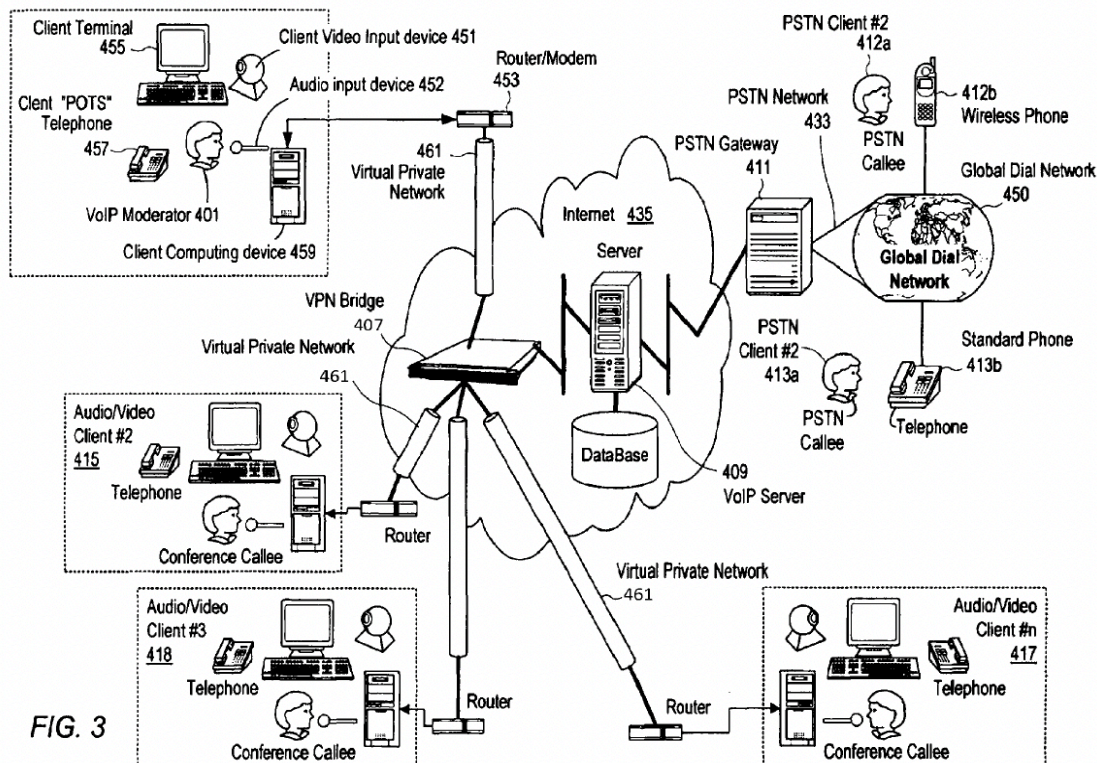


FIG. 3

Id. at FIG. 3.

51. In the above figure, the Internet network 435 includes “moderator 401,” “audio video client 415, audio video client 418, and audio video client number and 417.” *Id.* at 5:1-13. According to the ’612 patent, “the Internet-based clients 401, 415, 418, and 417 are connected through routers or modems 453 preferably in a virtual private network configuration 461” with the “VPN bridge 407” located in the center. *Id.* at 5:22-25. The ’612 patent explains that this forms a “star topology...[with] VPN bridge(s) at the center of the network topology.” *Id.* at 5:45-48.

52. Again, referring to Figure 3, the ’612 patent further discloses that the PSTN network includes “telephony clients using standard wired 413 or wireless

telephone 412 systems.” *Id.* at 5:14-15. The PSTN clients are connected to the virtual private network 461 via PSTN gateway 411, “thus bridging the PSTN and VPN together.” *Id.* at 5:30-34, FIG. 3. The ’612 patent further explains, in connection with the more detailed diagrams in Figures 4 and 5, that a connection to the center VPN bridge 407 corresponds to a “VPN tunnel.” *See, e.g., id.* at 5:42-51, 6:40-62.

53. Regarding the data transmitted, the ’612 patent claims recite encoding/compressing and decoding/decompressing the conference data, but the only disclosure for performing the encoding and decoding is a reference to the prior art H.323 codec. *See, e.g., id.* at 2:63-3:1 (“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.”). Further, although the ’612 patent claims collaboration data, the only disclosure for collaboration data is a reference to the prior art T.120 standard. *Id.* at 3:17-22 (“Lastly, the H.323 recommendation defines a data collaboration capability as known and outlined in the T.120 data collaboration unit 225.”).

B. The Challenged Claims

54. Claims 10-29 are the Challenged Claims of the ’612 patent. Claims 10, 19, and 25 the independent claims. Independent claims 10 and 19 are method

claims directed to a method for establishing a secure multi-participant conference call including full duplex audio:

10. A method for establishing a secure multi-participant conference call including full duplex audio, the method comprising:
initiating a dial-in process to establish a connection with a local moderator client in response to an input from a Public Switched Telephone Network (PSTN) client;
receiving first audio data at the local moderator client from the PSTN client through a PSTN gateway;
receiving second audio data at the local moderator client from a remote client through a Virtual Private Network (VPN) tunnel;
transmitting third audio data, first video data, and first collaboration data from the local moderator client to the remote client through the first VPN tunnel; and
transmitting fourth audio data from the local moderator client to the PSTN client through the PSTN gateway.

19. A method for conducting a secure multi-participant conference call including full duplex audio, the method comprising:
receiving first audio data, first video data, and first collaboration data at a server from a first remote client through a first Virtual Private Network (VPN) tunnel;
receiving second audio data at the server from a second remote client through a second VPN tunnel;
receiving third audio data at the server from a Public Switched Telephone Network (PSTN) client through a PSTN gateway;
mixing the first audio data with the second audio data into a first mixed audio data;
mixing the second audio data with the third audio data into a second mixed audio data;
mixing the first audio data with the third audio data into a third mixed audio data;
transmitting the first mixed audio data from the server to the PSTN client through the PSTN gateway;
transmitting the second mixed audio data from the server to the first remote client through the first VPN tunnel;
transmitting the third mixed audio data from the server to the second remote client through the second VPN tunnel; and
transmitting the first video data and the first collaboration data from the server to the second remote client.

Id. at claims 10, 19.

55. Independent claim 25 is a system implementing the method of claim 19. *See id.* at 12:12-39. The dependent claims add additional limitations requiring receiving video and collaboration data at the local moderator client, a PSTN client that is a cellular phone, compressing/decompressing data, and mixing audio data. *See id.* at 10:65-11:21, 11:48-12:12, 12:40-49.

VII. CLAIM CONSTRUCTION

56. It is my understanding that in order to properly evaluate the '612 Patent, the terms of the claims must first be interpreted.

57. It is my understanding that for purposes of *Inter Partes* Review, the claim terms must be construed according to their ordinary and customary meaning

as would have been understood by one of ordinary skill in the art. I have also been informed that claim terms need to be construed only to the extent necessary to resolve the obviousness inquiry. It is my opinion that for purposes of applying the prior art presented herein to evaluate the patentability of the claims, no term requires express construction. Accordingly, I have analyzed the claims consistent with their ordinary and customary meaning as would have been understood by a POSITA.

VIII. SCOPE AND CONTENT OF THE PRIOR ART

A. Audio, Video, and Collaboration Data Conferencing Technology

58. 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 '612 patent – multiparty conferencing, including mixing and transmitting audio, video, and collaboration data over hybrid networks, was already well known.

59. 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.

60. 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.

61. Video conferencing systems were also well known before the '612 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.

62. The conferencing art that existed prior to the '612 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.

63. 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.

64. 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.

65. 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 ’612 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

66. 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.

67. 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 '612 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

68. 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.

69. 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.

70. 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.

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

D. Packet-Switched Networks

72. 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.

73. 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

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."

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.

74. 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.

75. 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

76. 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.

77. 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

78. 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.

79. 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.

80. 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.

81. 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.

82. 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

83. 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.

84. 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

85. 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.

86. 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.

87. 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

88. 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.

89. 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.

90. 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 simple (unidirectional) tunnels are established, and together the pair of tunnels form a full-duplex bi-directional encrypted tunnel.

6. Virtual Private Network

91. 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.

92. 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)

93. 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.

94. 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.

95. 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 '612 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 '612 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.

96. 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

97. 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.

98. The '612 patent incorporates by reference the entire H.261 and H.263 specifications which specify video codecs. *See* '612 patent (Ex. 1001) at 4:29-46. 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

99. 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 '612 patent, we released a line card that interfaced to a digital T1 line.

100. 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.

101. 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

102. 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

103. 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.*

104. 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

105. 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.*

106. 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.”).

107. 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; ’612 patent (Ex. 1001) at 3:18-21. 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).

F. Half-Duplex and Full-Duplex Communications

108. When a single shared communications channel exists among a number of endpoints, only one device may typically transmit. If multiple devices transmitted in the channel simultaneously, the signals would interfere, thus corrupting the transmission at the receiver. Accordingly, a device communicating

on the channel could either transmit or receive but not at the same time. This type of channel is known as a half-duplex.

109. Early versions of ethernet from the late 1980's and early 1990's, known as "10BASE2," "thin ethernet" or "thinnet" were half-duplex. This version of ethernet used coaxial cables to chain computers together, using "T" connectors to connect computers to the middle of a coaxial cable. Multiple host computers were connected to a single length of coaxial cable, but only one host could be transmitting at any given time. Though all hosts would receive the transmission, hosts would only process ethernet packets addressed to the host, or multicast/broadcast ethernet packets, and ignore ethernet packets addressed to other hosts.

110. A full-duplex channel is one on which a device can be simultaneously transmitting and receiving. It was well known before the earliest claimed priority date to utilize full duplex communications. For example, switched ethernet, which began appearing in the late 1990's, supported full duplex communications such that a host computer could both be transmitting and receiving at the same time. Mobile phone systems also allowed for full duplex voice communications. In order to do so, pair of half-duplex frequency channels were typically assigned to the call, one channel in each direction between each cell tower and each active mobile phone.

IX. INVALIDITY ANALYSIS

111. As noted previously, I have been asked to provide my opinion as to whether the Challenged Claims (claims 10-29) of the '612 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 '612 patent and render the Challenged Claims obvious to a POSITA.

112. 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 '612 patent: March 10, 2003. I have also considered the level of ordinary skill in the pertinent art as of that date.

113. 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 an limitation-by-limitation basis for each Challenged Claim of the '612 patent. This analysis supports my opinion that the differences between the claims of the '612 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 '612 patent to a person having ordinary skill in the art to which the subject matter pertains.

114. I will now describe, in the grounds below, on an limitation-by-limitation basis, how the prior art teaches and/or renders obvious all limitations of claims 10-29 of the '612 patent.

A. Claims 10-18 Are Obvious Over Drell in Combination with Knappe, Elliott, Hendricks, and Doganata

1. The Prior Art

a) Summary of Drell

115. 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).

116. Like the '612 patent, Drell describes multi-point (MP) conferencing, specifically video conferencing among a plurality of endpoints, as shown in Figure 1:

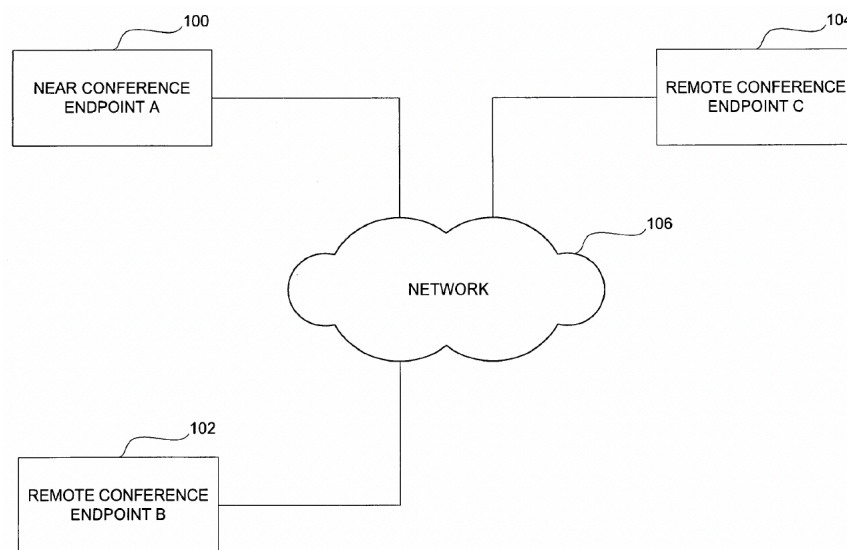
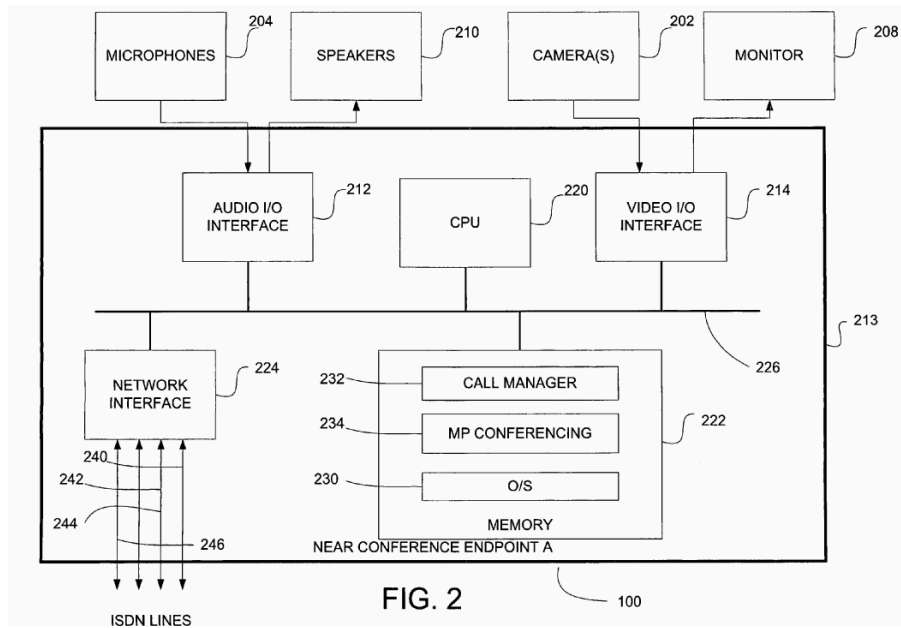


FIG. 1

Id. at FIG. 1.

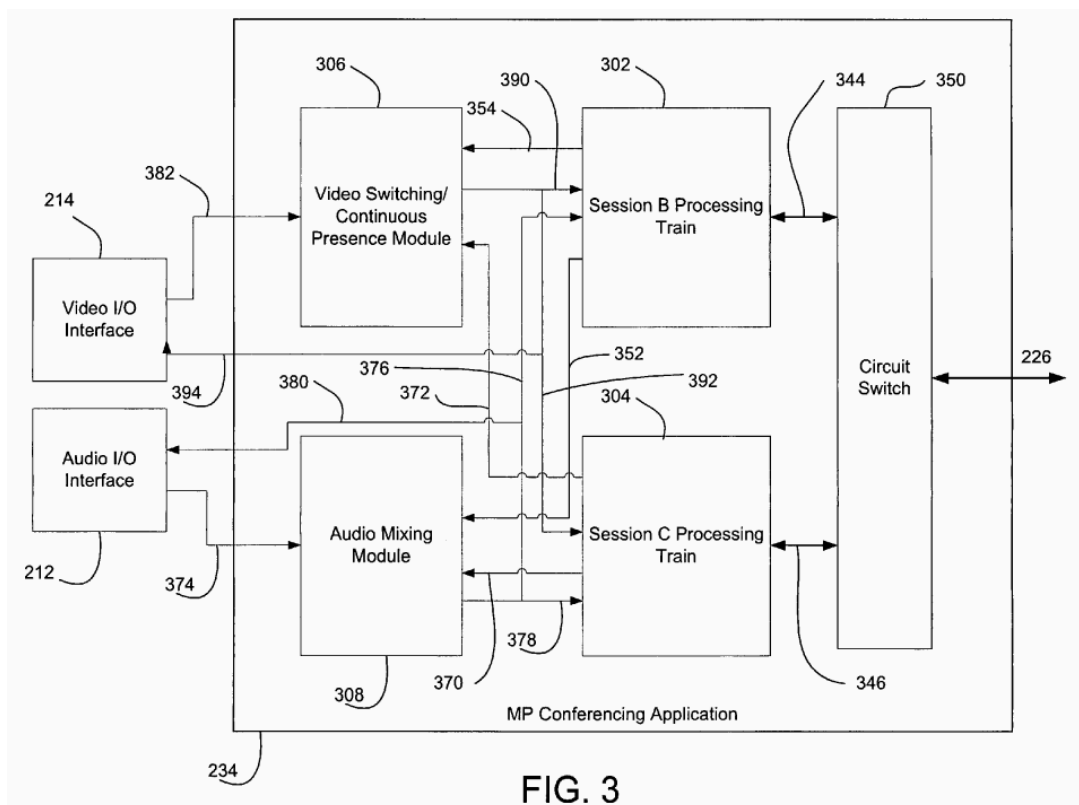
117. Specifically, Drell discloses a videoconferencing apparatus “to combine and distribute audio and video signals received from a plurality of remote conference endpoints[.]” *Id.* at Abstract. Drell explains that “network 106” facilitates the transmission of “audio and video data between and among near conference endpoint 100 and remote conference endpoints 102 and 104.” *Id.* at 3:20-23. Drell’s conferencing system, like the ’612 patent, also can be implemented across a hybrid network that includes the Internet and PSTN. Specifically, Drell notes that “[a]lthough 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. Drell further explains “that while only two remote conference endpoints are depicted in Fig. 1 for the purpose of clarity, a greater number of remote conference endpoints may be accommodated by near conference endpoint 100.” *Id.* at 3:15-19.

118. Drell’s system implements a software-based conference manager, explaining that centrally located conference bridges can have significant cost disadvantages. It thus teaches use of “MP conferencing application 234” that is located at a “near conference point 100.” *Id.* at 1:5-2:8, 4:6-20. Drell’s near conference endpoint 100 is illustrated in Figure 1, below:



Id. at FIG. 2.

119. Drell's MP conferencing application 234 is illustrated in Figure 3, reproduced below:



Id. at FIG. 3. 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. Once combined, the “audio and video streams are thereafter distributed to remote conferencing endpoints 102 and 104.” *Id.* at 4:29-31.

120. Drell further teaches that MP conferencing application 234 uses “a set of codecs” and that the communication process may “comprise an H.320 process (ISDN-based) or H.323 (packet-based) process.” *Id.* at 2:17-33, FIG. 4. To process received audio data, the “[a]udio codec 408 receives the audio data stream ... and applies ... decoding in accordance with a standard ... or proprietary audio compression algorithm.” *Id.* at 6:18-22. The decoded received audio and video data streams “together with the locally generated audio and video signals, are combined at an audio mixer and a video switching/continuous presence module.” *Id.* at 2:17-37. During transmission, “video codec 406 encodes the video data stream output by video switching/continuous presence module 306 (representative, for example, of a “continuous presence” image) using a standard or proprietary video compression algorithm (e.g., H.261) and...[a]udio codec 408 encodes the audio data stream output by audio mixing module 308 (representative, for example, of the blended speech of conference participants located at near conference

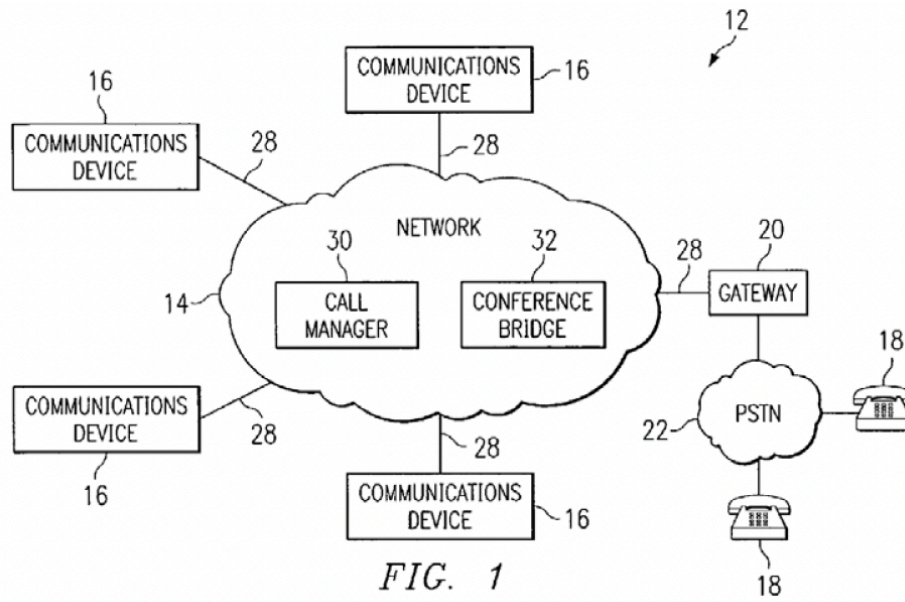
endpoint 100 and remote conference endpoints 102 and 104) using a standard or proprietary audio compression algorithm.” *Id.* at 6:33-44.

b) Summary of Knappe

121. 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.

122. Like the ’612 patent and Drell, 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.

123. 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. Knappe explains that “communication devices 16 comprise IP or other digital telephones, personal and other suitable computers or computing devices, personal digital assistants (PDAs), cell or other mobile telephones or handset or any other device or set of devices such as the telephone 18 and gateway 20 combination capable of communicating real- time audio, video and/or other information over the network 14.” *Id.* at 3:62-4:1.

124. 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.

125. 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.

126. 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:

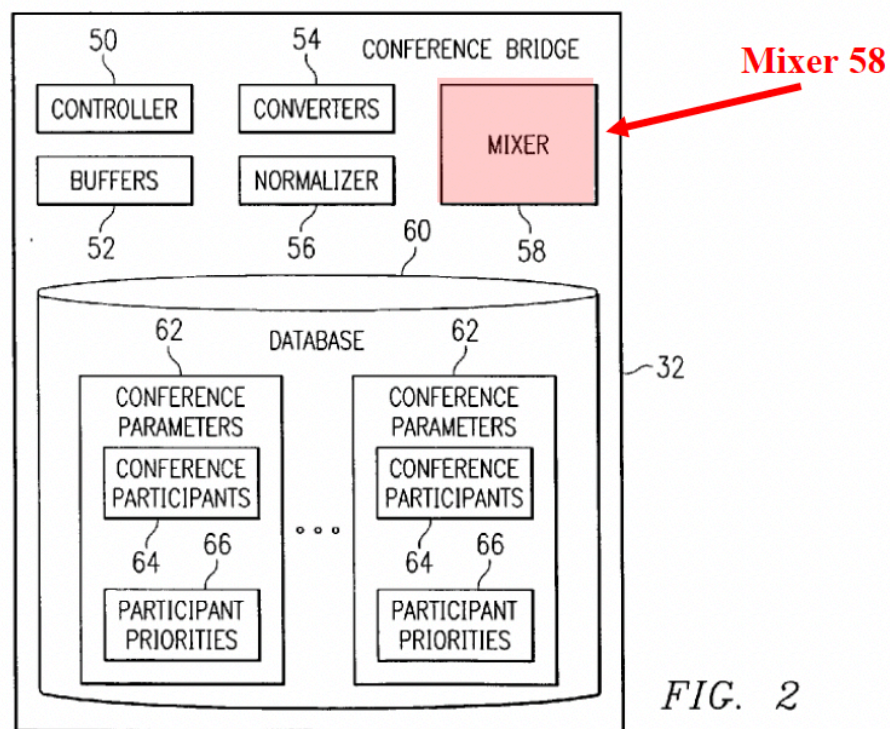
[a]t the gateway 20, the signals are converted to IP packets in the VoIP format. Similarly, VoIP packets received from the network 14 are converted into standard telephony signals for delivery to the destination telephone 18 through PSTN 22. The gateway 20 also translates between the network call control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

Id. at 4:21-28.

127. 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.

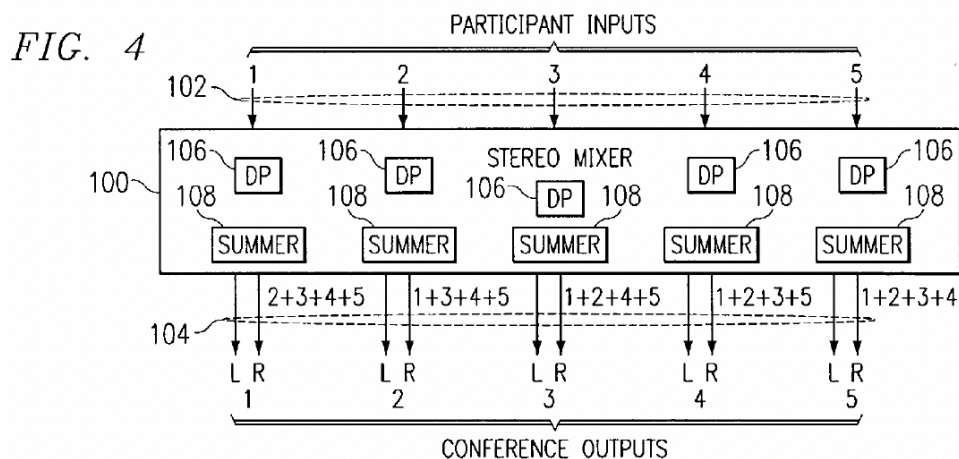
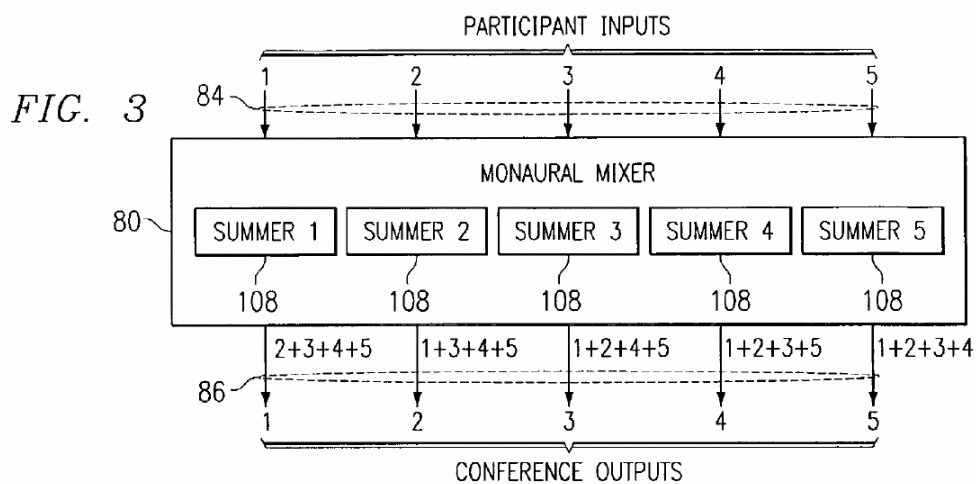
128. As illustrated in Figure 2 below, Knappe’s conference bridge 32 employs mixer 58, which “includes a plurality of summers or other suitable signal processing resources each operable to sum, add or otherwise combine a plurality of input streams into conference output streams for participants to a conference call.” *Id.* at 6:33-37.



Id. at FIG. 2 (annotated).

129. 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.

130. In the above figure, “the monaural mixer 80 [of the conference bridge 32] 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 6:59-62. Knappe further explains that “each participant is assigned a summer 82 that receives audio input streams from each other participant to the conference

call” which “combines the audio input streams to generate a conference output stream for delivery to the participant.” *Id.* at 6:62-67.

131. 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.*, 7:2-15. With respect to the 5-participant example of Figure 3, Knappe explains that “the conference output stream of participant 1 includes the audio inputs of participants 2–5.” *Id.* at 7:2-4. Similarly, the conference output stream of participants 2-5 includes the audio inputs of the other participants. *Id.* at 7:4-15.

c) Summary of Elliott

132. U.S. Patent No. 6,690,654 to Elliott et al. was filed on October 1, 1998, and issued on February 10, 2004. Like Drell, Elliott relates generally to the field of “communicating with a plurality of remote parties...for multi-media collaboration.” Elliott (Ex. 1007) at Abstract.

133. Elliott teaches that it was well known that the data in the transmitted stream includes “collaborative Web browsing, audio conferencing, video conferencing, and application sharing.” *Id.* at 1:40-44. Further, Elliott teaches utilizing 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.

d) Summary of Hendricks

134. WO2001018665A1 to Hendricks (Ex. 1040, “Hendricks”) was published on March 15, 2001. Hendricks is directed to video conference calls methods and systems. *Id.* at Abstract, 2:23-27. Specifically, Hendricks discloses “multi-party call...utilizing only one or a combination of several types of media (e.g., voice only or voice plus video and/or other media such as data, programs, etc.)” that communicate using “full duplex.” *Id.* at 3:4-10.

135. Like Drell, Hendricks discloses conferencing systems that support users “participating in a conference call,” including the “transmissions of video, audio and possibly other media.” *Id.* at 27:12-18. Hendricks, also like Drell, discloses that the network can be a “[h]ybrid... [of] the PSTN, and other networks, such as the Internet.” *Id.* at 23:13-14, Fig. 7.

e) Summary of Doganata

136. U.S. Patent No. 6,798,753 to Doganata (Ex. 1039) was filed on October 14, 1999 and published on September 28, 2004. Doganata is directed to “establishing a conference over a network, such as the Internet.” *Id.* at Abstract. Specifically, Doganata discloses how two types of conference connections can be scheduled. *Id.* at 4:38-39. In dial-in conference, a conference connection may be

established by participants dialing in to the conference. *Id.* at 4:39-42. In a dial-out conference, a connection may be established when “participants are contacted, or ‘dialed-out’ automatically at the time of the conference.” *Id.* at 4:42-45.

2. Reasons to Combine Prior Art

a) Reasons to Combine Drell and Knappe

137. A POSITA when considering the teachings of Drell would have also considered the teachings of Knappe, as they are analogous prior art both pertaining to the same field of endeavor of multiparty conferencing, including mixing audio from Internet and PSTN conference endpoints. Drell (Ex. 1008) at Abstract (“A videoconferencing apparatus includes a multi-point (MP) conference application that enables the apparatus to combine and distribute audio and video signals received from a plurality of remote conference endpoints.”); *see also id.* at 3:33-37 (“[T]he invention contemplates the use of two or more networks (for example, the PSTN and the Internet) to connect conference endpoints.”); Knappe (Ex. 1006) at 3:41-51 (“[C]ommunication system 12 includes...network 14 and/or PSTN 22...network 14 and/or PSTN 22 . . . network 14 is the Internet.”); *see also id.* at 6:59-62 (“[M]ixer 80 receives participant input streams 84 and combines the streams...to generate conference output streams 86 for each participant to a conference call.”).

138. A POSITA would have been motivated to combine the teachings of Drell with the teachings of Knappe for multiple reasons, including to produce obvious, beneficial, and predictable results, as discussed below.

(1) Moderator

139. As I discussed above, Drell teaches a conferencing system where audio and video data are transmitted between a “near conference endpoint” and “remote conference endpoints.” Drell (Ex. 1008) at 3:20-23. A POSITA would have been familiar with conferencing systems and would have recognized that it would be desirable for a person (e.g., a leader or a moderator) to initiate and organize the conference for other participants. Also, persons of skill would have recognized that 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.

140. Consistent with this knowledge in the art, Knappe teaches that at least “one participant may be given the priority of moderator for a particular conference call.” Knappe (Ex. 1006) at 8:52-56. According to Knappe, the “designated moderator” is able to “**control** the flow and/or **organization** of the conference.”

Id. at 1:53-55. Since Knappe teaches that at least one of the participants is a moderator, it would have been obvious to a POSITA either the user of the remote conference endpoint 104 or near conference endpoint 100 to be a moderator, depending on who initiates the conference. *See, e.g., id.* at Claim 1 (“...automatically identifying **a moderator initiating the multi-party communication session.**”). Moreover, designating an endpoint as the moderator would allow for the user (acting as the moderator) to control the flow and organization of the conference. *Id.* at 1:53-55 (“...a designated moderator(s) to **control** the flow and/or **organization** of the conference.”); *see also id.* at 3:35-37 (“The disclosed embodiments allow the moderator to exercise **control** and **influence** over the telephone conference.”); *see also* U.S. Patent No. 3,601,530 (Ex. 1014) at 2:34-40.

141. Accordingly, it would have been obvious to one of ordinary skill in the art for at least one of Drell’s conference participants, such as the user of the remote conference endpoint or near conference endpoint, to be a moderator per Knappe’s teaching, because doing so would beneficially allow the user (acting as the moderator) to initiate, organize, control, and influence the conference. Furthermore, as there are a finite number of participants in a conference, it would have been obvious to try designating one of Drell’s conference endpoint users as the moderator. The combination of Drell and Knappe is merely the ordinary use of

a common technique (identifying the user of the conference endpoint as a moderator per Knappe) in the similar conferencing system of Drell to yield predictable results of initiating, organizing, controlling, and influencing the conference. Each of these reasons separately and together would have motivated a POSITA.

142. The results would have been predictable and there would have been a reasonable expectation of success in the combination given that both Drell and Knappe describe similar conferencing systems. *See, e.g.*, Drell (Ex. 1008) at Abstract; Knappe (Ex. 1006) at Abstract, 3:47-51. Any modification needed to the teachings of Drell, including implementing software programming or hardware, to accommodate the teachings of Knappe, would have been within the level of ordinary skill in the art. *See, e.g.*, Knappe (Ex. 1006) at 5:49-62, 6:55-58. Indeed, Knappe itself recognizes that its teachings can be applied by POSITA to other network devices. *See, e.g., id.* at 4:30-33 (“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.”). Finally, the predictability and expectation of success for the combination are further confirmed by the use of a similar “conference leader” position that had long been known in the conferencing art. *See, e.g.*, U.S. Patent No. 3,601,530 (Ex. 1014) at 2:34-40.

(2) Mixing

143. As I discussed above, Drell teaches that its conference system “combine[s] and distribute[s] audio and video signals received from a plurality of remote conference endpoints.” Drell (Ex. 1007) at Abstract; *see also id.* at 5:18-20 (“[A]udio mixing module 308 may simply mix the received audio input data streams.”).

144. Knappe supplements and explains how to mix audio data from multiple conference endpoints. Specifically, in connection with Figure 3, Knappe teaches using a mixer 80 with a plurality of summers (e.g., summers 1 to 5, one for each participant), where each summer mixes audio data from the other participants and also “provide[s] higher prominence to the audio input [] of the conference moderator.” Knappe (Ex. 1006) at 6:59-7:15, FIGS. 1-3.

145. It would have been obvious to a POSITA, when implementing a mixer in Drell, to use a plurality of summers (one for each participant) as Knappe teaches, because this would allow for generating separate audio instances for each conference endpoint. Such an implementation would complement Drell’s teaching of generating separate instances for each endpoint. *See* Drell (Ex. 1008) at 2:19-21 (“Generally, the MP conferencing application generates separate processing trains for signal streams from/to each of the remote conference endpoints.”); *id.* at 6:66-7:4 (“By dynamically generating a separate instance of a processing train for each

remote endpoint session, a videoconferencing system embodying the invention may easily and flexibly accommodate endpoint sessions comprising a range of connection bandwidths and communication protocols.”).

146. Further, a POSITA would have applied Knappe’s mixing teachings to Drell because using separate summers (one for each participant) would facilitate providing higher prominence to the moderator and “increasing the intelligibility of the moderator relative to the remaining participants.” *See, e.g.*, Knappe (Ex. 1006) at 2:30-34, 6:59-7:17; *see also id.* at 2:40-46, 8:10-12. The increased intelligibility of the moderator enhances the moderator’s control over the conference. *See, e.g., id.* at 1:59-67.

147. Furthermore, a POSITA would have understood that it is undesirable for a conference participant to receive his or her own audio as part of the mixed audio. U.S. Patent No. 7,266,091 (Ex. 1034) at 6:31-33 (“It is undesirable for a conference participant to receive his or her own audio in the mixed audio stream.”). Including a participant’s own audio in the mixed audio for that participant creates the possibility of audio feedback or echo, which adversely affects the conference experience. It has long been known that “[e]choes are one of the primary factors affecting the perceived quality of voice connections.” *See* U.S. Patent No. 5,606,550 (Ex. 1035) at 1:20-22. Because network latencies are longer in packet-switched networks compared to circuit-switched networks (due in part to the

number of store-and-forward and queuing operations for packet data traversing a packet-switched network), “echoes are easily perceptible, annoying and need to be suppressed.” *Id.* at 1:23-26. Furthermore, feedback can occur when the transmitted audio signal is picked up by the far end receiver’s microphone and returned from the receiver back to the sender, such as may occur when the receiver side is on a speakerphone. While the delays in a packet-switched voice network are too long to produce the screeching such as occurs occasionally in a public-address audio system, echoes nevertheless may persist as they continue to be picked up and be amplified by both ends and continue to ping-pong back and forth, such as when both ends are using speakerphone. A POSITA would recognize that in a conference call, sending audio over a network to a mixer that includes the sender’s audio in the audio sent back to the sender may potentially generate an undesirable echo signal. Accordingly, a POSITA would have recognized that Knappe’s teachings of using a separate summer for each participant, would be beneficial to Drell since each participant’s mixed audio includes the other participants’ audio and excludes its own audio:

During normal operation, **each participant receives the audio input of each other participant.** Thus, for example, the conference output stream of participant 1 includes the audio inputs of participants 2–5. Similarly, the conference output stream of participant 2 includes the audio inputs of participants 1 and 3–5. The conference output stream of participant 3 includes the audio inputs of participants 1–2 and 4–5. The conference output stream of participant 4 includes the audio

inputs of participants 1–3 and 5. The conference output stream of participant 5 includes the audio inputs of participants 1-4.

Knappe (Ex. 1006) at 7:1-11. Thus, a POSITA would have been motivated to implement Drell's mixer, per Knappe's teachings to include a separate summer for each participant, because doing so would allow for excluding audio of the participant receiving the mixed audio, thereby avoiding undesirable audio effects in Drell's conferencing system.

148. In summary, a POSITA would have been motivated to apply Knappe's multiple-summer teachings when implementing a mixer in Drell because doing so would beneficially (1) further Drell's goal of generating separate instances for each conference endpoint, (2) increase intelligibility of the moderator (3) enhance the moderator's control over the conference, and (4) exclude audio of the participant receiving the mixed audio thereby avoiding undesirable audio effects, such as feedback or echo. Each of these reasons separately and together would have motivated a POSITA.

149. The results would have been predictable and there would have been a reasonable expectation of success in the combination given that both Drell and Knappe describe similar conferencing systems. *See, e.g.*, Drell (Ex. 1008) at Abstract; Knappe (Ex. 1006) at Abstract, 3:47-51. The reasonable expectation of success is also confirmed by other artisans, who for decades had been implementing conferencing systems with moderators. *See, e.g.*, U.S. Patent No.

3,601,530 (Ex. 1014) at 2:34-40. Any modification needed to the teachings of Drell, including implementing software programming or hardware, to accommodate the teachings of Knappe, would have been within the level of ordinary skill in the art. *See, e.g.*, Knappe (Ex. 1006) at 5:49-62, 6:55-58. Indeed, Knappe itself recognizes that its teachings can be applied by POSITAs to other network devices. *See, e.g. id.* at 4:29-33 (“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.”).

(3) Encapsulating VoIP Packets in RTP

150. 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.”); *id.* at 3:33-34 (“[N]etwork 106 may comprise a packet switched network, such as the Internet.”); *see also id.* at 5:8-25, FIG. 4. It would have been obvious that Drell is transmitting its audio data over the Internet using IP packets, because the Internet utilizes IP packets for communications. *See, e.g.*, Knappe (Ex. 1006) at 3:50-55 (“In the Internet embodiment, the network... transmits information in Internet

Protocol (IP) packets.”); Cisco Packetized Voice & Data Integration, Robert Caputo (Ex. 1021) at 12-13.

151. Further, Knappe supplements Drell and teaches that “Real-time IP packets such as VoIP packets are encapsulated in real-time transport protocol (RTP) packets for transmission over the network 14.” Knappe (Ex. 1006) at 3:50-58; *see also id.* at 4:9-12. In view of Knappe’s teachings, it would have been obvious to a POSITA to encode Drell’s real-time audio data in VoIP packets and encapsulate them in RTP packets for transmission over the Internet, as Knappe teaches, because RTP was specifically created for applications such as Drell’s real-time conferencing application. *See* Drell (Ex. 1008) at 3:20- 38 (“Network 106 may be of any type suitable for the transmission of audio and video data between and among near conference endpoint 100 and remote conference endpoints... [N]etwork 106 may comprise a packet switched network, such as the Internet...to connect conference endpoints..”); *see* Caputo (Ex. 1021) at 22 (“The Real- Time Protocol (**RTP**)...**was created to transport real-time traffic**, such as video and audio, over the Internet.”).

152. Furthermore, Drell teaches that its “communication process...may for example comprise an...H.323 (packet-based) process.” Drell (Ex. 1008) at 2:26-28. A POSITA would have known that the VoIP standard community adopted H.323 standards to foster interoperability of equipment from multiple vendors. *See*

Caputo (Ex. 1021) at 23 (“H.323 provides standards for voice-encoding...The voice over IP community has adopted H.323 standards in an effort to foster the interoperability of equipment from multiple vendors. H.323...has become critical for developing large-scale VoIP networks.”); 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.”). Accordingly, a POSITA would have been motivated to apply Knappe’s VoIP and RTP teachings to Drell because they are well suited for use with H.323 packet processing.

153. The result would have been predictable and there would have been a reasonable expectation of success in applying Knappe’s VoIP and RTP teachings to Drell, since the utilization of both VoIP and RTP were known to be applied in the conferencing context. *See* Caputo (Ex. 1021) at 22-23. Also, the results would have been predictable and there would have been a reasonable expectation of success given that both references describe similar conferencing systems that utilize real-time audio and communicate over the Internet. *See, e.g.,* Drell (Ex. 1008) at 1:27-28, 3:33-34; Knappe (Ex. 1006) at 3:29- 31, 3:50-58. Furthermore, any modification needed to the teachings of Drell, including implementing software programming or hardware, to accommodate the teachings of Knappe, would have been within the level of ordinary skill in the art. *See, e.g.,* U.S. Patent No. 6,850,496 (Ex. 1024) at 1:23-48; Knappe (Ex. 1006) at 5:49-62, 6:55-58;

Caputo (Ex. 1021) at 22-23. Indeed, Knappe itself recognizes that its teachings can be applied by POSITAs to other network devices. *See, e.g.*, Knappe (Ex. 1006) at 4:30-33 (“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.”).

(4) Gateway

154. Drell teaches that “the invention contemplates the use of two or more networks (for example, the PSTN and the Internet) to connect conference endpoints utilizing different communication protocols.” Drell (Ex. 1008) at 3:20-37. A POSITA would have recognized that since Drell’s Internet network and PSTN utilize different protocols, it would be beneficial to utilize a device that can convert data between the two networks. *See, e.g.*, U.S. Patent No. 6,553,515 (Ex. 1025) at 4:9-14. To that end, Knappe teaches using a “gateway... [that] provides conversion between analog and/or digital formats.” Knappe (Ex. 1006) at 4:18-20. Specifically, Knappe explains that “standard analog telephones [] communicate standard telephony signals through PSTN [network] to the gateway []. At the gateway [], the signals are converted to IP packets in the VoIP format. Similarly, VoIP packets received from the [Internet] network [] are converted into standard telephony signals for delivery to the destination telephone [] through PSTN.” *Id.* at 4:20-25; *see also id.* at 3:50-58.

155. In view of Knappe's teachings, it would have been obvious to a POSITA to use in Drell's system a gateway that receives standard audio data from PSTN conference endpoints and converts the audio data into a format (e.g., VoIP packets encapsulated in RTP) suitable for conference endpoints on the Internet network. Doing so would beneficially allow conference data to be transmitted between Drell's PSTN conference endpoints and near conference endpoint 100 over networks (PSTN and Internet) that utilize different protocols, thereby furthering Drell's objective. *See* Drell (Ex. 1008) at 3:20-37.

156. The result would have been predictable and there would have been a reasonable expectation of success in applying Knappe's gateway teachings to Drell, since the utilization of gateways was conventional and known to be applied in the conferencing context. *See, e.g.*, U.S. Patent No. 6,553,515 (Ex. 1025) at 4:9-14. Also, the results would have been predictable and there would have been a reasonable expectation of success given that both references describe similar conferencing systems that utilize both Internet networks and PSTN. *See, e.g.*, Drell (Ex. 1008) at 3:33-38; Knappe (Ex. 1006) at 4:18-25. Furthermore, any modification needed to the teachings of Drell, including implementing software programming or hardware, to accommodate the teachings of Knappe, would have been within the level of ordinary skill in the art.

157. For each of the above discussed combinations of Knappe and Drell, I rely on the teachings of the references and do not suggest that physical incorporation of Knappe's elements into Drell's system is required.

b) Reasons to Combine Drell and Elliott

158. A POSITA would have been motivated to combine the teachings of Drell and Elliott for multiple reasons, including to produce the obvious, beneficial, and predictable results of utilizing different types of collaboration data and VPN tunneling within Drell's system.

159. A POSITA when considering the teachings of Drell would have also considered the teachings of Elliott as they are analogous prior art both pertaining to the same field of endeavor of multiparty conferencing. Drell (Ex. 1008) at Abstract ("A videoconferencing apparatus includes a multi- point (MP) conference application that enables the apparatus to combine and distribute audio and video signals received from a plurality of remote conference endpoints."); Elliott (Ex. 1007) at Abstract ("A method for communicating with a plurality of remote parties...for multi-media collaboration."). Also, both Drell and Elliott support computer and PSTN telephone conference participants. *See, e.g.*, Drell (Ex. 1008) at 3:33-38; Elliott (Ex. 1007) at 5:35-39, 6:1-11, 8:16-39, FIG. 1.

160. Further, a POSITA would have combined the teachings of Elliott with Drell to obtain numerous predictable and beneficial results.

(1) Collaboration Data

161. Drell teaches a communications system capable of “transmission of audio and video data between and among near conference endpoint...and remote conference endpoints,” including “still images.” Drell (Ex. 1008) at 3:20-23, 6:53-56. A POSITA would have recognized that Drell’s disclosure of transmitting “still images” is an example of collaboration data. A POSITA would have recognized that sharing collaboration data in Drell is desirable since it beneficially provides additional ways (in addition to audio and video) to convey information among participants in a multiparty conference.

162. Notably, Elliott evidences this understanding and teaches the sharing of various types of collaboration data, including T.120 data. In one example, Elliott teaches that during a conference, the transmitted stream includes “collaborative Web browsing, audio conferencing, video conferencing, and application sharing” and that “[o]ther types of collaboration...data can be performed,” including transmitting data per the “T.120 standard.” Elliott (Ex. 1007) at 1:40-44, 5:49-56, 6:53-54. To provide context, the T.120 standard was established by the International Telecommunications Union (ITU) and contains a series of application protocols (i.e., T.122, T.123, T.124, and T.125), including standardizing the exchange of collaboration data. *See, e.g.*, A Primer on the T.120

Series Standard (Ex. 1016); '612 patent (Ex. 1001) at 3:19-21; *see also supra* ¶¶ 58-65, 105-107.

163. It would have been obvious to a POSITA in view of Elliott's teaching, to utilize collaboration data in Drell's system, including "Web browsing...and application sharing" (Elliott (Ex. 1007) at 1:40-44) and "T.120 standard" collaboration data (*id.* at 5:49-56), because this type of collaboration data would allow for conveying a greater variety of information among the participants thereby improving the conferencing experience and enhancing the effectiveness of collaboration in a variety of contexts. *See, e.g.*, U.S. Patent No. 6,237,025 (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 videotapes 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."). Furthermore, sharing collaborative information such as written or graphic material during a video conference had been known in the conferencing art for decades. *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).

164. Accordingly, a POSITA would have recognized that collaboration data (e.g., web browsing data, application data, and data provided by the T.120 standard, as Elliott teaches) would be beneficial to Drell's participants. For example, when the user of one of Drell's conference endpoints connects to the conference via a computer, the user may share web browsing data when doing so helps the other participants better understand the context for the discussion, such as when discussing a web page at a particular URL. Drell (Ex. 1008) at 3:54-64 (CPU, memory, and network interface), FIG. 2; Elliott (Ex. 1007) at 5:58-61 ("During the course of a collaborative session, one party can push a URL to another party. In other words, one party can cause the browser of a remote party to connect to a certain web site."). As another example, when the user of one of Drell's conference endpoints is the presenter, the user may share an application, such as sharing a presentation slide, so that other participants can follow along with the discussion (similar to presenting slides at an in-person meeting). *See* U.S. Patent No. 6,237,025 (Ex. 1018) at 9:23-26 ("'application sharing' (sharing both the control and display of running applications)"). In short, Elliott's collaboration data (e.g., web browsing, application sharing, and other collaboration data) beneficially provides Drell's participants distinct ways to share information during

a conference, thereby enhancing the effectiveness of the collaboration among the participants.

165. Such a combination would have simply been combining prior art elements (sharing collaboration data taught by Elliott in the conferencing system of Drell) according to known methods (disclosed by Elliott) to yield predictable results (allow for participants in Drell's system to share various types of data during a conference for collaboration purposes). Additionally, the combination of Drell and Elliott is merely the ordinary use of common techniques (transmitting collaboration web browsing data, application sharing data, and collaboration data per the T.120 standard as Elliott teaches) to improve the similar conferencing system of Drell in the same way by allowing for enhanced collaboration among participants.

(2) VPN Tunneling

166. Drell also teaches a communications system capable of "transmission of audio and video data between and among near conference endpoint[s]...and remote conference endpoints" across "network 106 [which] may comprise a packet switched network, such as the Internet." Drell (Ex. 1008) at 3:20-38, 6:53-56. It was known, however, that data transmitted over public IP networks is vulnerable to eavesdropping and other problems since the transmission is not secure. This is evidenced, for example, by the VPN Textbook which states that "more and more

business-to-business traffic takes place over the Internet” and that this type of communication is “riddled with security issues.” VPN Textbook (Ex. 1010) at 119. A solution for securing communications over public networks was also known. Specifically, it was known to protect communications using a VPN. *See, e.g., id.* at 115 (“A Virtual Private Network (VPN)...[provides] connectivity amongst multiple sites...with the same access or security policies as a private network.”); *see also id.* at 118, 129-130 (describing that the most common VPN implementation is hub-and-spoke). Thus, it would have been obvious to a POSITA that VPNs would beneficially provide security to the system of Drell, which describes conferencing communications for business over the Internet. Drell (Ex. 1008) at 1:20-25.

167. The obviousness of adding a VPN feature to Drell’s conferencing system is further demonstrated by the work of other artisans making similar feature combinations in the prior art timeframe. For example, U.S. Patent No. 6,980,526 addresses the “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.” Ex. 1020 at 2:65-3:3, 5:52-55. As another example, U.S. Patent No. 6,782,413 describes a “conference bridge application” that transmits data “over public networks, using technologies such as...[a] Virtual Private Network (VPN)

or any other conventional communication method.” Ex. 1019 at 4:23-27, 6:36-41. That skilled artisans were combining conferencing systems and VPNs confirms that a POSITA would have found it obvious to combine Drell’s conferencing system with the VPN technology described in Elliott and the VPN Textbook (further discussed below).

168. Consistent with well-known approaches, Elliott teaches utilizing 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 utilize VPN tunneling when communicating audio, video, and other data across Drell’s Internet network 106 because VPN tunneling would make the transmitted data secure from outside access and interference. *See, e.g.*, 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.”). Additionally, a POSITA would have understood that VPNs are beneficial because they reduce the need to lease private data lines, which reduces or avoids the costs for implementing a private network by communicating securely over the Internet. *See, e.g.*, 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.”).

169. As such, one of ordinary skill in the art would have been motivated to apply the techniques of using VPN tunneling for conferencing, as Elliott teaches, to Drell's conferencing system because VPN tunneling provides secure communications while also reducing or avoiding costs associated with leased private data lines. Indeed, the combination of Drell and Elliott is merely the ordinary use of a common technique (using VPN tunneling for conferencing data) in the similar conferencing system of Drell to yield predictable results of (1) securing the transmitted data, and (2) reducing or avoiding costs associated with leasing private data lines. Each of these benefits separately and together would have motivated a POSITA.

170. I rely on the teachings of the references and do not suggest that physical incorporation of Elliott's elements into Drell's system is required in the proposed combination.

171. Lastly, I also note that the results would have been predictable and there would have been a reasonable expectation of success in the combination given that both Drell and Elliott describe similar conferencing systems. *See, e.g.*, Drell (Ex. 1008) at Abstract; Elliott (Ex. 1007) at Abstract. This is especially so since VPN tunneling was conventional. U.S. Patent No. 6,782,413 (Ex. 1019) at 4:23-27, 6:36-41; Elliott (Ex. 1007) at 5:49-56. Any modification needed to the

teachings of Drell, to accommodate the teachings of Elliott, would have been within the level of ordinary skill in the art.

172. For example, the modifications would have been accomplished with software programming or employing commercially available hardware, which would have been within the level of skill in the art and predictable. I personally, before the earliest claimed priority date of the '612 patent, configured, deployed and utilized VPN technology for carrying VoIP traffic using the NBX 100 system, which was also capable of performing conference calls.

c) Reasons to Combine Drell and Hendricks

173. A POSITA would have been motivated to combine the teachings of Drell and Hendricks for multiple reasons, including to produce the obvious, beneficial, and predictable results of providing full duplex audio to the conference participants.

174. I initially note that one of ordinary skill in the art when considering the teachings of Drell would have also considered the teachings of Hendricks, as they are analogous prior art both pertaining to the same field of endeavor of multiparty conference calling. Drell (Ex. 1008) at 1:16-18 (“The present invention relates generally to conferencing systems, and more particularly to a videoconferencing apparatus for use with multi-point conferences.”); Hendricks (Ex. 1040) at 3:4-10 (“...multi-party call...utilizing only one or a combination of

several types of media (e.g., voice only or voice plus video and/or other media such as data, programs, etc.)”).

175. As already noted, Drell contemplates audio data communication over “two or more networks (for example, the PSTN and the Internet).” Drell (Ex. 1008) at 3:35-37. A POSITA would have known that audio data could be transmitted in two modes; namely, full duplex and half duplex. U.S. Patent No. 6,768,914 (Ex. 1041) at 1:20- 21 (“Speakerphones may be designed to operate in either half-duplex or full-duplex.”); U.S. Patent No. 6,501,740 (Ex. 1042) at 7:17-20 (“[T]he mode can be specified to be full duplex (full participation, speaking and listening) or half duplex (listen only).”).

176. A POSITA, however, would have recognized that for conferencing applications full duplex is preferred. U.S. Patent No. 6,768,914 (Ex. 1041) at 1:28-35 (“[M]any speakerphones are designed to operate in full-duplex mode. The full-duplex mode allows incoming and outgoing parties voices to be simultaneously transmitted so that **there is no clipping or choppiness**. This results in much more **natural and spontaneous** flowing conversation. **Full-duplex is even more important** when more than one person is in a room using the speakerphone for a **conference call**.”). Thus, it would have been obvious to a POSITA that full duplex audio would be beneficial to the communication system in Drell that conducts multiple party conferencing.

177. Hendricks discloses “multi-party call...utilizing only one or a combination of several types of media (e.g., voice only or voice plus video and/or other media such as data, programs, etc.)” that communicate over PSTN and Internet networks using “**full duplex.**” Hendricks (Ex. 1040) at 3:4-10, 23:13-14, Fig. 7, 17:19-22 (“The network [] provides **full duplex interconnectivity** between viewers [] involved in a given video conference call. The network [] can support many simultaneous calls.”). It would have been obvious for a POSITA to utilize full duplex audio (as Hendricks teaches) in Drell’s conferencing system because doing so would beneficially (1) reduce clipping or choppiness and (2) facilitate natural and spontaneous flowing conversation among the conference participants. U.S. Patent No. 6,768,914 (Ex. 1041) at 1:28-33. As an additional third (3) reason, a POSITA would have utilized full duplex in Drell’s system because doing so would improve audio when more than one person is in a room using the speakerphone for a conference call.

178. Accordingly, a POSITA would have been motivated to apply the full duplex voice communication teachings of Hendricks to Drell’s communication system to obtain the beneficial results of (1) reducing clipping or choppiness, (2) facilitating natural and spontaneous flowing conversation, and (3) improving audio when more than one person is in a room using the speakerphone for a conference call. Each of these benefits separately and together would have motivated a

POSITA. Indeed, the use of a known full duplex technique, as taught by Hendricks, improves the similar conferencing system of Drell in the same way; namely, it (1) reduces clipping or choppiness, (2) facilitates natural and spontaneous flowing conversation, and (3) improves audio when more than one person is in a room using the speakerphone for a conference call. I also note that it would have been obvious to try full duplex in Drell's conferencing system since there are a finite number of identified predictable solutions (full duplex and half duplex) and there would have been a reasonable expectation of success.

179. That the result would have been predictable and that there would have been a reasonable expectation of success is evidenced by that fact that at the time "many speakerphones [were already] designed to operate in full-duplex mode." U.S. Patent No. 6,768,914 (Ex. 1041) at 1:28-33. Further, other artisans had implemented full duplex systems. U.S. Patent No. 6,501,740 (Ex. 1042) at 7:14-20, 7:51-55. Moreover, any modification needed to the teachings of Drell to accommodate the full duplex teachings of Hendricks, such as utilizing software programing and hardware, would have been within the level of ordinary skill in the art. U.S. Patent No. 6,424,647 (Ex. 1043) at 12:20-32 ("It can be appreciated that the above hardware and software configuration can be practiced to provide tele-conference capabilities between multiple phone parties. The tele-conference can be with all parties interconnected through an Internet connection and an Internet

phone and the above software and hardware configurations modified to provide for full duplex communications between all the parties...two parties can be connected over the Internet as described above, and multiple parties can be brought into the phone call connection through connections over the PSTN.”).

180. Lastly, I rely on the teachings of the references and do not suggest that physical incorporation of Hendricks’s elements into Drell’s system is required in the proposed combination.

d) Reasons to Combine Drell with Doganata

181. A POSITA would have been motivated to combine the teachings of Drell and Doganata for multiple reasons, including to produce the obvious, beneficial, and predictable results of establishing a connection between conference participants.

182. I initially note that one of ordinary skill in the art when considering the teachings of Drell would have also considered the teachings of Doganata, as they are analogous prior art with both pertaining to the same field of endeavor of multiparty conferencing. Drell (Ex. 1008) at 1:17-18 (“[a] videoconferencing apparatus for use with multi-point conferences.”); Doganata (Ex. 1039) at Abstract (“A system and method for providing automatic scheduling and establishment of telephone conference over a network such as the Internet.”).

183. Further, a POSITA would have combined the teachings of Doganata with Drell to obtain numerous predictable and beneficial results.

184. As I discussed above, Drell teaches a conferencing system for transmitting data between various conferencing endpoints over one or more networks, such as the PSTN and the Internet. Drell (Ex. 1008) at 3:35-36. To setup the conference, Drell utilizes a call manager application 232 that “controls the establishment and termination of connections between near conference endpoint 100 and remote conference endpoints 102 and 104.” *Id.* at 4:15-18.

185. Doganata supplements Drell and teaches a dial-in technique where “a user requests assignment of a dial-in number and a passcode so that the participants may dial in to the conference when the time comes.” Doganata (Ex. 1039) at 4:39-42. Doganata further explains that dial-in process may be used by participants with telephones connected to the PSTN. *Id.* at 3:62-63.

186. It would have been obvious to a POSITA, when establishing a connection in Drell between the near conference endpoint 100 and the conference endpoints connected to the PSTN, to use a dial-in technique as Doganata teaches. Doing so would enable the participants to individually establish a conference call connection and join the conference when they chose to do so. *See, e.g., id.* at 3:10-12.

187. Indeed, the combination of Drell and Doganata is merely a use of an ordinary technique (dialing in), in Drell's similar conference system to yield the predictable result of establishing conference call connections. I also note that it would have been obvious to dial-in to establish connections in Drell's conferencing system since there are a finite number of identified predictable solutions (dial-out and dial-in) and there would have been a reasonable expectation of success in establishing the connection.

188. The result would have been predictable and there would have been a reasonable expectation of success in applying Doganata's teachings in Drell, since the utilization of dial-in techniques was conventional and known to be applied in establishing conference connections. *See, e.g.*, U.S. Patent No. 6,961,416 (Ex. 1044) at 11:26-31 ("Once a conference has been scheduled and the conference start time has arrived, callers 16 may join or be joined to the conference. FIG. 7 illustrates an exemplary method of joining caller 16 to a scheduled conference, **using either a 'dial-in' procedure, a 'dial-out' procedure, or a combination of these.**"); *see also id.* at 12:14-19. Further, the results would have been predictable and there would have been a reasonable expectation of success given that both references describe similar conferencing systems. Furthermore, any modification needed to the teachings of Drell, including implementing software programming or

hardware, to accommodate the teachings of Doganata, would have been within the level of ordinary skill in the art.

189. Lastly, I rely on the teachings of the references and do not suggest that physical incorporation of Doganata's elements into Drell's system is required in the proposed combination.

3. Claim 10

[10.P] A method for establishing a secure multi-participant conference call including full duplex audio, the method comprising:

190. The prior art renders obvious a method for establishing a secure multi-participant conference call including full duplex audio.

191. **First**, Drell discloses “[a] **method for conducting a conference**” that uses “conferencing systems, and more particularly to a videoconferencing apparatus for use with **multi-point conferences.**” Drell (Ex. 1008) at 1:16-18. Drell's system that performs the method is illustrated below at Figure 1:

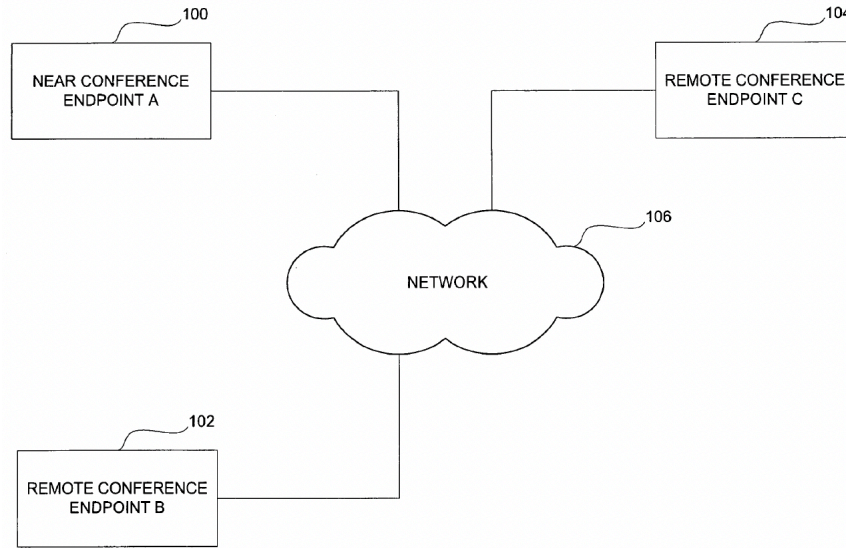


FIG. 1

Id. at FIG. 1.

192. With reference to the above figure, Drell explains that its system is a multi-point (MP) system that includes a network 106, which can be a combination of two networks such as the Internet and Public Switched Telephone Network (PSTN), for communicating audio/video data between near conference endpoint 100 and remote conference endpoints 102 and 104:

FIG. 1 depicts an exemplary operating environment of the multi-point (MP) conferencing application of the present invention. **A near conference endpoint 100**, embodying the MP conferencing application, is coupled to **remote conference endpoints 102 and 104 via a network 106**. Remote conference endpoints 102 and 104 may comprise, for example, conventional videoconferencing devices equipped to **transmit and receive** both video (image) data and **audio (speech) data**. Alternatively, one or more of remote endpoints 102 and 104 may comprise conventional audio conferencing devices limited to reception and transmission of audio data. It should be appreciated that while only two remote conference endpoints are depicted in FIG. 1 for the purpose of clarity, a greater number of

remote conference endpoints may be accommodated by near conference endpoint 100.

Network 106 may be of any type suitable for the transmission of **audio** and video data between and among near conference endpoint 100 and remote conference endpoints 102 and 104. Typically, network 106 will comprise the public switched telephone network (PSTN) or comparable circuit switched network to which each of the conference endpoints is connected by one or more ISDN lines. A multi-point conference is initiated by establishing a connection between near conference endpoint 100 and remote conference endpoint 102, and between near conference endpoint 100 and remote conference endpoint 104. Establishment of the connections may be effected through a dial-up procedure, or through use of a dedicated line.

Alternatively, **network 106 may comprise a packet switched network, such as the Internet.** Although a single network 106 is shown, the invention 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 3:4-47.

193. In view of the above, a POSITA would have understood that Drell's system utilizes network 106, which can comprise the Internet and PSTN, to facilitate video/audio communications during a conference call between near and remote conference endpoints, including IP conference endpoints and PSTN conference endpoints. Thus, Drell discloses a system for supporting a multi-participant audio/video conference call.

194. **Second**, Elliott teaches transmitting audio and video conference data over Internet network 18 using a VPN tunnel:

Multi-media collaboration refers to the use of more than one **media stream (e.g.: voice, fax, data, video, etc.) used in collaboration** with more than one party.

Elliott (Ex. 1007) at 2:66-3:1.

As alluded to above, communications between the parties can be made over a variety of networks 18. A network 18 is at least two (and preferably more) computers interconnected together so that communication between them is possible. In the illustrated embodiment, the public Internet is shown as an exemplary **Internet Protocol-based (IP-based) network 18 for multi-media collaboration.**

Id. at 4:6-12.

For example, network 18 can be either a public or a carrier supported **virtual private network (VPN)...which uses IP or IP-tunneling.**

Id. at 4:31-34.

195. Based on the above disclosure, a POSITA would have understood that Elliott utilizes VPN tunneling for communicating conferencing 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 for a POSITA to utilize VPN tunneling in Drell's system for communicating audio and video data over Drell's Internet network 106. This is because the VPN tunneling would make Drell's system secure from outside access and interference. *See, e.g.,* VPN Textbook (Ex. 1010) at 115 ("**A Virtual Private Network (VPN)**...[provides] the same access or **security** policies as a private network."), 122 ("Overlay VPN networks can be implemented with...IP-over-IP tunneling methods...[including] Generic Route

Encapsulation (GRE) tunneling and IP Security (IPSec) encryption.”); 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.”); *see also* Reasons to Combine Drell and Elliott, *supra* ¶¶ 158-172. Accordingly, Elliott’s VPN tunneling teachings renders obvious establishing a “*secure*” multi-participant conference, as recited in the preamble.

196. **Third**, Drell teaches that its “communication process...may for example comprise an...**H.323** (packet-based) process.” Drell (Ex. 1008) at 2:26-33. It would have been obvious to a POSITA that Drell includes utilizing “*full duplex audio*,” because H.323 was a well-known standard that supported both full-duplex or half- duplex transmission. *See, e.g.*, U.S. Patent No. 6,553,515 (Ex. 1025) 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.”).

197. In addition to Drell’s disclosure of full- duplex audio, Hendricks discloses that a multi-participant conference call over Internet/PSTN network includes full duplex audio:

Hybrids of cable television networks, cellular telephony networks, PCNs, the **PSTN**, and other networks, such as the **Internet**, are possible.

Hendricks (Ex. 1040) at 23:13-14, FIG. 7.

[T]he term “**conference call**,” or sometimes more simply “call,” denotes a traditional two-party call as well as a **multi-party call** or a single party call with storage and future access by others, includes calls utilizing only one or a combination of several types of media (e.g., voice only or **voice plus video and/or other media such as data**, programs, etc.), and includes one way broadcast calls (i.e., view only) as well as **full duplex calls or mixed media** combinations of one-way broadcast and full duplex (e.g., broadcast video from one party and shared return audio 10 from all parties).

Id. at 3:4-10.

Figure 7 illustrates a video conferencing network 1000 and various configurations for connecting viewers 266 (or other video conferencing end equipment 1005) to the network 1000. The network 1000 provides **full duplex interconnectivity** between viewers 266 involved in a given video conference call. The network 1000 can support many simultaneous calls.

Id. at 17:19-22.

In a preferred embodiment, a video conferencing central node performs central combining for a conference call. Central combining simplifies the communication among the viewers 266 **participating in a conference call by establishing full duplex connections** having fixed or bounded bandwidth, one between each participating viewer 266 and the central node. The central node manages all aspects of the call and is ultimately responsible for the selection of information received by each participating viewer on its single channel. The central node receives all uplink transmissions of **video, audio and possibly other media**. The central node then makes various combinations of the uplink data, one combination being transmitted to each participating viewer 266.

Id. at 27:12-19. Accordingly, Hendricks renders obvious that the conference call includes “*full duplex*” audio.

198. It would have been obvious for a POSITA to utilize full duplex audio in Drell's system (as Hendricks teaches) because doing so would provide the participants of Drell's near conference end point 100, remote conference end points 102 and 104, and PSTN conference endpoints a more natural way to communicate with each other during the conference. *See* U.S. Patent No. 6,768,914 (Ex. 1041) at 1:20-39; U.S. Patent No. 6,501,740 (Ex. 1042) at 7:14-34; Reasons to Combine Drell and Hendricks, *supra* ¶¶ 173-180.

199. Thus, Drell in combination with Elliott and Hendricks teaches a method for supporting multiple participants in a secure (VPN tunneling) conference call including full duplex audio, which renders obvious “[a] *method for establishing a secure multi-participant conference call including full duplex audio*” as recited in the preamble.

[10.1.1] initiating a dial-in process to establish a connection with a local moderator

200. The prior art renders obvious initiating a dial-in process to establish a connection with a local moderator computer.

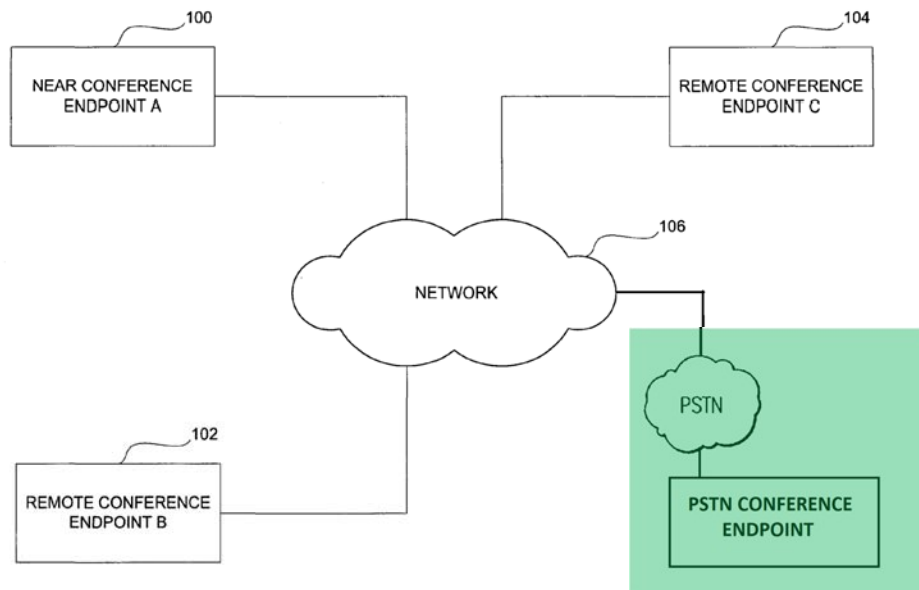
201. **First**, Drell teaches that a PSTN conference endpoint participates in a conference with a near conference endpoint 100:

Network 106 may be of any type suitable for the transmission of audio and video data between and among near conference endpoint 100 and remote conference endpoints...the invention contemplates the use of two or more networks (for example, the PSTN and the Internet) to

connect conference endpoints utilizing different communication protocols.

Drell (Ex. 1008) at 3:20-37. Drell further explains that “while only two remote conference endpoints are depicted in FIG. 1 for the purpose of clarity, a greater number of remote conference endpoints may be accommodated by near conference endpoint 100.” *Id.* at 3:15-19.

202. In view of the above disclosure, as I have annotated in Figure 1 below, a POSITA would have understood that Drell contemplates that a PSTN conference endpoint participates in the conference:



Id., FIG. 1 (annotated).

203. **Second**, Knappe teaches designating at least one of the conference participants as a moderator:

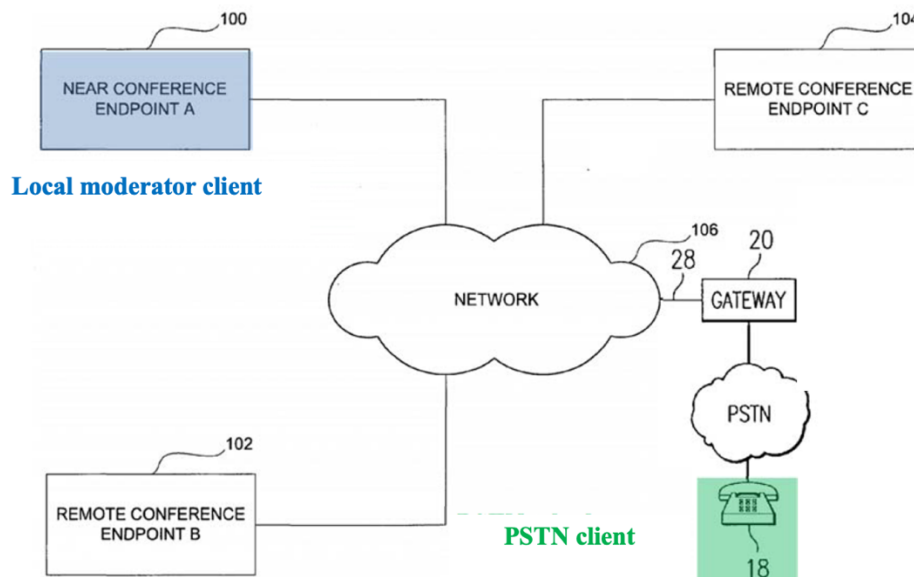
[A]t least one participant may be given the priority of moderator for a particular conference call. Typically, the moderator

will receive the highest priority in a conference call. Therefore, in accordance with one embodiment, the priorities simply include moderator for one.

Knappe (Ex. 1006) at 8:52-56.

204. Since Knappe teaches that at least one participant may be given the priority of a moderator, it would have been obvious to a POSITA to designate the user of Drell's near conference endpoint 100 as the moderator (as shown in annotated Figure 1 below), when the near conference endpoint 100 is facilitating the conference call. Doing so would allow for the user of the near endpoint (acting as the moderator) to control the flow and organization of the conference. *See id.* at 1:53-55 (“...a designated moderator(s) to **control** the flow and/or **organization** the conference.”); *see also* Reasons to Combine Drell and Knappe, *supra* ¶¶ 137-157.

205. Accordingly, Drell in combination with Knappe teaches that the user of the near conference endpoint 100 is a “*local moderator client*” as claimed:



Drell (Ex. 1008) at FIG. 1 (annotated).

206. **Third**, Drell teaches that the moderator near conference endpoint 100 includes within its memory 222 a call manager application 232:

Reference is now directed to FIG. 2, which depicts in block form various components of near conference endpoint 100...

Depending on bandwidth requirements, data communicated between **near conference endpoint 100** and a **remote conference endpoint** may be carried on a single ISDN line, or may be distributed (for higher bandwidth connections) among a plurality of ISDN lines.

Stored within memory 222 are an operating system 230, a call manager application 232, and the MP conferencing application 234. Operating system 230 controls the allocation and usage of hardware resources, such as CPU 220 and 15 memory 222. Call 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 3:39-4:20.

207. Drell teaches that the call manager application 232 controls the establishment and termination of connections between the near conference endpoint 100 and remote conference endpoints 102/104:

Call 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.

208. Doganata teaches how multiple participants, including PSTN conference endpoints, use a “dial-in” processes to establish a conference:

Two types of conferences can be scheduled, a dial-in conference or a dial-out conference. For a dial-in conference, a user requests assignment of a dial-in number and a passcode so that the participants may dial in to the conference when the time comes. On the other hand, the moderator may request establishment of a dial-out conference in which the participants are contacted, or “dialed-out”, automatically at the time of the conference.

Doganata (Ex. 1039) at 4:38-45.

The desktop application automatically **contacts participants** through e-mail messages **informing them of the time, date, dial-in number**, etc., to which the participants must respond with an e-mail either accepting or rejecting the conference. Generally such notifications by e-mail are fed directly into the recipient's calendar once the e-mail has been-opened and must be deleted by the recipient if the conference time is being rejected.

Id. at 5:25-33.

The conference Service Provider 15 receives the service requests 17 over the Internet 14, allocates an audio bridge for the conference, and either dials out to the **participants phones via the Public**

Switch Telephony Network 16 or awaits dial-in calls for connecting to the audio bridge.

Id. at 3:60-65.

209. In view of Doganata's teachings, it would have been obvious to a POSITA to initiate a dial-in process to establish a connection with the call manager application 232 of the moderator near conference endpoint 100. Doing so would beneficially allow Drell's conference participants and the moderator to individually establish a conference call connection using the conference endpoints and join in a conference as each participant desires. *See* Reasons to Combine Drell and Doganata, *supra* ¶¶ 181-189.

210. Thus, Drell in combination with Knappe and Doganata teaches initiating a dial-in process by a PSTN conference endpoint to establish a connection with a moderator near conference endpoint 100, which renders obvious "*initiating a dial-in process to establish a connection with a local moderator client,*" as claimed.

[10.1.2] in response to an input from a Public Switched Telephone Network (PSTN) client;

211. The prior art renders obvious in response to an input from a Public Switched Telephone Network (PSTN) client.

212. **First**, as discussed at [10.1.1], Drell in combination with Knappe and Doganata teaches initiating a dial-in process by a PSTN conference endpoint to establish a connection with a moderator near conference endpoint 100.

213. **Second**, Doganata teaches that the dial-in processes is initiated in response to the conference participants inputting a dial-in number and a passcode:

Two types of conferences can be scheduled, a dial-in conference or a dial-out conference. For a dial-in conference, a user requests assignment of a **dial-in number** and a **passcode** so that the participants may dial in to the conference when the time comes. On the other hand, the moderator may request establishment of a dial-out conference in which the participants are contacted, or “dialed-out”, automatically at the time of the conference.

Doganata (Ex. 1039) at 4:38-46.

The conference Service Provider 15 receives the service requests 17 over the Internet 14, allocates an audio bridge for the conference, and either dials out to the **participants phones via the Public Switch Telephony Network 16 or awaits dial-in calls for connecting to the audio bridge.**

Id. at 3:60-65. Doganata’s dial-in number and the passcode/password, separately and together, render obvious the claimed “*input*.”

214. In view of Doganata’s teaching, it would have been obvious to a POSITA that in the combination, Drell’s PSTN conference endpoint would use a dial in number and passcode as input to establish a connection with the moderator near conference endpoint 100. A POSITA would have been motivated to make the combination since Doganata’s teaching would allow the PSTN conference endpoint participant to connect to the conference and exchange audio data with the

other participants. Moreover, a POSITA would have recognized that inputting a passcode/password is desirable since it provides security to the conference and identifies which of multiple conferences which to join. *See* Reasons to Combine Drell and Doganata, *supra* ¶¶ 181-189.

215. Thus, Drell in combination with Knappe and Doganata teaches that initiating the dial-out process (discussed at [10.1.1]) is in response to an input (e.g., dial-in number and/or passcode) from the PSTN conference endpoint, which renders obvious “in response to an input from a Public Switched Telephone Network (PSTN) client,” as claimed.

[10.2] receiving first audio data at the local moderator client from the PSTN client through a PSTN gateway;

216. The prior art renders obvious receiving first audio data at the local moderator client from the PSTN client through a PSTN gateway.

217. **First**, as discussed at [10.1.1]-[10.1.2], Drell in combination with Knappe and Doganata teaches initiating a dial-in process to establish a connection with local moderator near conference endpoint in response to input from the PSTN conference endpoint.

218. **Second**, Drell teaches that the PSTN conference endpoint participates in a conference with a near conference endpoint 100:

Network 106 may be of any type suitable for the transmission of audio and video data between and among **near conference endpoint 100** and remote conference endpoints...the invention contemplates the

use of two or more networks (for example, **the PSTN** and the Internet) **to connect conference endpoints** utilizing different communication protocols.

Drell (Ex. 1008) at 3:20-37. In view of the above, a POSITA would have understood that Drell's near conference endpoint 100 local moderator establishes communications with PSTN conference endpoints.

219. Drell further describes receiving audio data from remote conference endpoints:

As will be described in further detail below, MP conferencing application 234 is configured to instantiate a processing train for each remote conference endpoint 102 and 104 to which near conference endpoint 100 is connected. The processing trains process **audio and video data streams received from remote conferencing endpoints** 102 and 104.

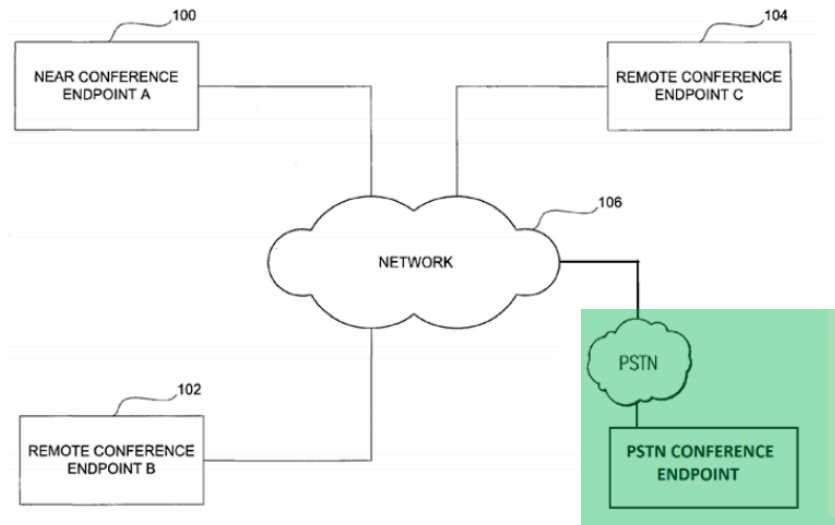
Id. at 4:21-26.

Audio mixing module 308 is configured to combine **audio data received from remote conference endpoints** 102 and 104 with locally generated audio data (received from audio I/O interface 212 via audio data path 374, and typically being representative of the speech of the near conference participant(s)).

Id. at 5:8-13; *see also* FIGs. 1-3. Drell further explains that “while only two remote conference endpoints are depicted in FIG. 1 for the purpose of clarity, a greater number of remote conference endpoints may be accommodated by near conference endpoint 100.” *Id.* at 3:15-19.

220. In view of the above disclosure, as I have annotated in Figure 1 below, a POSITA would have understood that Drell contemplates that a PSTN

conference endpoint participates in the conference and that the near conference endpoint receives and transmits audio data to and from the PSTN conference endpoint:



Id. at FIG. 1 (annotated).

221. Thus, Drell’s moderator near conference endpoint 100 that receives audio data from the PSTN conference endpoint over a PSTN renders obvious “*receiving first audio data at the local moderator client from the PSTN client,*” as claimed.

222. **Third**, Knappe similarly describes “multiparty communications sessions” where conference media streams include “audio, video, voice, data and other suitable types of real-time and/or non real-time traffic.” Knappe (Ex. 1006) at 3:27-28, 3:65-4:1, 5:46-47. Knappe further teaches that when a PSTN conference endpoint participates in a conference, over a PSTN network and an IP

network, audio data from the PSTN conference endpoint would be communicated through a gateway that provides conversion of data between the two different networks:

The gateway [] provides conversion between analog and/or digital formats. The standard analog telephones [] communicate standard telephony signals through PSTN [network] to the gateway []. At the gateway [], the signals are converted to IP packets in the VoIP format. Similarly, VoIP packets received from the [Internet] network [] are converted into standard telephony signals for delivery to the destination telephone [] through PSTN [network].

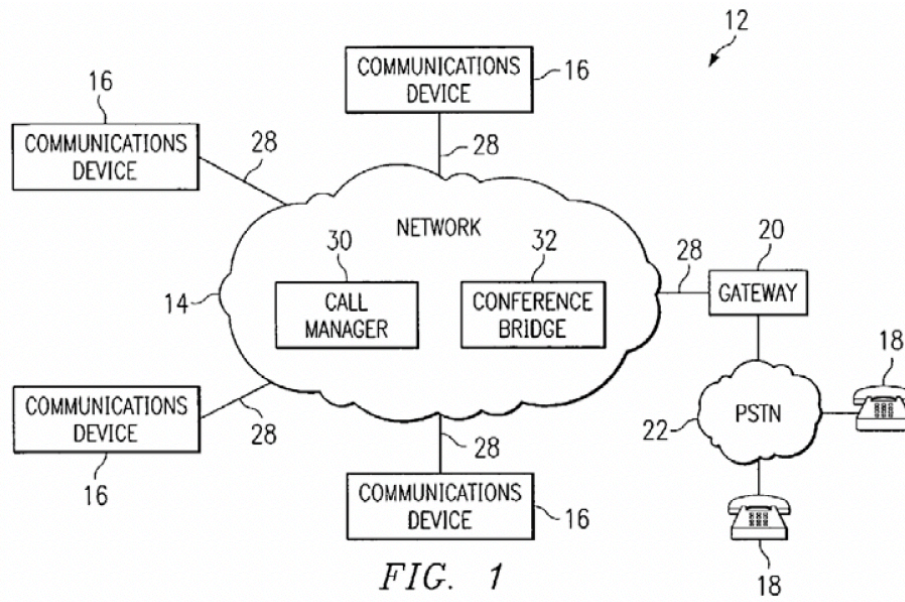
Id. at 4:18-25.

[A] system and method for improving the intelligibility of a moderator during a multi-party communications session includes **receiving a plurality of participant voice streams** from a plurality of respective conference participants.

Id. at 2:1-6.

[T]he **telephone 18 and gateway 20 combination [is] capable of communicating real-time audio, video** and/or other information over the network 14.

Id. at 3:65-4:1; *see also id.* at 3:50-58. Knappe's Figure 1 is reproduced below, illustrating that the gateway interfaces between the Internet network 14 and the PSTN 22:

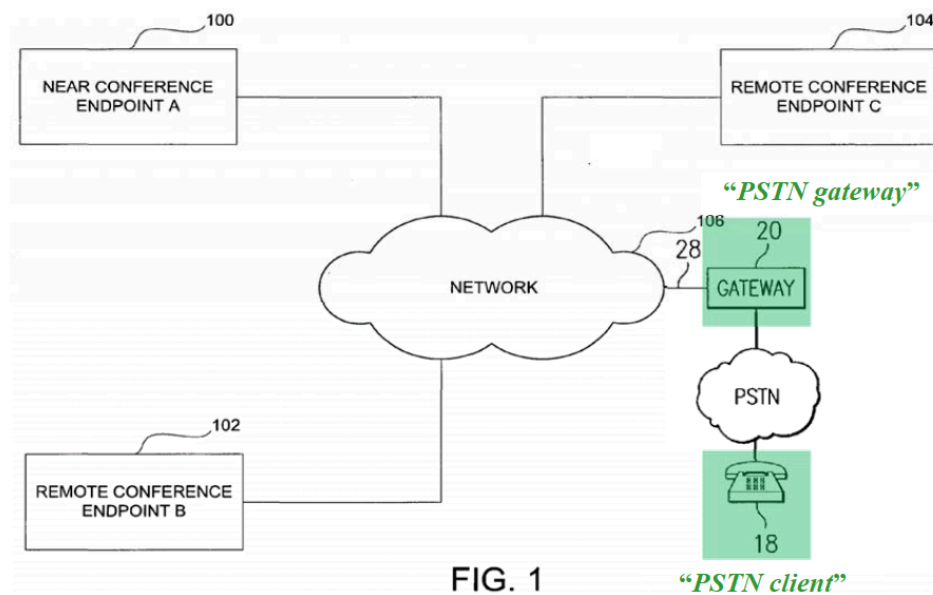


Id. at FIG. 1.

223. In view of Knappe’s teachings, it would have been obvious to a POSITA to utilize in Drell’s system a gateway that receives standard audio data from PSTN conference endpoints and converts the audio data into a format suitable for conference endpoints on an Internet network. Doing so would beneficially allow Drell’s PSTN conference endpoints to communicate, during the conference, audio data to the near conference endpoint 100, over a PSTN network and an Internet network 106. *See also* Reasons to Combine Drell and Knappe, *supra* ¶¶ 137-157.

224. The combined teachings of Drell and Knappe on this point are illustrated in the annotated Figure 1 below, showing how a POSITA would have found obvious to apply Knappe’s teaching of interconnecting the PSTN and the

Internet to Drell's system. As illustrated, a POSITA would understand that Drell's PSTN conference endpoint would transmit audio data through a gateway, as Knappe teaches.



Drell (Ex. 1008) at FIG. 1 (annotated).

225. Thus, Drell in combination with Doganata and Knappe teaches that the moderator near conference endpoint 100 receives audio data from a PSTN conference endpoint through a gateway, which renders obvious *“receiving first audio data at the local moderator client from the PSTN client through a PSTN gateway,”* as claimed.

[10.3.1] receiving second audio data at the local moderator client from a remote client;

226. The prior art renders obvious receiving second audio data at the local moderator client from a remote client.

227. **First**, as discussed at [10.1.1]-[10.1.2], Drell in combination with Knappe and Doganata teaches a moderator near conference endpoint 100.

228. **Second**, Drell teaches that an audio mixing module 308, of the moderator near conference endpoint, receives audio data from remote endpoint 102:

FIG. 1 depicts an exemplary operating environment of the multi-point (MP) conferencing application of the present invention. A near conference endpoint 100, embodying the MP conferencing application, is coupled to remote conference endpoints 102 and 104 via a network 106. **Remote conference endpoints 102 and 104 may comprise, for example, conventional videoconferencing devices equipped to transmit and receive both video (image) data and audio (speech) data.**

...

Network 106 may be of any type suitable for the transmission of **audio and video data** between and among near conference endpoint 100 and remote conference endpoints 102 and 104.

Drell (Ex. 1008) at 3:4-23.

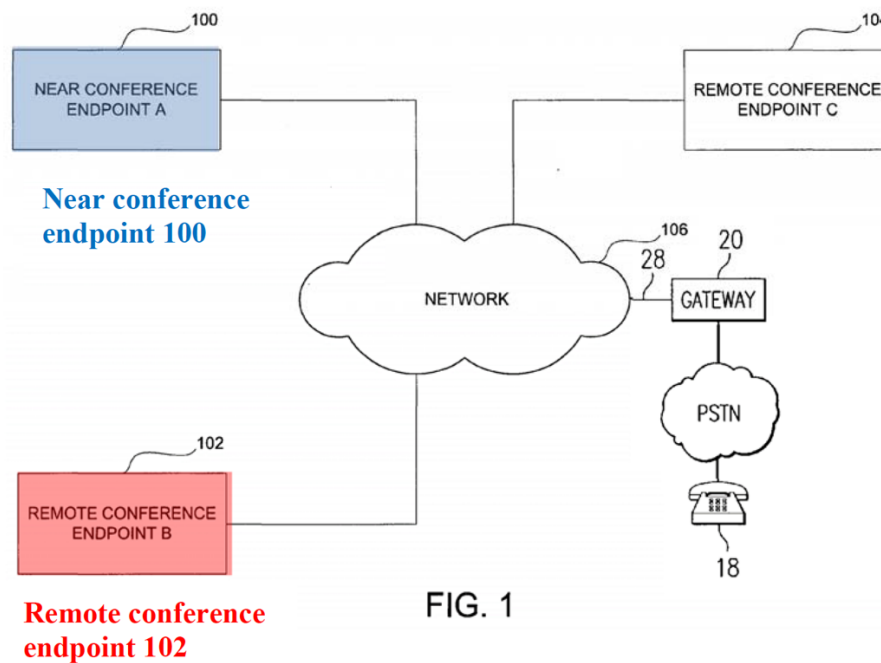
As will be described in further detail below, MP conferencing application 234 is configured to instantiate a processing train for each remote conference endpoint 102 and 104 to which near conference endpoint 100 is connected. **The processing trains process audio and video data streams received from remote conferencing endpoints 102 and 104.**

Id. at 4:21-26.

Audio mixing module 308 is configured to combine audio data received from remote conference endpoints 102 and 104 with locally generated audio data (received from audio I/O interface 212 via audio data path 374, and typically being representative of the speech of the near conference participant(s)). The term “combine” is used in its broadest and most general sense and is intended to cover any operation wherein audio mixing module 308 generates an output

audio data stream (or plurality of output audio data streams) based on information contained in the remotely and locally generated audio data input streams. For example, audio mixing module 308 may simply mix the received audio input data streams.

Id. at 5:8-20; *see also id.* at Abstract, 2:34-52, 3:45-48. Remote conference endpoint 102 corresponds to “a remote client,” as claimed.



Id. at FIG. 1 (annotated).

229. As shown below at Drell’s Figure 3, the MP conferencing application 234 includes processing train 302 and audio mixing module 308 that receive audio data from remote conference endpoint 102:

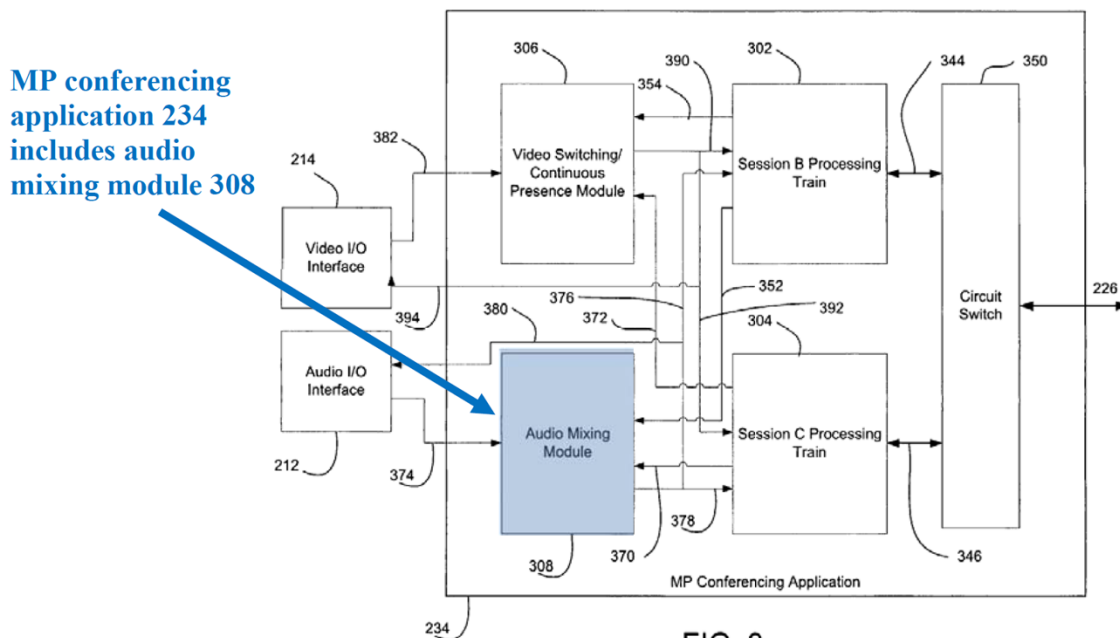


FIG. 3

Id. at FIG. 3 (annotated).

230. Thus, Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 receives audio data from the remote conference endpoint 102, which renders obvious “*receiving second audio data at the local moderator client from a remote client,*” as claimed.

[10.3.2] through a Virtual Private Network (VPN) tunnel;

231. The prior art renders obvious that the second audio data is received through a Virtual Private Network (VPN) tunnel.

232. **First**, as discussed at [10.3.1], Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 receives audio data from the remote conference endpoint 102.

233. **Second**, Drell teaches that network 106, such as the Internet is suitable for transmission of audio data between the moderator near conference endpoint 100 and the remote conference endpoint 102:

Network 106 may be of any type suitable for the transmission of audio and video data between and among near conference endpoint **100 and remote conference endpoints 102** and 104. Typically, network 106 will comprise the public switched telephone network (PSTN) or comparable circuit switched network to which each of the conference endpoints is connected by one or more ISDN lines. A multi- point conference is initiated by establishing a connection between near conference endpoint 100 and remote conference endpoint 102, and between near conference endpoint 100 and remote conference endpoint 104.

Id. at 3:20-30.

Alternatively, **network 106 may comprise a packet switched network, such as the Internet**. Although a single network 106 is shown, the invention 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.

234. Further, Elliott teaches a conferencing system that transmits audio, video, and collaboration data over Internet network:

Multi-media collaboration refers to the use of more than one media stream (e.g.: **voice, fax, data, video, etc.) used in collaboration** with more than one party.

Elliott (Ex. 1007) at 2:66-3:1.

Other types of collaboration...data conferencing can also be performed.

Id. at 6:53-54.

As alluded to above, communications between the parties can be made over a variety of networks 18. A network 18 is at least two (and preferably more) computers interconnected together so that communication between them is possible. In the illustrated embodiment, the public Internet is shown as an exemplary Internet Protocol-based (IP-based) network 18 for multi-media collaboration.

Id. at 4:6-12.

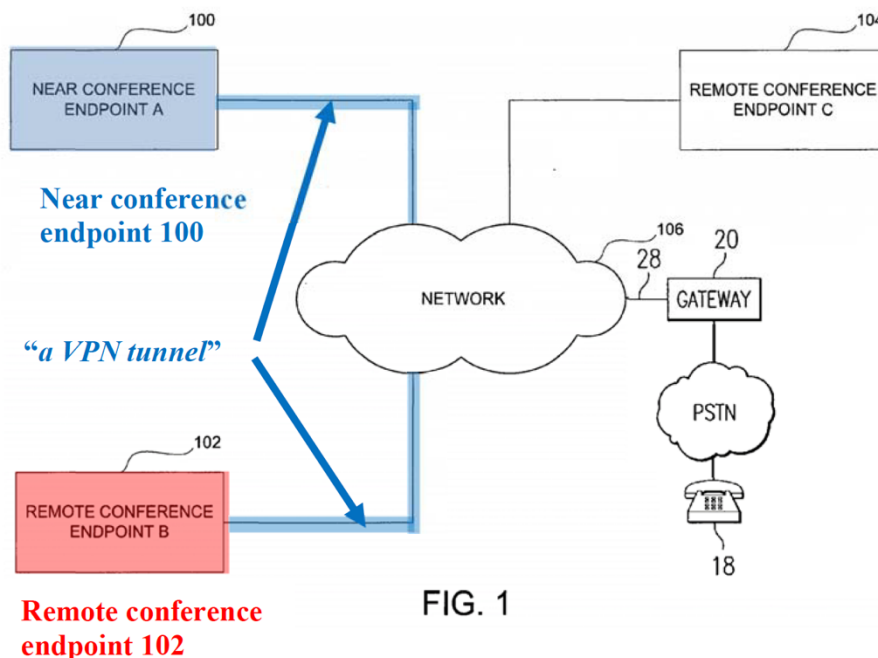
235. Furthermore, Elliott teaches that the data is transmitted over Internet 18 using VPN tunneling:

For example, network 18 can be either a public or a carrier supported virtual private network (VPN)...which uses IP or IP- tunneling.

Id. at 4:31-34.

236. Based on the above disclosure, a POSITA would have understood that Elliott utilizes VPN tunneling for communicating conferencing 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 utilize VPN tunneling when communicating audio, video, and collaboration data to and from the moderator near conference endpoint 100, because it would provide Drell with secure communications as the communications traverse Internet network 106. *See also* Reasons to Combine Drell and Elliott, *supra* ¶¶ 158-172.

237. As shown below in annotated Figure 1, when applying Elliot's teachings to Drell, communications from the moderator near conference endpoint 100 to the remote conference endpoint 102 would take place over a VPN tunnel:



Drell (Ex. 1008) at FIG. 1 (annotated).

238. Thus, Drell in combination with Knappe, Doganata, and Elliott teaches that the moderator near conference endpoint 100 receives audio data from the remote conference endpoint 102 through a VPN tunnel, which renders obvious that audio data is received from the remote client “*through a Virtual Private Network (VPN) tunnel,*” as claimed.

[10.4.1] transmitting third audio data, first video data, and first collaboration data from the local moderator computer to the remote client

239. The prior art renders obvious transmitting third audio data, first video data, and first collaboration data from the local moderator client to the remote client.

240. **First**, as discussed at [10.1.1]-[10.1.2], Drell in combination with Knappe and Doganata teaches a moderator near conference endpoint 100. Further, as discussed at [10.3.1], Drell teaches a remote conference endpoint 102.

241. **Second**, Drell teaches that the audio mixing module 308 of the moderator near conference endpoint generates mixed audio data:

Audio mixing module 308 is configured to combine audio data received from remote conference endpoints 102 and 104 with locally generated audio data (received from audio I/O interface 212 via audio data path 374, and typically being representative of the speech of the near conference participant(s)). The term “combine” is used in its broadest and most general sense and is intended to cover any operation wherein **audio mixing module 308 generates an output audio data stream** (or plurality of output audio data streams) based on information contained in the remotely and locally generated audio data input streams. For example, audio mixing module 308 may simply mix the received audio input data streams.

Drell (Ex. 1008) at 5:8-23; *see also id.* at Abstract, 2:34-52, 3:25-31. Locally generated audio is input to the near conference endpoint 100 using a “microphone” that generates “audio signals representative of the... speech of the near conference participant.” *Id.* at 3:45-48.

MP conferencing application 234 includes audio mixing module 308

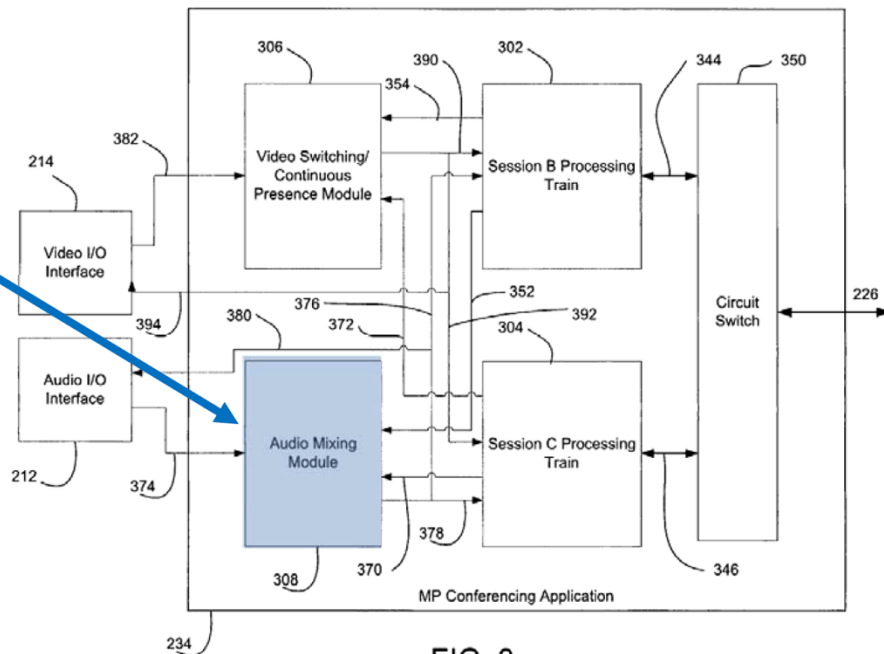


FIG. 3

Id. at FIG. 3 (annotated).

242. A POSITA would understand that the audio data received from the PSTN conference endpoint would also be combined in the audio mixing module 308.

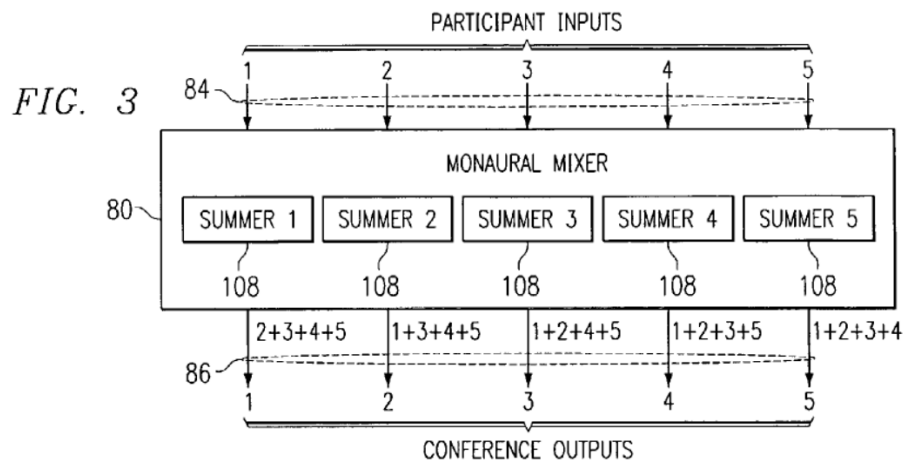
243. Knappe further supplements and explains how to mix audio data from multiple conference endpoints, including PSTN conference endpoints. *See* Knappe (Ex. 1006) at FIGs. 1-3, 5:49-62, 6:59-7:15. Specifically, in connection with Figure 3, Knappe teaches using a mixer 80 with a plurality of summers (e.g., summers 1 to 5), where each summer mixes audio data from the other participants (e.g., participants 1 to 5) while providing higher prominence to the moderator audio:

Referring to FIG. 3, the monaural mixer 80 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. In one embodiment, each participant is assigned a summer 82 that receives audio input streams from each other participant to the conference call. The summer 82 combines the audio input streams to generate a conference output stream for delivery to the participant.

During normal operation, each participant receives the audio input of each other participant. Thus, for example, the conference output stream of participant 1 includes the audio inputs of participants 2–5. Similarly, the conference output stream of participant 2 includes the audio inputs of participants 1 and 3–5. The conference output stream of participant 3 includes the audio inputs of participants 1–2 and 4–5. The conference output stream of participant 4 includes the audio inputs of participants 1–3 and 5. The conference output stream of participant 5 includes the audio inputs of participants 1–4.

The audio input 84 of the **conference moderator may be amplified** and/or the audio input 84 of the remaining participants attenuated to focus on or **provide higher prominence** to the audio input 84 of the conference moderator.

Id. at 6:59-7:15.



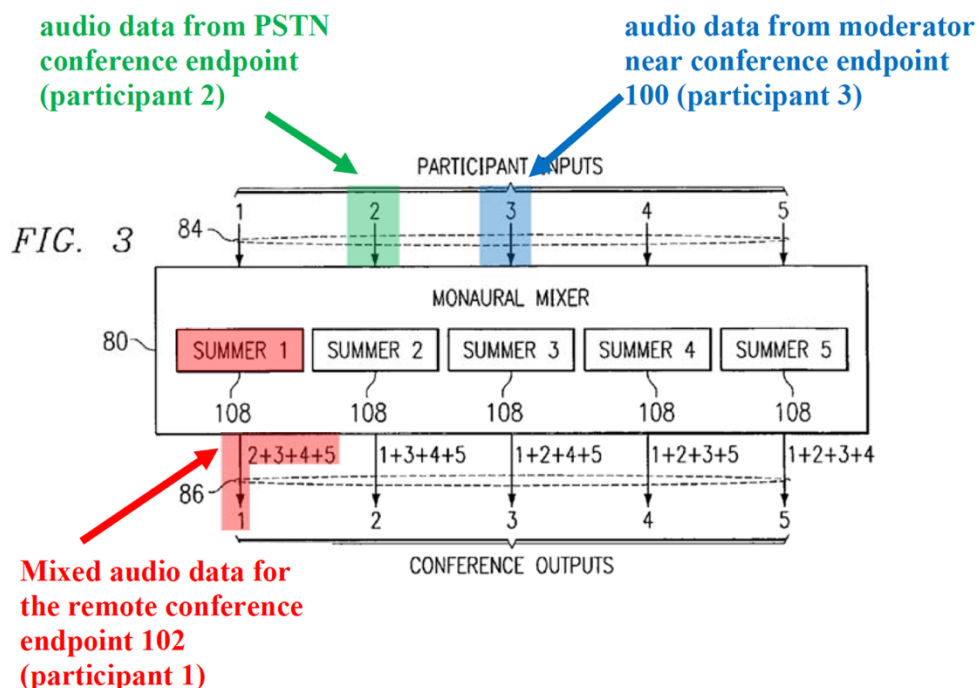
Id. at FIG. 3.

244. It would have been obvious to a POSITA to utilize in the mixing module 308 of the MP conferencing application 234 a plurality of summers, as Knappe teaches. Using a plurality of summers (one for each participant) would provide for separately mixing audio data from the PSTN conference endpoint with audio data from the moderator near conference endpoint 100, while providing higher prominence to the moderator audio and increasing the intelligibility. *See, e.g., id.* at 2:40-46, 7:11-17. By enhancing the intelligibility of the moderator during the conference, Knappe's teachings beneficially enhance the moderator's control over the conference. *See, e.g., id.* at 1:59-67. Additionally, Knappe's teachings would also provide the benefit of reducing undesirable audio effects since each participant's mixed audio excludes its own audio data. *See* U.S. Patent No. 7,266,091 (Ex. 1034) at 6:31-33 ("It is undesirable for a conference participant to receive his or her own audio in the mixed audio stream."); *see also* Reasons to Combine Drell and Knappe, *supra* ¶¶ 137-157.

245. In the combination, Drell's audio mixing module 308 would utilize a plurality of summers (one summer per participant) to obtain the functionality of generating a separate mix for each participant, which is consistent with Drell's desire to generate a separate instance for each remote endpoint. *See* Drell (Ex. 1008) at 6:66-7:4 ("By dynamically generating a separate instance of a processing train for each remote endpoint session, a videoconferencing system embodying the

invention may easily and flexibly accommodate endpoint sessions comprising a range of connection bandwidths and communication protocols.”).

246. For explanation purposes, and without limiting Knappe’s disclosure, I have provided below an annotated example of Figure 3 identifying which of Drell’s endpoints corresponds to which of participants 1-5. In the figure below, participant 2 corresponds to the PSTN conference endpoint that provides audio data to input 2 of mixer 80, and participant 3 corresponds to the moderator near conference endpoint 100 that provides audio data to input 3 of mixer 80, which summer 1 mixes into a mixed audio data comprising audio from participants 2+3+4+5:



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

247. In the combination of Drell and Knappe, the mixed audio data output by the audio mixing module 308, for the remote conference endpoint 102, corresponds to a “*third audio data*,” since it is a mix of locally generated audio from the moderator and the audio from the PSTN conference endpoint and does not include the remote conference endpoint 102 audio.

248. **Third**, Drell teaches that video switching/continuous presence module 306 of the multi-point (MP) conferencing application 234 outputs video data of the currently speaking participant:

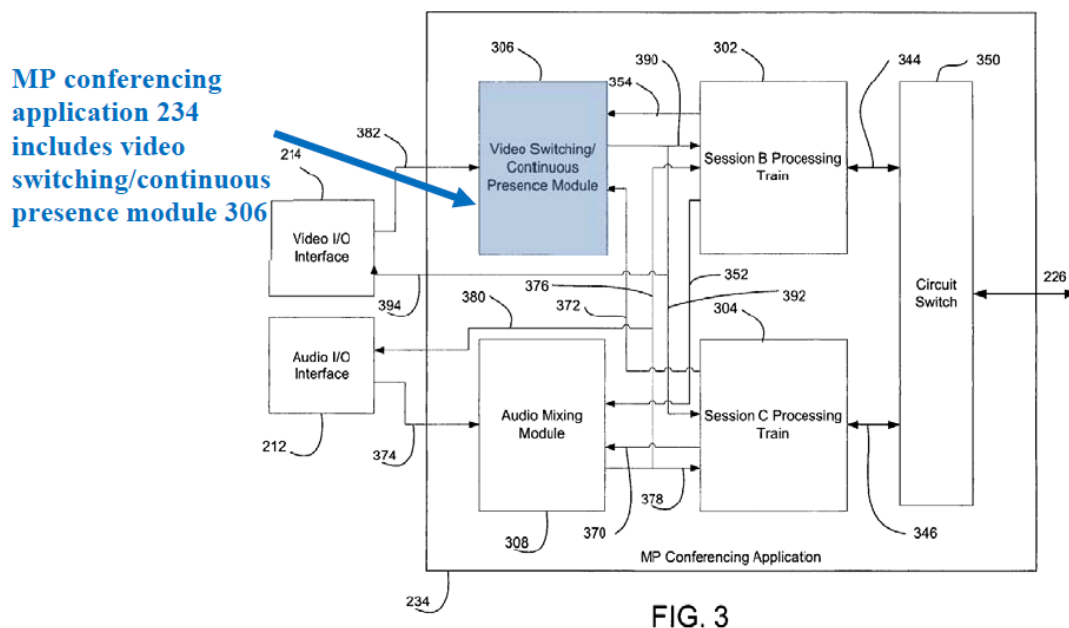
FIG. 3 is a block diagram showing the various components of an embodiment of MP conferencing application 234 and the flow of data between and among the various components. MP conferencing application 234 includes a circuit switch 350, a plurality of processing trains 302 and 304, **a video switching/continuous presence module 306**, and an audio mixing module 308.

Drell (Ex. 1008) at 4:32-38.

Video switching/continuous presence module 306 combines video data received from remote conference endpoints 102 and 104 with locally generated video data (received from video I/O interface 214 via video data path 382, and being typically representative of images of the near conference participants). Again, the term “combine” is used in its broadest and most general sense. In one mode of operation, video switching/continuous presence module 306 may select one of the video data input streams for output based on predetermined criteria (for example, it may select for output the video data stream corresponding to the conference endpoint of the currently speaking participants. In a second mode of operation (referred to as the “continuous presence mode”), video switching/continuous presence module 306 may construct a composite image wherein images corresponding to conference endpoints are displayed in different sectors of the composite

image. **The video data stream output (or plurality of outputs) from video switching continuous presence module 306** is thereafter distributed to processing trains 302 and 304 and video I/O interface 214 via video data paths 390, 392 and 394.

Id. at 5:27-47; *see also id.* at Abstract, 2:34-52, 3:45-48, 4:25-31. A POSITA would have understood that when the moderator is speaking, the moderator's video data would be selected and output by the video switching continuous presence module 306.



Id. at FIG. 3 (annotated).

249. The selected video output by the video switching continuous presence module 306, e.g., video of moderator, corresponds to a “*first video data.*”

250. **Fourth**, Drell teaches that the processing train 302, of the moderator near conference endpoint 100, transmits the mixed audio data and selected video data to the remote conference endpoint 102, which receives the data:

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.

Id. at 2:43-48.

The outputs of the audio mixer and video switching module are sent to each of the plurality of signal processing trains, which process the combined signals according to a transmit mode for distribution to the remote endpoints over the network.

Id. at Abstract.

In the transmission mode, processing train 302 is configured to receive the audio...streams output by audio mixing module 308...The received data streams are then encoded and combined to form a mixed encoded audio/video data stream, and the encoded **audio/video data stream is transmitted.**

Id. at 5:48-54.

FIG. 1 depicts an exemplary operating environment of the multi-point (MP) conferencing application of the present invention. A near conference endpoint 100, embodying the MP conferencing application, is coupled to remote conference endpoints 102 and 104 via a network

Remote conference endpoints 102 and 104 may comprise, for example, conventional videoconferencing devices equipped to transmit and receive both video (image) data and audio (speech) data...It should be appreciated that while only two remote conference endpoints are depicted in FIG. 1 for the purpose of clarity, a greater number of remote conference endpoints may be accommodated by near conference endpoint 100.

Id. at 3:4-19; *see also id.* at FIGs. 1-5. Thus, Drell in combination with Knappe teaches that the near conference endpoint 100 transmits mixed audio data and selected video data to remote conference endpoint 102, which renders obvious “*transmitting third audio data, first video data... from the local moderator client to the remote client.*”

251. **Fifth**, Drell teaches that the processing train 302, of the near conference endpoint, utilizes a data codec to encode “still images” and transmits the still images to remote conference endpoint 102:

It is noted that, while not depicted in the Figures, 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 102 and 104.**

Id. at 6:53-56.

The MP conference application is configured to generate, for each remote conference endpoint participating in a conference, discrete instances of a signal processing train by means of dynamically allocable IMUXs, each processing train including a communication process and audio/video/**data codecs**.

Id. at Abstract.

252. It would have been obvious to a POSITA that during the conference, the moderator near conference endpoint 100 receives still images from remote conference endpoint 104 and to facilitate collaboration transmits the still images from remote conference endpoint 104 to the other conference endpoints, including remote conference endpoint 102. Accordingly, Drell’s “still images” from remote

conference endpoint 104 which are transmitted to the remote conference endpoint 102, correspond to “*first collaboration data*,” as claimed. *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”).

253. **Additionally**, Drell teaches that its “communication process...may for example comprise an...H.323 (packet-based) process.” Drell (Ex. 1008) at 2:26-33. A POSITA would understand that employing an H.323 process also discloses “*first 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). This is consistent with the ’612 patent, which also relies on the prior art H.323/T.120 standards for transmitting collaboration data as part of the packetizing process. *See, e.g.*, ’612 patent (Ex. 1001) at 3:19-21 (“[T]he **H.323** recommendation defines a **data collaboration** capability as known and outlined in the **T.120 data collaboration unit**.”), FIG. 2 (Prior Art).

254. **Additionally**, Elliott teaches collaboration data, such as “application sharing” data, “web browsing” data, and collaboration data per the “T.120 standard”:

Multi-media collaboration services allow two or more parties to exchange information in a conference setting. These services can be readily provided over the Internet and include **collaborative Web browsing**, audio conferencing, video conferencing, and **application sharing**.

Elliott (Ex. 1007) at 1:40-44.

Once a session is established, multi-media collaboration between the agent and client is enabled as illustrated in step 208. For example, the Java applet running on the client computer 20 will monitor the client computer’s IP socket connection for incoming URLs. When a URL is pushed by an agent or another client, the applet will feed that URL to the client's Web browser, which will then go to the Web site addressed by the URL. **Other types of collaboration such as...data conferencing can also be performed.**

Id. at 6:46-54.

The T.120 standard contains a series of communication and application protocols and services that provide support for real-time, multi-point data communications. Session server 12 is preferably able to set up to conduct video conferences, audio conferences, and **data conferences with application sharing. Collaborative Web browsing** can also be enabled with URL push technology. During the course of a collaborative session, one party can push a URL to another party. In other words, one party can cause the browser of a remote party to connect to a certain web site.

Id. at 5:49-61.

...multi-media collaboration comprises **application sharing**.

Id. at 11:12-13.

...multi-media collaboration comprises collaborative **web browsing**.

Id. at 11:6-7.

255. Any one of Elliott’s types of collaboration data (such as application sharing, web browsing, and other collaboration data per the T.120 standard) corresponds with “*first collaboration data*,” as claimed.

256. It would have been obvious to a POSITA for the moderator near conferencing endpoint 100 to transmit collaboration data (e.g., application data, web browsing data, and other collaboration data per the T.120 standard) as Elliott teaches, to the remote conference endpoint 102, because including collaboration data in transmissions to other conference endpoints would allow different ways for the participants to collaborate thereby enhancing the effectiveness of the collaboration among the participants. *See* Reasons to Combine Drell and Knappe, *supra* ¶¶ 137-157; Reasons to Combine Drell and Elliott, *supra* ¶¶ 158-172.

257. Thus, Drell in combination with Knappe, Doganata, and Elliott teaches transmitting mixed audio data (mixed specifically for remote conference endpoint 102), selected video data, and collaboration data (e.g., still image data, application data, web browsing data, and other collaboration data per the T.120 standard) from the moderator near conference endpoint 100 to the remote conference endpoint 102, which renders obvious “*transmitting third audio data, first video data, and first collaboration data from the local moderator client to the remote client*,” as claimed.

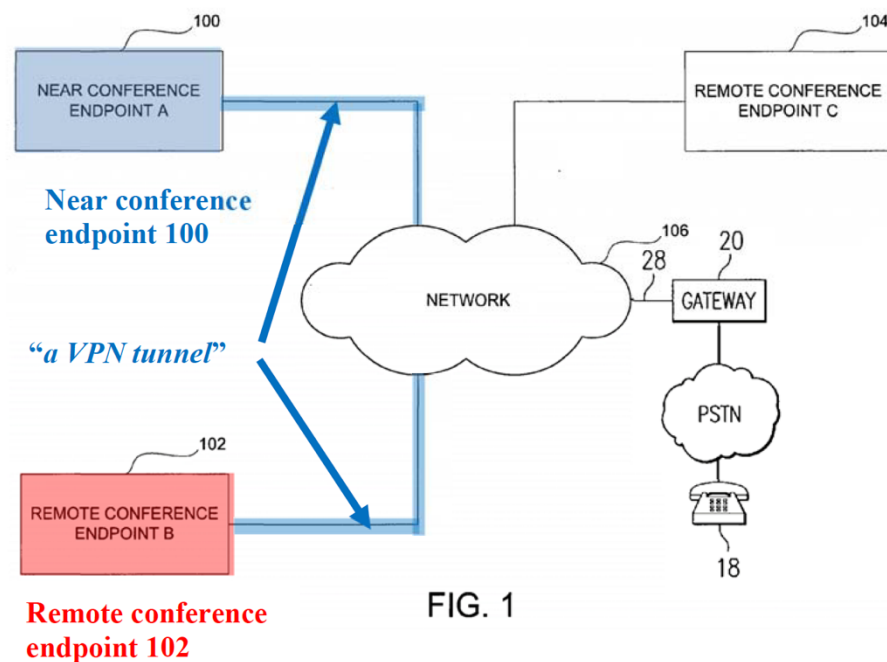
[10.4.2] through the first VPN tunnel

258. The prior art renders obvious that the third audio data, first video data, and the first collaboration data are transmitted to the remote client through the first VPN tunnel.

259. As a threshold matter, the phrase “*the first VPN tunnel*” is not previously recited in the claim, which only discusses “*a VPN tunnel*” (see element [10.3.2]). For analysis purposes only, I will assume that the claimed “*first VPN tunnel*” in [10.4.2] corresponds to the “*VPN tunnel*” discussed at [10.3.2].

260. **First**, as discussed at [10.4.1], Drell in combination with Knappe, Doganata, and Elliott teaches that the moderator near conference endpoint 100 transmits mixed audio data, selected video data, and collaboration data to the remote conference endpoint 102.

261. **Second**, as discussed at [10.3.2], Drell in combination with Knappe, Doganata, and Elliott, teaches that the moderator near conference endpoint 100 communicates with the remote conference endpoint 102 through the VPN tunnel. Specifically, in the combination, the VPN tunnel is established between the moderator near conference endpoint 100 and the remote conference endpoint 102:



Drell (Ex. 1008) at FIG. 1 (annotated).

262. As illustrated above, when the moderator near conference endpoint 100 transmits data to or receives data from the remote conference endpoint 102, the data would be received and sent through the same VPN tunnel, which corresponds to the claimed “*VPN tunnel*.” It would have been obvious to a POSITA to utilize the VPN tunnel for communications between the moderator near conference endpoint 100 and the remote conference endpoint 102, because it would provide Drell secure communications. *See also* Reasons to Combine Drell and Elliott, *supra* ¶¶ 158-172.

263. Thus, Drell in combination with Knappe, Doganata, and Elliott teaches that the moderator near conference endpoint 100 transmits the mixed audio data, selected video data, and collaboration data to the remote conference endpoint

102 through the VPN tunnel, which renders obvious that third audio data, first video data, and first collaboration data are transmitted to the remote client “*through the first VPN tunnel,*” as claimed.

[10.5] transmitting fourth audio data from the local moderator client to the PSTN client through the PSTN gateway.

264. The prior art renders obvious transmitting fourth audio data from the local moderator client to the PSTN client through the PSTN gateway.

265. **First**, as discussed at [10.1.1]-[10.2], Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 communicates with the PSTN conference endpoint through the PSTN gateway.

266. **Second**, as noted previously, Drell teaches that the audio mixing module 308 mixes the locally generated audio from the near conference endpoint 100 with the audio received from the remote conference endpoint 102:

Audio mixing module 308 is configured to combine audio data received from remote conference endpoints 102 . . . with locally generated audio data (received from audio I/O interface 212 via audio data path 374, and typically being representative of the speech of the near conference participant(s)). The term “combine” is used in its broadest and most general sense and is intended to cover any operation wherein audio mixing module 308 generates an output audio data stream (or plurality of output audio data streams) based on information contained in the remotely and locally generated audio data input streams. For example, **audio mixing module 308 may simply mix the received audio input data streams.**

Drell (Ex. 1008) at 5:8-23; *see also id.* at Abstract, 2:34-52, 3:25-38; FIGS. 1-3.

267. Knappe supplements and explains how to mix audio data from multiple conference endpoints, including remote endpoints. *See* Knappe at FIGs. 1-3, 5:49- 62, 6:59-7:15. Specifically, in connection with Figure 3, Knappe teaches using a mixer 80 with a plurality of summers (e.g., summers 1 to 5), where each summer mixes audio data from the other participants (e.g., participants 1 to 5) while providing higher prominence to the moderator audio:

Referring to FIG. 3, the monaural mixer 80 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. In one embodiment, each participant is assigned a summer 82 that receives audio input streams from each other participant to the conference call. The summer 82 combines the audio input streams to generate a conference output stream for delivery to the participant.

During normal operation, each participant receives the audio input of each other participant. Thus, for example, the conference output stream of participant 1 includes the audio inputs of participants 2–5.

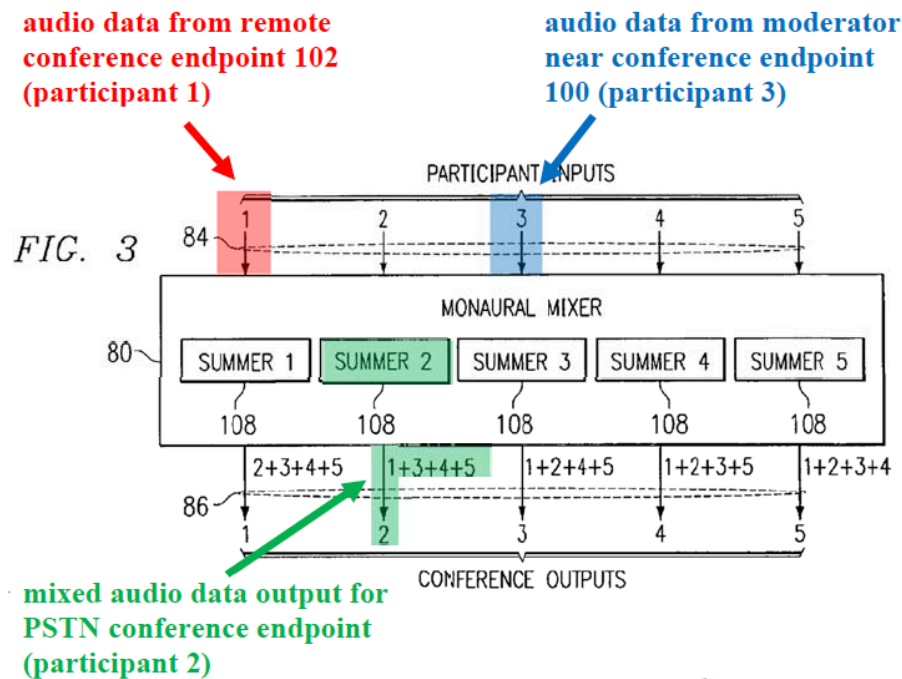
Similarly, the conference output stream of participant 2 includes the audio inputs of participants 1 and 3–5. The conference output stream of participant 3 includes the audio inputs of participants 1–2 and 4–5. The conference output stream of participant 4 includes the audio inputs of participants 1–3 and 5. The conference output stream of participant 5 includes the audio inputs of participants 1-4.

The audio input 84 of the conference moderator may be amplified and/or the audio input 84 of the remaining participants attenuated to focus on or **provide higher prominence to the audio input 84 of the conference moderator.**

Knappe (Ex. 1006) at 6:59-7:15; FIG. 3.

268. As noted previously, it would have been obvious to a POSTIA to use a plurality of summers in Drell's mixing module 308 (one summer for each participant) for the reasons stated in *supra* ¶¶ 244-245.

269. For explanation purposes, and without limiting Knappe's disclosure, I have provided below an annotated example of Figure 3 identifying which of Drell's endpoints corresponds to which of participants 1-5. In the figure below, participant 1 corresponds to the remote conference endpoint 102 that provides audio data to input 1 of mixer 80, and participant 3 corresponds to the moderator near conference endpoint 100 that provides audio data to input 3 of mixer 80, which summer 2 mixes into mixed audio data comprising audio from participants 1+3+4+5:



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

270. In the combination of Drell and Knappe, the mixed audio data output by the audio mixing module 308, for the PSTN conference endpoint, corresponds to a “*fourth audio data*,” since it is a mix of locally generated audio from the moderator and the audio from the remote conference endpoint 102.

271. **Third**, Drell teaches the moderator near conference endpoint 100 transmits data over the Internet and PSTN to the PSTN conference endpoint:

Network 106 may be of any type suitable for the **transmission of audio and video data between and among near conference endpoint 100 and remote conference endpoints 102 and 104**....network 106 may comprise a packet switched network, such as the Internet. Although a single network 106 is shown, **the invention contemplates the use of two or more networks (for example, the PSTN and the Internet) to connect conference endpoints** utilizing different communication protocols.

Drell (Ex. 1008) at 3:20-38.

The outputs of the audio mixer and video switching module are sent to each of the plurality of signal processing trains, which process the combined signals according to a **transmit mode for distribution to the remote endpoints over the network**.

Id. at Abstract.

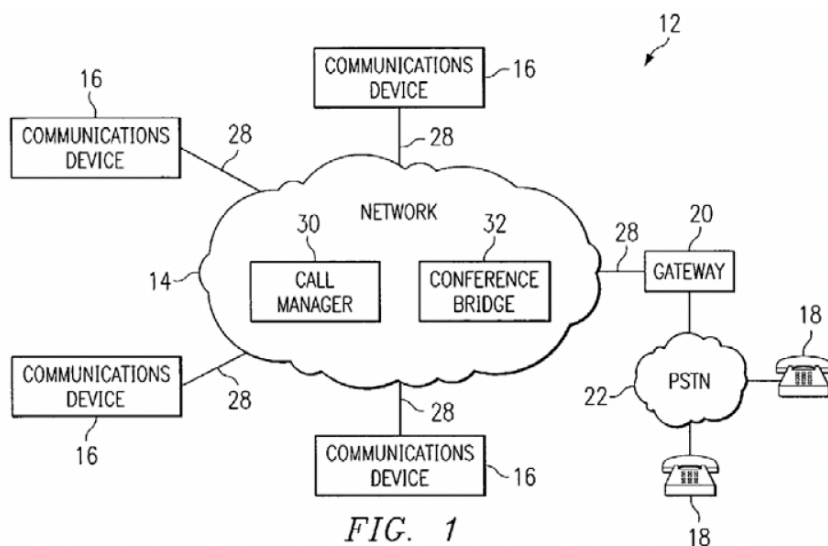
In the transmission mode, processing train 302 is configured to receive the audio...streams output by audio mixing module 308...The received data streams are then encoded and **combined to form a mixed encoded audio/video data stream, and the encoded audio/video data stream is transmitted**.

Id. at 5:48-54.

272. Knappe supplements Drell and teaches how data is transmitted to the endpoints in the PSTN. Specifically, Knappe teaches that the PSTN gateway converts data into a suitable format for the PSTN conference endpoint:

The gateway [] provides conversion between analog and/or digital formats. The standard analog telephones [] communicate standard telephony signals through PSTN [network] to the gateway []. At the gateway [], the signals are converted to IP packets in the VoIP format. **Similarly**, VoIP packets received from the [Internet] network [] are converted into standard telephony signals for delivery to the destination telephone [] through PSTN [network].

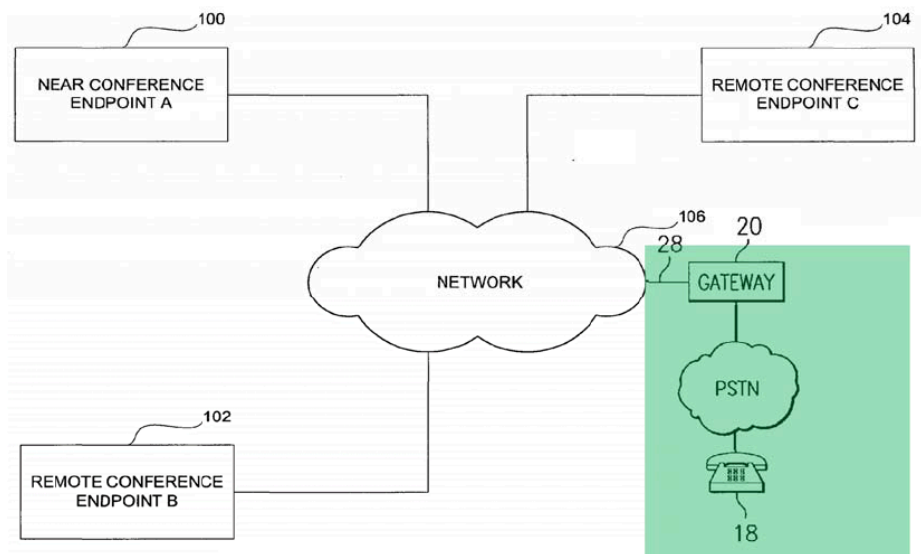
Knappe (Ex. 1006) at 4:18-25, FIG. 1; *see also id.* at 3:50-58. This is illustrated in Knappe's Figure 1, where gateway 20 is positioned between Internet network 14 and PSTN 22 to perform conversion between analog and/or digital formats for each network:



Id. at FIG. 1.

273. Since Drell contemplates that its conference endpoints communicate over “PSTN and the Internet” networks (Drell (Ex. 1008) at 3:20-38), it would have been obvious to a POSITA to utilize a gateway positioned between the PSTN and Internet networks, as Knappe teaches, because it would provide Drell the benefit of converting IP packets from the Internet network 106 to suitable signals for the PSTN network, and vice-versa (Knappe (Ex. 1006) at 4:18-25). *See also* Reasons to Combine Drell and Knappe, *supra* ¶¶ 137-157.

274. In the combination, it would have been obvious to a POSITA that after Drell’s moderator near conference endpoint 100 generates mixed audio data for the PSTN conference endpoint (see discussion directly above), the moderator near conference endpoint 100 would transmit packets including that mixed audio data over Internet network 106 to a gateway that converts the received packets to suitable signals that are delivered to the PSTN conference endpoint, consistent with Knappe’s teachings. *See* Knappe (Ex. 1006) at 4:18-25.



Drell (Ex. 1008) at FIG. 1 (annotated).

275. Thus, Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 transmits mixed audio data (mixed specifically for the PSTN conference endpoint) to the PSTN conference endpoint through a gateway, which renders obvious “*transmitting fourth audio data from the local moderator client to the PSTN client through the PSTN gateway,*” as claimed.

4. Claim 11

[11.P] The method of claim 10, further comprising:

276. See analysis for [10.P]-[10.5].

[11.1] receiving second video data at the local moderator client from the remote client

277. The prior art renders obvious receiving second video data at the local moderator client from the remote client.

278. **First**, as discussed at [10.1.1]-[10.1.2], Drell in combination with Knappe and Doganata teaches a moderator near conference endpoint 100. Further, as discussed at [10.3.1], Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 receives audio data from the remote conference endpoint 102.

279. **Second**, Drell further teaches that the moderator near conference endpoint 100 receives video data from the remote conference endpoint 102:

FIG. 1 depicts an exemplary operating environment of the multi-point (MP) conferencing application of the present invention. A near conference endpoint 100, embodying the MP conferencing application, is coupled to remote conference endpoints 102 and 104 via a network 106. **Remote conference endpoints 102 and 104 may** comprise, for example, conventional videoconferencing devices equipped to **transmit and receive both video (image) data and audio (speech) data**...Network 106 may be of any type suitable for the transmission of audio and video data between and among near conference endpoint 100 and remote conference endpoints 102 and 104.

Drell (Ex. 1008) at 3:4-23.

As will be described in further detail below, MP conferencing application 234 is configured to instantiate a processing train for each remote conference endpoint 102 and 104 to which near conference endpoint 100 is connected. The processing trains process **audio and video data streams received from remote conferencing endpoints 102 and 104**.

Id. at 4:21-26.

Video switching/continuous presence module 306 combines video data **received from remote conference endpoints 102 and 104** with locally generated video data (received from video I/O interface 214 via video data path 382, and being typically

representative of images of the near conference participants). Again, the term “combine” is used in its broadest and most general sense. In one mode of operation, video switching/continuous presence module 306 may select one of the video data input streams for output based on predetermined criteria (for example, it may select for output the video data stream corresponding to the conference endpoint of the currently speaking participants. In a second mode of operation (referred to as the “continuous presence mode”), video switching/continuous presence module 306 may construct a composite image wherein images corresponding to conference endpoints are displayed in different sectors of the composite image. The video data stream output (or plurality of outputs) from video switching continuous presence module 306 is thereafter distributed to processing trains 302 and 304 and video I/O interface 214 via video data paths 390, 392 and 394.

Id. at 5:27-47; *see also id.* at Abstract, 2:34-52, 3:45-48. The video data from the remote conference endpoint 102 corresponds with the claimed “*second video data.*”

280. As shown below at Drell’s Figure 3, the MP conferencing application 234 (of the moderator near conference endpoint 100) includes processing train 302 that receives video data from the remote conference endpoint 102 (*see id.* at 4:61-64) which is then provided to the video switching/continuous presence module 306 via path 354 (*id.* at 4:64-67):

MP conferencing application 234 includes processing train 302 which receives video data from remote conference endpoint 102 and provides the received video data to the switching/continuous presence module 306

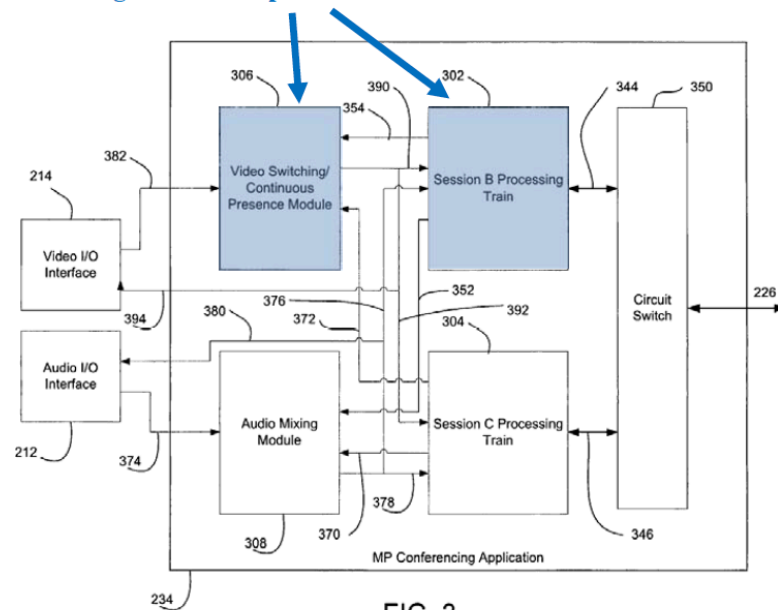


FIG. 3

Id. at FIG. 3 (annotated).

281. In view of the above disclosure, a POSITA would understand that in the combination during conferencing the moderator near conference endpoint 100, including the processing train 302 and the video switching/continuous presence module 306, receives video data from the remote conference endpoint 102.

282. Thus, Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 receives video data from the remote conference endpoint 102, which renders obvious “*receiving second video data at the local moderator client from the remote client,*” as claimed.

5. Claim 12

[12.P] *The method of claim 10, further comprising:*

283. See analysis for [10.P]-[10.5].

[12.1] receiving second collaboration data at the local moderator client from the remote client.

284. The prior art renders obvious receiving second collaboration data at the local moderator client from the remote client.

285. **First**, as discussed at [10.1.1]-[10.1.2], Drell in combination with Knappe and Doganata teaches a moderator near conference endpoint 100. Further, as discussed at [10.3.1], Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 receives audio data from the remote conference endpoint 102.

286. **Second**, Drell's MP conferencing application 234 includes a processing train 302 with a "data codec" used with still images:

It is noted that, while not depicted in the Figures, processing train 302 may include a **data codec for coding and encoding still images** and the like **received from ... remote conference endpoints 102** and 104.

Drell (Ex. 1008) at 6:53-56.

The MP conference application is configured to generate, for each remote conference endpoint participating in a conference, discrete instances of a signal processing train by means of dynamically allocable IMUXs, each processing train including a communication process and audio/video/**data codecs**.

Id. at Abstract.

287. The "still images," received from remote conference endpoint 102 correspond to "*second collaboration data*," as claimed. *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 (“Still Image”).

288. **Additionally**, Drell teaches that its “communication process...may for example comprise an...H.323 (packet-based) process.” *Id.* at 2:26-33. For the same reasons articulated with respect to “*first collaboration data*,” it would have been obvious to a POSITA that employing an H.323 process also discloses “*second collaboration data*,” because H.323 was a well-known standard that recommended T.120 collaboration data. *See supra* ¶ 253.

289. **Additionally**, as discussed at [10.4.1], Elliott teaches collaboration data, such as “application sharing” data, “web browsing” data, and collaboration data per the “T.120 standard.” *See, e.g.*, Elliott (Ex. 1007) at 1:40-44, 5:49-61, 6:46-54, 11:6-13. Any one of Elliott’s types of collaboration data (such as application sharing, web browsing, and other collaboration data per the T.120 standard), corresponds to the claimed “*second collaboration data*.”

290. It would have been obvious to a POSITA for the moderator near conferencing endpoint 100 to receive collaboration data (e.g., application data, web browsing data, and other collaboration data per the T.120 standard) as Elliott teaches, from the remote conference endpoint 102, because including collaboration data in transmissions to and from remote conference endpoints would allow different ways for the participants to collaborate thereby enhancing the

effectiveness of the collaboration among the participants. *See* Reasons to Combine Drell and Elliott, *supra* ¶¶ 158-172.

291. Thus, Drell in combination with Knappe, Doganata, and Elliott teaches receiving collaboration data (e.g., still image data, application data, web browsing data, and other collaboration data per the T.120 standard) from the remote conference endpoint 102, which renders obvious “*receiving second collaboration data at the local moderator client from the remote client,*” as claimed.

6. Claim 13

[13.P] The method of claim 10, further comprising:

292. See analysis for [10.P]-[10.5].

[13.1] wherein the PSTN client is a cellular telephone.

293. The prior art renders obvious that the PSTN client is a cellular phone.

294. **First**, as discussed at [10.1.1], Drell and Knappe disclose a PSTN conference endpoint.

295. **Second**, Knappe teaches implementing a PSTN conference endpoint as a wireless device that connects to PSTN 22 over wireless link 28:

FIG. 1 illustrates a communication system 12 in accordance with one embodiment of the present invention. In this embodiment, the communication system 12 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 **endpoints**.

Knappe (Ex. 1006) at 3:24-29.

Referring to FIG. 1, communication system 12 includes a network 14 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 (PSTN) 22. The communication devices 16, standard analog **telephones 18 and gateway 20 are connected to the...PSTN 22 through...radio frequency...and/or any other suitable type of...wireless links 28.**

Id. at 3:41-49.

296. Knappe further teaches that one type of a wireless link is a cellular link used by a cellular telephone:

The communication devices 16 comprise IP or other digital telephones, personal and other suitable computers or computing devices, personal digital assistants (PDAs), **cell or other mobile telephones** or handset or any other device or set of devices such as the telephone 18 and gateway 20 combination capable of communicating real-time audio, video and/or other information over the network 14.

Id. at 3:62-4:1.

297. Knappe's disclosure that telephone 18 communicates over the PSTN network 22 using "any...suitable type of...wireless link," and its example of a cellular telephone, which a POSITA would have understood uses a cellular wireless link, renders obvious that the PSTN conference endpoint is a "*cellular telephone.*"

298. In view of Knappe's teachings, it would have been obvious to a POSITA to utilize in Drell's system a PSTN conference endpoint that is a cellular phone. Doing so would beneficially allow Drell's PSTN conference endpoint to

communicate, during the conference, audio data to the moderator near conference endpoint 100, over a PSTN network and an Internet network 106 as the participant moves among different locations. *See also* Reasons to Combine Drell and Knappe, *supra* ¶¶ 137-157.

299. Thus, Drell in combination with Knappe teaches that the PSTN conference endpoint is a cellular telephone, which renders obvious that “*the PSTN client is a cellular phone.*”

7. Claim 14

[14.P] The method of claim 10, further comprising:

300. See analysis for [10.P]-[10.5].

[14.1] compressing the third audio data prior to the transmitting the third audio data.

301. The prior art renders obvious compressing the third audio data prior to the transmitting the third audio data.

302. **First**, as discussed at [10.4.1]-[10.4.2], Drell combined with Knappe teaches that the near conference endpoint 100 transmits mixed audio data (“*third audio data*”) to the remote conference endpoint 102 through a VPN tunnel.

303. **Second**, Drell teaches using a compression algorithm to compress mixed audio data before transmission to a remote endpoint:

In the transmit mode...[a]udio codec 408 encodes the audio data stream output by audio mixing module 308
(representative, for example, of the blended speech of conference

participants located at near conference endpoint 100 and remote conference endpoints 102 and 104) using a standard or proprietary audio **compression algorithm** (e.g., G.711) and delivers the encoded audio data to communication process 404.

Drell (Ex. 1008) at 6:33-45, FIG. 4.

By dynamically generating a separate instance of a processing train for each remote endpoint session, a videoconferencing system embodying the invention may easily and flexibly accommodate endpoint sessions comprising a range of connection bandwidths and communication protocols.

Id. at 6:66-7:4.

304. **Additionally**, Drell 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. 1025) at 4:35-56 (“**H.323** is adapted to support many aspects of multimedia... This standard also specifies lowest-common denominator **compression protocols**, such as G.711...G.723 and G.729.”). In fact, the ’612 patent’s compressor is simply a well-known prior art H.323 codec. *See, e.g.,* ’612 patent (Ex. 1001) at 2:63-3:1 (“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.”)

305. Thus, Drell teaches using a compression algorithm and/or compression protocols for compressing the mixed audio data before the mixed audio data is transmitted to the remote conference endpoint 102, which renders obvious “*compressing the third audio data prior to the transmitting the third audio data,*” as claimed.

8. Claim 15

[15.P] *The method of claim 10, further comprising:*

306. See analysis for [10.P]-[10.5].

[15.1] *wherein the second audio data received from the remote client is compressed digital audio data.*

307. The prior art renders obvious wherein the second audio data received from the remote client is compressed digital audio data.

308. **First**, as discussed at [10.3.1], Drell in combination with Knappe and Doganata teaches a moderator near conference endpoint 100 receives “*second audio data*” from the remote conference endpoint 102.

309. **Second**, Drell teaches that the audio data from the remote conference endpoint 102 is decoded by audio codec 408 (using a standard or proprietary compression algorithm):

FIG. 4 depicts components of an exemplary processing train 302. **Processing train 302 includes a communication process 404 and**

video and audio codecs 406 and 408. In the receive mode, the combined data stream 344 is directed to communication process 404 which carries out a predetermined set of functions with respect to data stream 344

Drell (Ex. 1008) at 5:64-6:2.

Audio codec 408 receives the audio data stream from communication process 404 and applies redundancy reduction **decoding in accordance with a standard (e.g., ITU Recommendation G.711) or proprietary audio compression algorithm.** The decoded audio data stream is then sent to audio mixing module 308, as described above.

Id. at 6:18-23; *see also id.* at FIGs. 3, 4. As already discussed at [10.3.1], processing train 302 (including audio codec 408), is utilized to receive and decode audio from the remote conference endpoint 102. *See, e.g., id.* at 4:46-47 (“...processing trains 302 and 304 respectively corresponding to remote conference endpoints 102 and 104.”). A POSITA would have understood that audio codec 408 (of the processing train 302) decoding audio using an audio compression algorithm is performing decompression of the audio data from the remote conference endpoint 102.

310. Thus, it would have been obvious to a POSITA that the audio received from the remote conference endpoint 102 (discussed at [1.3.1]) is “*compressed...audio data*,” since it is decompressed by audio codec 408 using a standard or proprietary algorithm.

311. **Additionally**, as noted in *supra* ¶ 304, Drell teaches that the communication process may include a standard process such as H.323. *See id.* at

2:26-28. Thus, it would have been obvious to a POSITA that performing the H.323 process includes using the compression protocols and codecs described in the standard. *See supra* ¶ 304.

312. **Third**, Drell teaches that the near conference endpoint 100 receives the audio data over a packet switched network, such as the Internet:

[N]etwork 106 may comprise a **packet** switched network, such as the **Internet**. Although a single network 106 is shown, the invention contemplates the use of two or more networks (for example, the PSTN and the Internet) to connect conference endpoints utilizing different communication protocols.

Drell (Ex. 1008) at 3:33-38, FIG. 4; *see also id.* at 2:26-28 (“The communication process...may for example comprise an...H.323 (**packet-based**) process.”).

313. Further, Knappe teaches that data communicated over the Internet is digitally encoded. Knappe (Ex. 1006) at 4:11-14 (“The **digital** encoded data is encapsulated into IP [Internet Protocol] 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.”). It would have been obvious to a POSITA that Drell’s system transmits data over Internet network 106 in digital form because communications over the Internet utilize digital data encapsulated in packets. *See, also* Caputo (Ex. 1021) at 12-13. Thus, it would have been obvious to a POSITA that when the moderator near conference endpoint 100 receives audio data from the remote conference endpoint 102 over the Internet network 106 (see

discussion at [1.3.1]), the audio data is “*digital*,” as claimed. I also note that the ’612 patent specification’s only reference to “digital” data is a description of how “digital IP is routed via Internet.” ’612 patent (Ex. 1001) at 6:18-20.

314. Thus, Drell in combination with Knappe and Doganata teaches that the audio data received from the remote conference endpoint 102 is compressed digital audio data, which renders obvious that “*the second audio data received from the remote client is compressed digital audio data*,” as claimed.

9. Claim 16

[16.P] The method of claim 15, further comprising:

315. See analysis for [15.P]-[15.1].

[16.1] decompressing the second audio data received from the remote client.

316. The prior art renders obvious decompressing the second audio data received from the remote client.

317. As discussed at [15.1], Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 receives compressed digital audio data from the remote conference endpoint 102. Further, as discussed at [15.1], Drell teaches that the received compressed digital audio data is decompressed with a compression algorithm by the processing train 302, including audio codec 408. *See, e.g.,* Drell (Ex. 1008) at 6:18-23, FIGs. 3, 4. As an additional example discussed at [15.1], Drell teaches that its communication

process includes a H.323 (packet-based) process which renders obvious decompressing the received compressed digital audio data from the remote conference endpoint 102. *See, e.g., id.* at 2:17-33; U.S. Patent No. 6,553,515 (Ex. 1025) at 4:35-56; U.S. Patent No. 6,850,496 (Ex. 1024) at 3:56-67. It would have been obvious to a POSITA to decompress the received audio data because decompression is required before the audio can be played back to a user.

318. Thus, Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 uses a compression algorithm and/or protocol to decompress the compressed digital audio data received from the remote conference endpoint 102, which renders obvious “*decompressing the second audio data received from the remote client,*” as claimed.

10. Claim 17

[17.P] *The method of claim 16, further comprising:*

319. See analysis for [16.P]-[16.1].

[17.1] *mixing the second audio data received from the remote client with the first audio data received from the PSTN client.*

320. The prior art renders obvious mixing the second audio data received from the remote client with the first audio data received from the PSTN client.

321. **First**, as discussed at [10.2]-[10.3.1], Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 receives audio data from the PSTN conference endpoint and also receives audio

data from the remote conference endpoint 102. Further, as discussed at [16.1], the audio data from the remote conference endpoint 102 is decompressed by the processing train 302 of the moderator near conference endpoint 100.

322. **Second**, Drell teaches that the audio mixing module 308 (of the near conference endpoint 100) mixes the audio received from conference endpoints:

The term “combine” is used in its broadest and most general sense and is intended to cover any operation wherein audio mixing module 308 generates an output audio data stream (or plurality of output audio data streams) based on information contained in the remotely and locally generated audio data input streams. For example, **audio mixing module 308 may simply mix the received audio input data streams.**

Drell (Ex. 1008) at 5:13-21.

By **dynamically generating a separate instance** of a processing train for each remote endpoint session, a videoconferencing system embodying the invention may easily and flexibly accommodate endpoint sessions comprising a range of connection bandwidths and communication protocols.

Id. at 6:66-7:4; *see also id.* at Abstract, 2:34-52, 3:45-48, 4:25-38, FIGs. 3, 4. As discussed at [10.4.1], it would have been obvious to a POSITA that the audio data received from the PSTN conference endpoint would also be combined in the audio mixing module 308.

323. Further, as discussed at [10.4.1] and [10.5], Knappe explains how to mix audio data from multiple conference endpoints. For example, Knappe supplements and explains how to mix audio data from multiple conference endpoints, including remote conference endpoints and PSTN conference endpoints.

See Knappe (Ex. 1006) at FIGs. 1-3, 5:49-62, 6:59-7:15. Specifically, in connection with Figure 3, Knappe teaches using a mixer 80 with a plurality of summers (e.g., summers 1 to 5), where each summer mixes audio data from the other participants (e.g., participants 1 to 5) while providing higher prominence to the moderator audio:

Referring to FIG. 3, the monaural mixer 80 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. In one embodiment, each participant is assigned a summer 82 that receives audio input streams from each other participant to the conference call. The summer 82 combines the audio input streams to generate a conference output stream for delivery to the participant.

During normal operation, each participant receives the audio input of each other participant. Thus, for example, the conference output stream of participant 1 includes the audio inputs of participants 2–5. Similarly, the conference output stream of participant 2 includes the audio inputs of participants 1 and 3–5. The conference output stream of participant 3 includes the audio inputs of participants 1–2 and 4–5. The conference output stream of participant 4 includes the audio inputs of participants 1–3 and 5. The conference output stream of participant 5 includes the audio inputs of participants 1-4.

The audio input 84 of the conference moderator may be amplified and/or the audio input 84 of the remaining participants attenuated to focus on or **provide higher prominence to the audio input 84 of the conference moderator.**

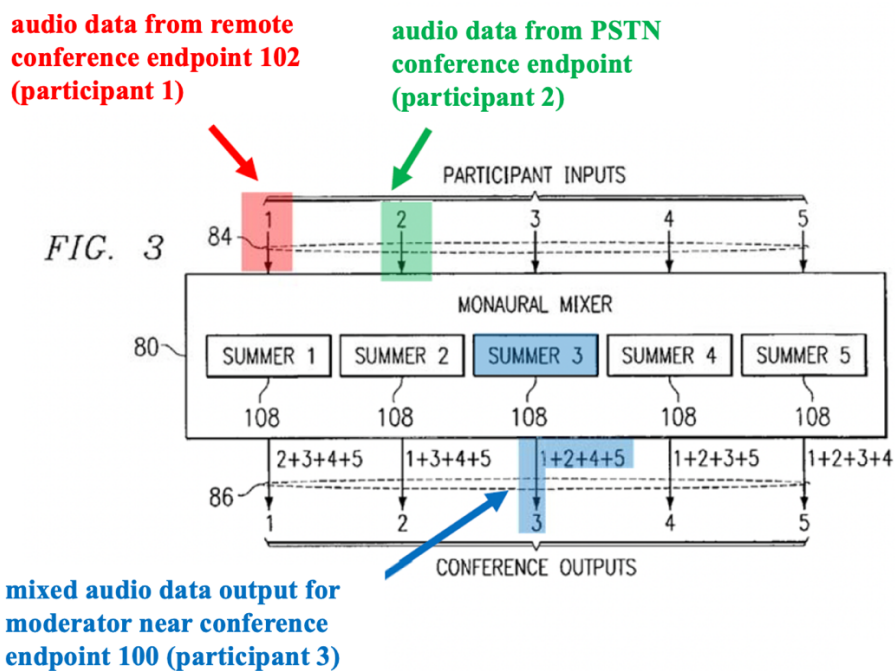
Id. at 6:59-7:15.

324. It would have been obvious to a POSITA to utilize in Drell's mixing module 308 of the MP conferencing application 234 a plurality of summers, as

Knappe teaches. Using a plurality of summers (one for each participant) would provide for separately mixing audio data from the remote conference endpoint 102 with audio data from the PSTN conference endpoint. Knappe's teachings would provide the benefit of reducing undesirable audio effects since each participant's mixed audio excludes its own audio data. *See also* Reasons to Combine Drell and Knappe, *supra* ¶¶ 137-157.

325. In the combination, Drell's MP conferencing application 234 of the moderator near conference endpoint 100 would utilize a plurality of summers as part of its mixing module 306 to obtain the functionality of separately mixing audio data from the participants, which is consistent with Drell's desire to generate a separate instance for each remote endpoint. *See* Drell (Ex. 1008) at 6:66-7:4.

326. For explanation purposes, and without limiting Knappe's disclosure, I have provided below an annotated example of Figure 3 identifying which of Drell's endpoints corresponds to which of participants 1-5. In the figure below, participant 1 corresponds to the remote conference endpoint 102 that provides audio data to input 1 of mixer 80, participant 2 corresponds to the PSTN conference endpoint that provides audio data to input 2 of mixer 80, and participant 3 corresponds to the moderator near conference endpoint 100 that receives mixed audio data from summer 3 comprising audio from participants 1+2+4+5:



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

327. Thus, Drell in combination with Knappe and Doganata teaches that the moderator near conference endpoint 100 includes a mixing module 308 (with separate summers including summer 3) configured to mix audio data from the remote conference endpoint 102 with audio data from the PSTN conference endpoint, which renders obvious “*mixing the second audio data received from the remote client with the first audio data received from the PSTN client,*” as claimed.

11. Claim 18

[18.P] *The method of claim 17, further comprising:*

328. See analysis for [17.P]-[17.1].

[18.1] *playing the mixed audio data from speakers connected to the local moderator client.*

329. The prior art renders obvious playing the mixed audio data from speakers connected to the local moderator client.

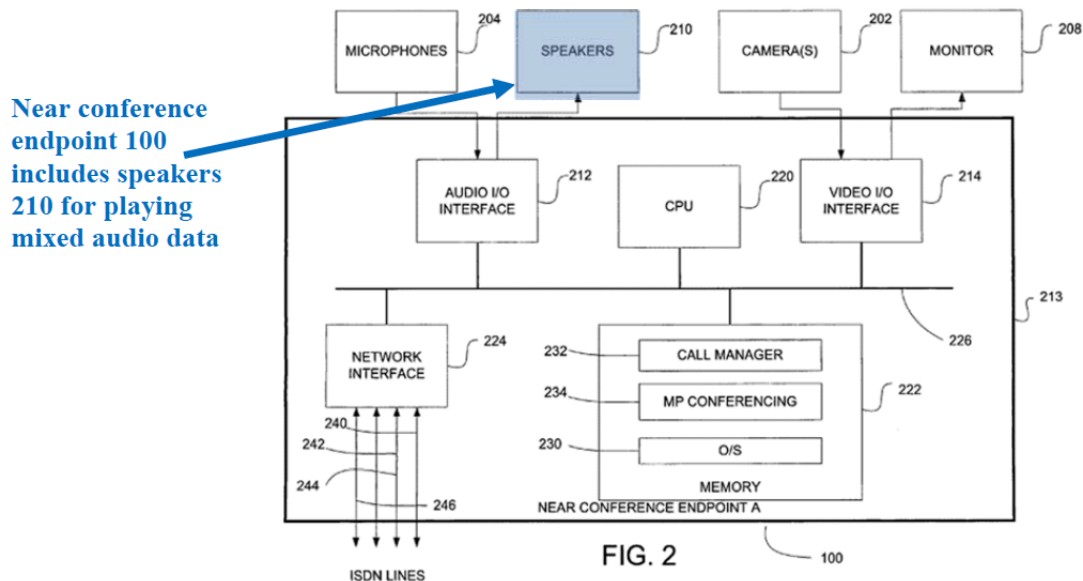
330. **First**, as discussed at [10.1.2], Drell in combination with Knappe and Doganata teaches a moderator near conference endpoint 100. Further, as discussed at [17.1], Drell in combination with Knappe teaches that the moderator near conference endpoint 100 generates mixed audio data.

331. **Second**, Drell teaches that speakers 210 present the mixed audio data (speech) to the moderator:

A video monitor 208 and **loudspeaker 210 present** images and **speech** of the remote conference participants.

Drell (Ex. 1008) at 3:45-46; *see also* Drell at FIGs. 2, 3. It would have been obvious to a POSITA that presenting audio from the speakers 210 includes “*playing*” the audio as claimed.

332. The summed speech of the remote conference participants corresponds to the “*mixed audio data*,” which as analyzed in [8.1] includes audio from the PSTN conference endpoint and the remote conference endpoint 102. As shown below at Drell’s Figure 2, the near conference endpoint 100 includes speakers 210 connected to the near conference endpoint 100 that play the mixed audio data to the moderator:



Id. at FIG. 2 (annotated).

333. Thus, Drell in combination with Knappe and Doganata teaches playing the mixed audio data from speakers 210 connected to the moderator near conference endpoint 100, which renders obvious “*playing the mixed audio data from speakers connected to the local moderator client,*” as claimed.

B. Claims 19-21 and 25-27 Are Obvious Over Knappe in Combination with Elliott, the VPN Textbook, and Hendricks

1. The Prior Art

a) Summary of Knappe

334. See Summary of Knappe, *supra* ¶¶ 121-131.

b) Summary of Elliott

335. See Summary of Elliott, *supra* ¶¶ 132-133.

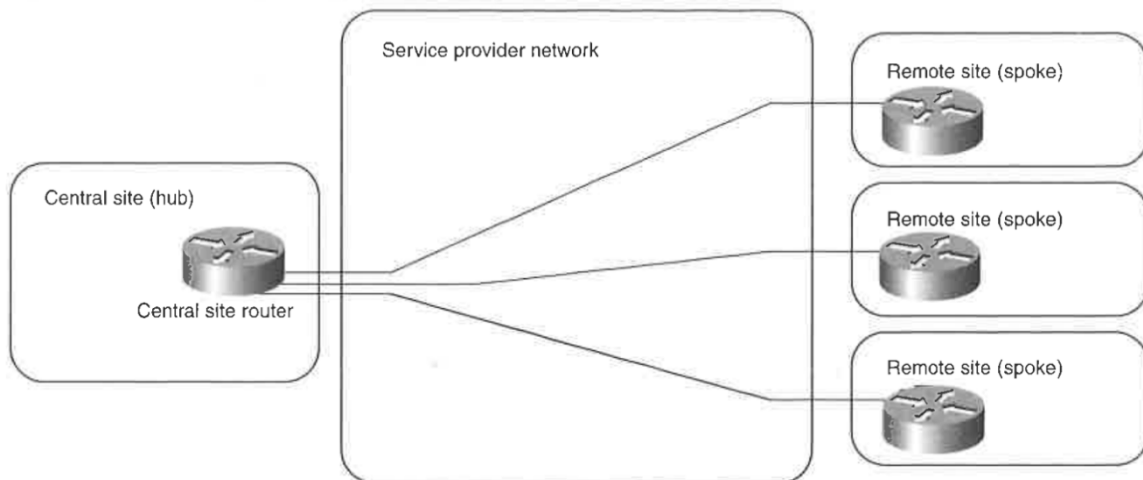
c) Summary of VPN Textbook

336. The textbook “MPLS and VPN Architectures, A practical guide to understanding, designing and deploying MPLS and MPLS-enabled VPNs,” (Ex. 1010, “VPN Textbook”) by Guichard et al., was published in 2001 by Cisco Press.

337. The VPN Textbook describes well-known topologies for implementing VPNs. Specifically, under the heading “Typical VPN Network Topologies,” 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. The “hub-and-spoke” topology is also referred to as a “star” topology. *See* Computer Desktop Encyclopedia (Ex. 1027) at 438.

338. The hub-and-spoke (or star) topology is illustrated in Figure 7-10 of the VPN Textbook, below:

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



VPN Textbook (Ex. 1010) at 130.

339. The VPN Textbook explains that “hub-and-spoke VPN topology is more common than you might expect.” *Id.* at 129. According to the VPN Textbook, even where an implementation may “benefit from a different topology” “the hub-and-spoke topology [is chosen] for cost or complexity reasons.” *Id.*

d) Summary of Hendricks

340. WO2001018665A1 to Hendricks was published on March 15, 2001. Hendricks is directed to video conference calls methods and systems. Hendricks (Ex. 1040) at Abstract, 2:23-27. Specifically, Hendricks discloses “multi-party call...utilizing only one or a combination of several types of media (e.g., voice only or voice plus video and/or other media such as data, programs, etc.)” that communicate using “full duplex.” *Id.* at 3:4-10.

341. Like Knappe, Hendricks uses a “conferencing central node” to support users “participating in a conference call,” including the “transmissions of video, audio and possibly other media.” *Id.* at 27:12-18. Hendricks, also like Knappe, discloses that the network can be a “[h]ybrid... [of] the PSTN, and other networks, such as the Internet.” *Id.* at 23:13-14, Fig. 7.

2. Reasons to Combine Prior Art

a) Reasons to Combine Knappe and Elliott

342. 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.

343. 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 ("A system and method for ... multi-party communication session"), 3:26-29 ("[T]he communication system 12...[facilitates] transmitting, audio, video, voice, data and other suitable types of real-time and/or non real-time traffic between source and destination points"); Elliott (Ex. 1007) at Abstract ("A method for communicating with a plurality of remote parties...for multi-media collaboration"). 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.

344. 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.

345. 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.

346. 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.

347. 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 '612 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. 1017). Before the earliest possible priority date of the '612 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.

348. 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," (Elliott (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.

349. 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 '612 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).

350. 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 '612 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).

351. 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.

352. 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. Additionally, the combination of Knappe and Elliott is merely the ordinary use of common techniques (transmitting collaboration web browsing data, application sharing data, and collaboration data per the T.120

standard as Elliott teaches) to improve the similar conferencing system of Knappe in the same way by allowing for enhanced collaboration among participants.

353. 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.

354. 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").

355. 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.").

356. 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.

357. 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”)).

358. 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.”).

359. 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.”).

360. 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.

361. 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 VPN Textbook

362. 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.

363. 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.

364. 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.

365. 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.

366. 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.

367. 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").

368. 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).

369. 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 '612 patent.

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

c) Reasons to Combine Knappe and Hendricks

371. A POSITA would have been motivated to combine the teachings of Knappe with Hendricks, for multiple reasons, including to produce the obvious, beneficial, and predictable results of providing full duplex audio to the conference participants.

372. A POSITA when considering the teachings of Knappe would have also considered the teachings of Hendricks as they are analogous prior art both pertaining to the same field of endeavor of multiparty conference calling. Knappe

(Ex. 1006) at Abstract (“A system and method for...multi-party communication session.”); Hendricks (Ex. 1040) at 3:4-10 (“...multi-party call...utilizing only one or a combination of several types of media (e.g., voice only or voice plus video and/or other media such as data, programs, etc.)”).

373. As already noted, Knappe communicates audio data over Internet network 14 and PSTN 22 to multiple conference participants (communications devices 16 and PSTN telephone 18). Knappe (Ex. 1006) at 3:40-51, FIG. 1. A POSITA would have known that audio data could be transmitted in two modes; namely, full duplex and half duplex. U.S. Patent No. 6,768,914 (Ex. 1041) at 1:20-21 (“Speakerphones may be designed to operate in either half-duplex or full-duplex.”); U.S. Patent No. 6,501,740 (Ex. 1042) at 7:17-20 (“[T]he mode can be specified to be full duplex (full participation, speaking and listening) or half duplex (listen only).”).

374. A POSITA, however, would have recognized that for conferencing applications full duplex is preferred. U.S. Patent No. 6,768,914 (Ex. 1041) at 1:28-35 (“[M]any speakerphones are designed to operate in full-duplex mode. The full-duplex mode allows incoming and outgoing parties voices to be simultaneously transmitted so that **there is no clipping or choppiness**. This results in much more **natural and spontaneous** flowing conversation. **Full-duplex is even more important** when more than one person is in a room using the speakerphone for a

conference call.”). Thus, it would have been obvious to a POSITA that full duplex audio would be beneficial to the communication system in Knappe that conducts multiple party conferencing.

375. Hendricks discloses “multi-party call...utilizing only one or a combination of several types of media (e.g., voice only or voice plus video and/or other media such as data, programs, etc.)” that communicate over PSTN and Internet networks using “**full duplex**.” Hendricks (Ex. 1040) at 3:4-10, 23:13-14, Fig. 7; *see also id.* at 17:19-22 (“The network [] provides **full duplex interconnectivity** between viewers [] involved in a given video conference call. The network [] can support many simultaneous calls.”). It would have been obvious for a POSITA to utilize full duplex audio (as Hendricks teaches) in Knappe’s conferencing system because doing so would beneficially (1) reduce clipping or choppiness and (2) facilitate natural and spontaneous flowing conversation among the conference participants. U.S. Patent No. 6,768,914 (Ex. 1041) at 1:28-33. As an additional third (3) reason, a POSITA would have utilized full duplex in Knappe’s system because doing so would improve audio when more than one person is in a room using the speakerphone for a conference call.

376. Accordingly, a POSITA would have been motivated to apply the full duplex voice communication teachings of Hendricks to Knappe’s communication system to obtain the beneficial results of (1) reducing clipping or choppiness, (2)

facilitating natural and spontaneous flowing conversation, and (3) improving audio when more than one person is in a room using the speakerphone for a conference call. Each of these benefits separately and together would have motivated a POSITA. Indeed, the use of a known full duplex technique, as taught by Hendricks, improves the similar conferencing system of Knappe in the same way; namely, it (1) reduces clipping or choppiness, (2) facilitates natural and spontaneous flowing conversation, and (3) improves audio when more than one person is in a room using the speakerphone for a conference call. I also note that it would have been obvious to try full duplex in Knappe's conferencing system since there are a finite number of identified predictable solutions (full duplex and half duplex) and there would have been a reasonable expectation of success.

377. That the result would have been predictable and that there would have been a reasonable expectation of success is evidenced by that fact that at the time “many speakerphones [were already] designed to operate in full-duplex mode.” U.S. Patent No. 6,768,914 (Ex. 1041) at 1:28-33. Further, other artisans had implemented full duplex systems. U.S. Patent No. 6,501,740 (Ex. 1042) at 7:14-20, 7:51-55. Moreover, any modification needed to the teachings of Knappe to accommodate the full duplex teachings of Hendricks, such as utilizing software programing and hardware, would have been within the level of ordinary skill in the art. U.S. Patent No. 6,424,647 (Ex. 1043) at 12:20-32 (“It can be appreciated that

the above hardware and software configuration can be practiced to provide tele-conference capabilities between multiple phone parties. The tele-conference can be with all parties interconnected through an Internet connection and an Internet phone and the above software and hardware configurations modified to provide for full duplex communications between all the parties...two parties can be connected over the Internet as described above, and multiple parties can be brought into the phone call connection through connections over the PSTN.”).

378. Lastly, I rely on the teachings of the references and do not suggest that physical incorporation of Hendricks’s elements into Knappe’s system is required in the proposed combination.

3. Claim 19

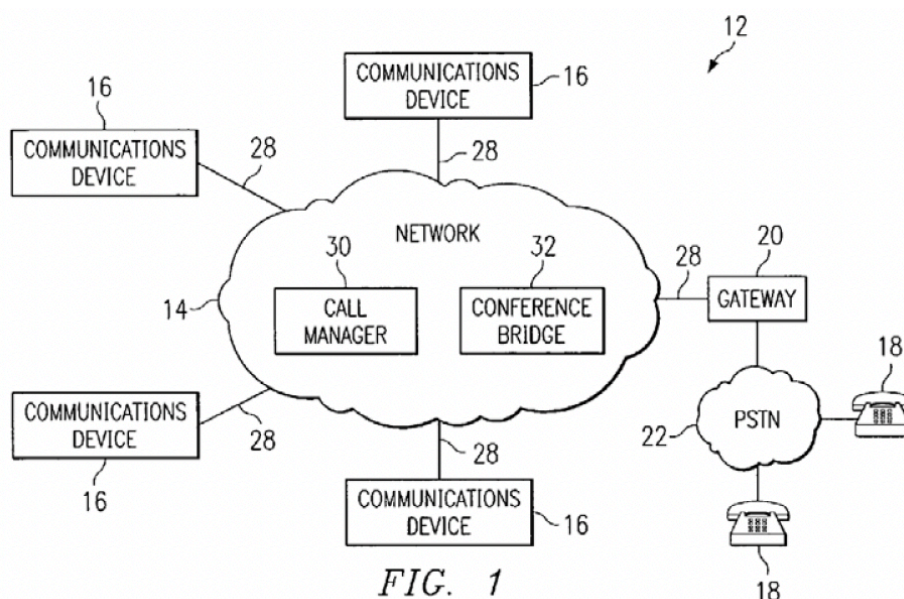
[19.P] A method of conducting a secure multi-participant conference call including full duplex audio, the method comprising:

379. The prior art renders obvious a method for conducting a secure multi-participant conference call including full duplex audio.

380. First, Knappe discloses “multiparty communications, and more particularly [] **a method and system** for improving the intelligibility of a moderator during a multiparty communication session...**Such communication networks also allow multiple people to participate in...a ‘conference call.’**”

Knappe (Ex. 1006) at 1:8-27; *see also id.* at 8:46-10:50, FIG. 8.

381. Knappe's system that performs the method for supporting multiparty communications is illustrated below in Figure 1:



Id. at FIG. 1.

382. In connection with above Figure 1, Knappe discloses that communication devices 16 may be implemented using a computing device, including a combination of telephone 18 and gateway 20:

The **communication devices 16** comprise IP or other digital telephones, personal and other suitable computers or computing devices, personal digital assistants (PDAs), cell or other mobile telephones or handset or any other device or set of devices such as the **telephone 18 and gateway 20 combination** capable of communicating real-time audio, video and/or other information over the network 14. The communication devices 16 also communicate control information with the network 14 to control call setup, teardown and processing as well as call services.

Id. at 3:62-4:4.

383. Further, Knappe discloses that its method and system for multiparty communications supports audio, video, voice, and other suitable data for conference sessions between communication devices 16:

FIG. 1 illustrates a communication system 12 in accordance with one embodiment of the present invention. In this embodiment, the communication system 12 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 endpoints. Communication system 12 may be used to conduct **multiple party telephone conference communication sessions**.

Id. at 3:24-31.

384. During operation, the conference bridge 32 facilitates the communication of audio, video, and other data between the conference devices:

The 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.

Id. at 4:58-63; *see also id.* at 5:42-48.

385. Further, Knappe discloses that the conference data is transmitted between the participants over Internet network 14.

The communication devices 16 comprise IP or other digital telephones, personal and other suitable computers or computing devices, personal digital assistants (PDAs), cell or other mobile telephones or handset or any other device or set of devices...**capable of communicating real-time audio, video and/or other information over the network 14**.

Id. at 3:62-4:1.

In accordance with a particular embodiment, **network 14 is the Internet....**In the Internet embodiment, the network 14 transmits information in Internet Protocol (IP) packets.

Id. at 3:50-54.

386. **Second**, as discussed in connection with [10.P], Elliott teaches transmitting audio and video conference data over Internetwork 18 using a VPN tunnel. *E.g.* Elliott (Ex. 1007) at 2:66-3:1, 4:6-12, 4:31-34; *see also* [10.P] *supra* ¶¶ 194-195. Thus, for the reasons stated in [10.P] *supra* ¶ 195, a POSITA would have understood that Elliott utilizes VPN tunnelling for communicating conferencing voice, video and collaboration data over a public network 18, such as the Internet. It would have been obvious to use in Knappe VPN tunneling as taught in Elliott to secure communications over Knappe's Internet network 14. *See supra* ¶¶ 194-195, 353-361.

387. **Third**, Knappe teaches that the conference participants communicate with each other through conference bridge 32, which provides a mixed audio for each participant that excludes its own audio:

Referring to FIG. 3, the monaural mixer 80 [of conference bridge 32] 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. In one embodiment, each participant is assigned a summer 82 that receives audio input streams from each other participant to the conference call. The summer 82 combines the audio input streams to generate a conference output stream for delivery to the participant.

During normal operation, each participant receives the audio input of each other participant. Thus, for example, the conference output stream of participant 1 includes the audio inputs of participants 2–5. Similarly, the conference output stream of participant 2 includes the audio inputs of participants 1 and 3–5. The conference output stream of participant 3 includes the audio inputs of participants 1–2 and 4–5. The conference output stream of participant 4 includes the audio inputs of participants 1–3 and 5. The conference output stream of participant 5 includes the audio inputs of participants 1–4.

Knappe (Ex. 1006) at 6:59-7:11

388. In the above portions, Knappe explains that the conference bridge 32 mixes the audio from the other participants while excluding the audio of the participant that receives the mixed audio. For example, the mixed output for participant 1 excludes its own audio and includes only the audio of participants 2–5. A POSITA would have understood that Knappe’s conference bridge 32 is excluding audio from participant 1 because it is not desirable for that participant to hear his or her voice while speaking. That is, participant 1 is capable of communicating audio to the conference bridge 32 (i.e., input 1) and also receive mixed audio (i.e., output 2+3+4+5). A person of ordinary skill in the art would recognize this simultaneous transmission and reception as a conference call that includes full-duplex audio.

389. Nevertheless, to the extent determined that Knappe does not disclose full-duplex audio, for the reasons stated in [10.P] (*supra* ¶ 197), Hendricks discloses that a multi-participant conference call over Internet/PSTN includes full

duplex audio and video and renders obvious that the conference call includes “*full duplex*” audio. *See supra* ¶ 197.

390. It would have been obvious for a POSITA to utilize full duplex audio in Knappe’s system (as Hendricks teaches) because doing so would provide the participants of Knappe’s communication devices 16 and telephones 18 a more natural way to communicate with each other during the conference. *See* U.S. Patent No. 6,768,914 (Ex. 1041) at 1:20-39; U.S. Patent No. 6,501,740 (Ex. 1042) at 7:14-34; Reasons to Combine Knappe and Hendricks, *supra* ¶¶ 371-378.

391. Thus, Knappe in combination with Elliott and Hendricks teaches a method and a system for supporting multiple participants in a secure (VPN tunneling) conference call including full duplex audio, which renders obvious “[a] *method for conducting a secure multi-participant conference call including full duplex audio,*” as recited in the preamble.

[19.1.1] receiving first audio data, first video data, and first collaboration data at a server from a first remote client

392. The prior art renders obvious receiving first audio data, first video data, and first collaboration data at a server from a first remote client.

393. **First**, Knappe discloses that the conference call manager 30 and conference bridge 32 provide services to conference participants, including real-time multiparty communication sessions between communication devices:

The call manager 30 is responsive to service requests from the communication devices 16, including the standard telephones 18 through the gateway 20. For example, the call manager 30 may provide voicemail, bridging, multicasting, call hold, conference call and other multiparty communications and/or other suitable services for the communications devices 16. The **call manager 30 provides services** by performing the services, controlling performance of the services, delegating performance of the services and/or by otherwise initiating the services.

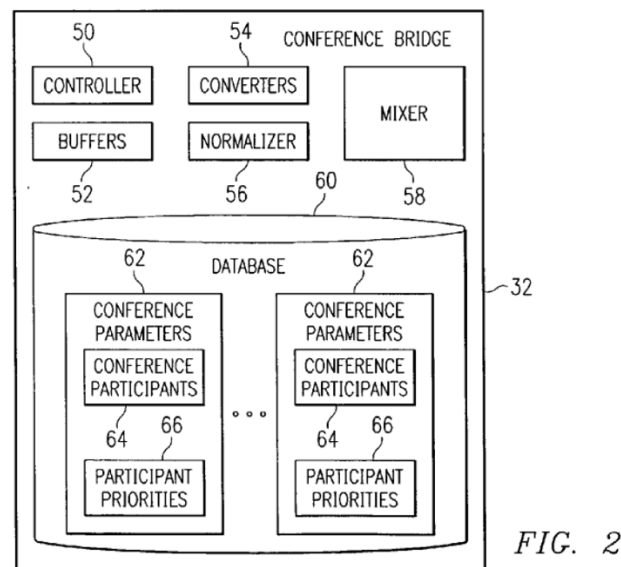
The **conference bridge 32 provides** conference call and other suitable audio, video, and/or real-time multiparty communication sessions between communication devices 16.

Knappe (Ex. 1006) at 4:48-61.

During the multiparty communications sessions, participants are connected and **stream media through the conference bridge 32**.

Id. at 5:31-33.

394. The conference bridge 32 is illustrated in Figure 2, replicated below:



Id. at FIG. 2.

395. Knappe's call manager 30, either alone or in combination with the conference bridge 32, corresponds to the claimed "*server*" because it (and they collectively) provide a conference call service to communication devices 16 and PSTN telephones 18.

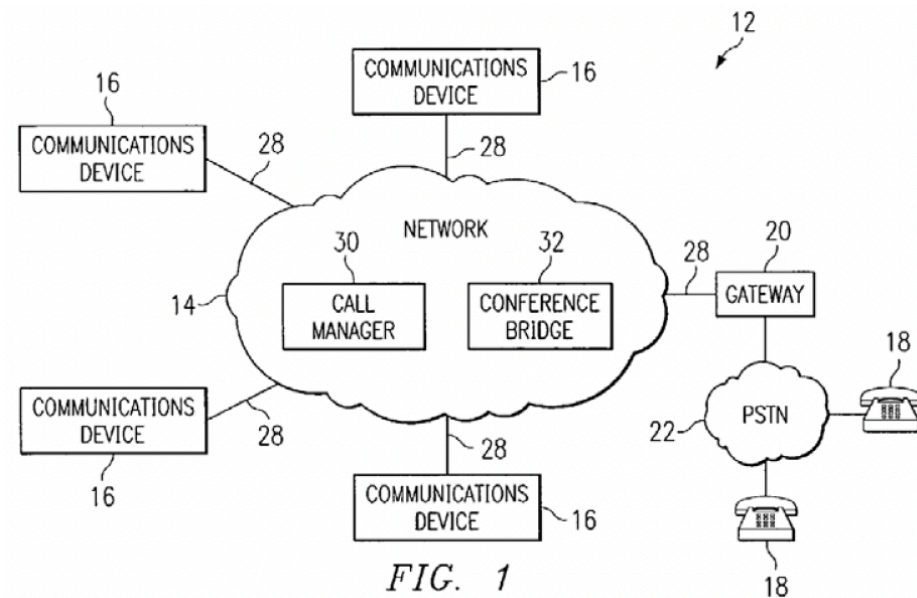
396. **Second**, Knappe discloses that its conference allows for people who are in different physical locations to participate:

Additionally, packet switched networks, such as the Internet, are able to perform a similar function by transporting packets containing data that represents audible sounds from one location to another...Modern business practices often require that several persons meet on the telephone to engage in a conference call. The conference call has introduced certain applications and techniques that are superior to those found in a meeting with persons physically present in the same location. For example, a conference call attendee who is not participating at the moment may wish to mute their audio output and simply listen to the other conferences. This allows the particular conferencee to work on another project while still participating in the conference.

Id. at 1:18-41. The remote locations of Knappe's participants is further confirmed by its reference to "minimizing travel expenses," as the conferencing system allows participants to communicate over Internet network 14 from wherever they are instead of requiring them to travel to a single location. *See id.* at 1:42- 44 (noting that conferencing is "helpful in minimizing travel expenses and other costs associated with business over long distances"), 3:50-51 ("In accordance with a particular embodiment, network 14 is the Internet."). A POSITA would have understood that Knappe contemplates that the users of the conference system are

not physically present in the same location; instead, they are remote from one another and communicate over the Internet and PSTN.

397. As shown below at Figure 1, Knappe's system includes communication devices 16 that are connected via links 28 to network 14:



Id. at FIG. 1.

398. As already noted, Knappe's network 14 may be the Internet. *Id.* at 3:50-51 ("In accordance with a particular embodiment, network 14 is the Internet."). It would have been obvious to a POSITA that the conference participant at any one of communications device 16 is remote with respect to the other devices.

399. **Third**, Knappe teaches that the remote communications device 16 transmits audio, video, data, and other information through a network 14 and to the conference bridge 32:

During the multiparty communications sessions, participants are connected and **stream media** through the conference bridge 32.

Id. at 5:31-33.

The **communication devices 16** comprise IP or other digital telephones, personal and other suitable computers or computing devices, personal digital assistants (PDAs), cell or other mobile telephones or handset or any other device or set of devices...**capable of communicating real-time audio, video and/or other information over the network 14.**

Id. at 3:62-4:1.

FIG. 1 illustrates a communication system 12 in accordance with one embodiment of the present invention. In this embodiment, the communication system 12 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 endpoints. Communication system 12 may be used to conduct multiple party telephone conference communication sessions.

Id. at 3:24-31.

The 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. In particular, the conference bridge 32 receives media from participating devices 16 and, using suitable signal processing techniques, mixes the media to produce conference signals. During normal operation, each device 16 receives a conference signal that includes contributions from all other participating devices.

Id. at 4:58-5:1. A POSITA would have understood Knappe's description of "stream[ing] media" to refer at least to its discussion of "communicating real-time **audio, video, and/or other information**" and "**data**."

400. Knappe further discloses that conference bridge 32 receives the transmitted conference data. *Id.* at 4:63-66 ("[C]onference bridge 32 **receives** media from participating devices 16."), 6:59-7:11 ("Referring to FIG. 3, the monaural mixer 80 [of the conference bridge 32] **receives** participant input streams."). In short, conference bridge 32 receives audio, video, and other data from the communications devices 16 for collaboration among the participants. *See, e.g., id.* at 1:29-41. Further, Knappe teaches that "[d]uring the multiparty communications sessions, participants are connected and stream media through the conference bridge 32." *Id.* at 5:31-33.

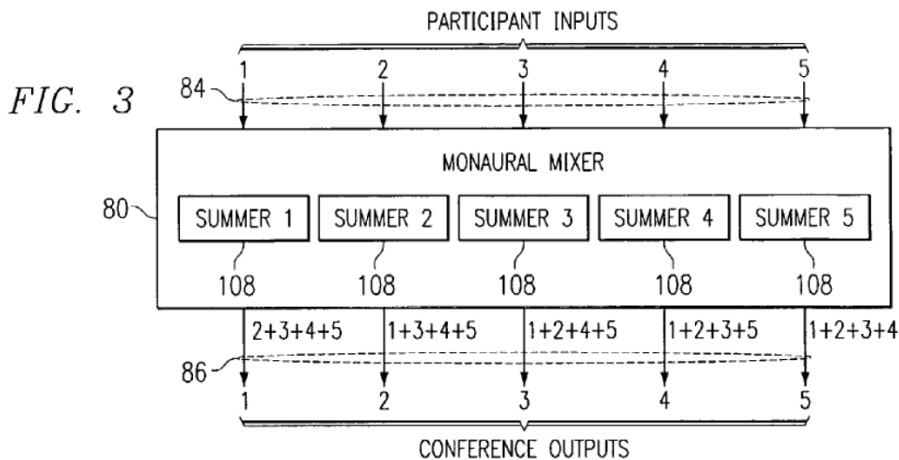
401. Furthermore, with regard to Figure 3, Knappe further explains that mixer 80 of conference bridge 32 receives input audio data from a plurality of participants (i.e., participants 1-5) and mixes the audio data using summers and outputs a mixed audio data stream for each participant:

Referring to FIG. 3, the monaural mixer 80 [of the conference bridge 32] **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. In one embodiment, each participant is assigned a summer 82 that receives audio input streams from each other participant to the conference call. The

summer 82 combines the audio input streams to generate a conference output stream for delivery to the participant.

During normal operation, each participant receives the audio input of each other participant. Thus, for example, the conference output stream of participant 1 includes the audio inputs of participants 2–5. Similarly, the conference output stream of participant 2 includes the audio inputs of participants 1 and 3–5. The conference output stream of participant 3 includes the audio inputs of participants 1–2 and 4–5. The conference output stream of participant 4 includes the audio inputs of participants 1–3 and 5. The conference output stream of participant 5 includes the audio inputs of participants 1-4.

Id. at 6:59-7:11; *see also id.* at 4:63-66. Because a monaural mixer is used to mix multiple audio inputs to a single audio output for each participant, a POSITA would have understood that the inputs to the monaural mixer include audio data.

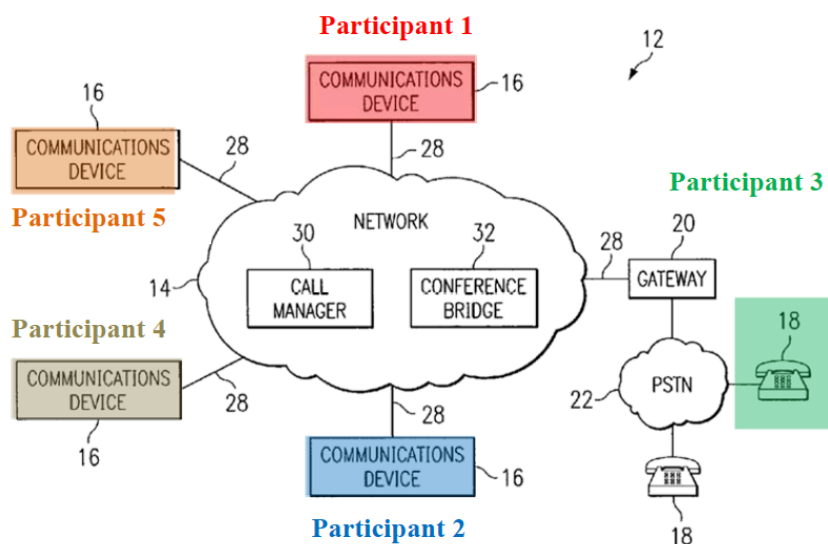


Id. at FIG. 3.

402. I note that Knappe does not identify which specific communications device 16 or PSTN telephone 18/gateway 20 (in Figure 1) corresponds with which specific participant 1-5 (in Figure 3). A POSITA reading Knappe, however, would have understood that this assignment of participant number is arbitrary and that

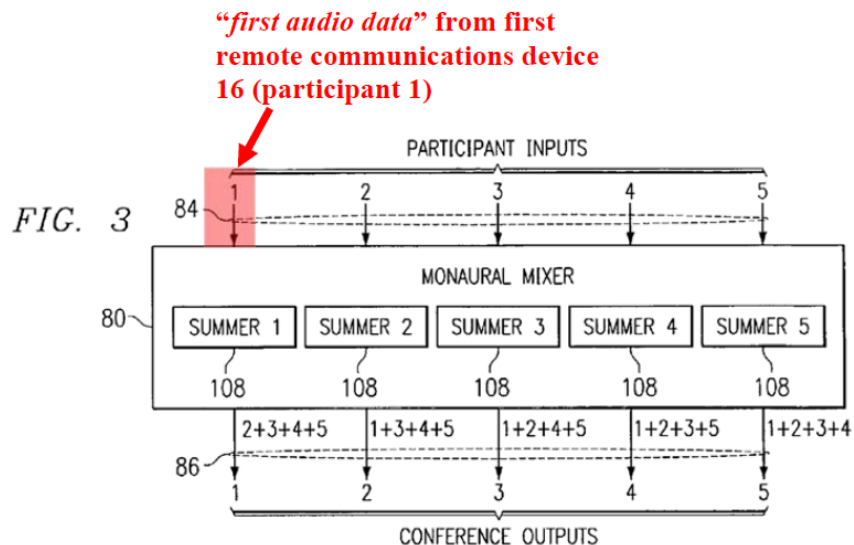
such detail is not necessary to understand the operation in the Figure 3 example.

For explanation purposes, and without limiting Knappe's disclosure, I have provided the below participant numbers corresponding to devices in Figure 1:



Id. at FIG. 1 (annotated). The communications device 16 of participant 1 corresponds to the claimed “*first remote client.*”

403. Further for explanation purposes, and without limiting Knappe's disclosure, I have provided an annotated example of Figure 3 identifying which device from Figure 1 corresponds to which of participants 1-5. In the figure below, participant 1 corresponds to the first remote communications device 16 that provides audio data to input 1 of mixer 80 of conference bridge 32:



Id. at FIG. 3 (annotated).

404. Thus, a POSITA would have understood that Knappe’s “audio, video, and/or other information” and “data” corresponds to “media” transmitted in real-time and received by the conference bridge 32 that is used for collaboration among the participants and therefore teaches “*receiving first audio data, first video data, and first collaboration data,*” as claimed.

405. Additionally, a POSITA would have understood that Knappe contemplated that the real-time “data” transmitted during conferencing (*id.* at 3:27-29, 3:62-4:1, 5:45-48) include well-known collaboration data, such as T.120 standard data.

406. **Fourth**, as discussed in [10.4.1] (*see supra* ¶ 254), Elliott teaches that it was well known that the data in the transmitted stream includes collaboration

data, such as “application sharing” data, “web browsing” data, and data per the “T.120 standard.” *See supra* ¶ 254.

407. Elliott’s teaching of collaboration, such as application sharing, web browsing, and other collaboration data per the T.120 standard, discloses that the stream includes “*collaboration data*,” as claimed. Indeed, the ’612 patent similarly uses the prior art T.120 standard to disclose collaboration data. *See, e.g.*, ’612 patent (Ex. 1001) at 3:19-21 (“[T]he H.323 recommendation defines a **data collaboration** capability as known and outlined in the **T.120 data** collaboration unit.”).

408. Knappe teaches that its communication devices 16 can “comprise...personal and other suitable computers or computing devices.” Knappe (Ex. 1006) at 3:62-64. A POSITA would have understood that when Knappe’s communication devices 16 are implemented as personal computers, Elliott’s teachings of web browsing and application sharing would be directly applicable. Accordingly, it would have been obvious to a POSITA that Knappe’s “other information” would include collaboration data as Elliott teaches (e.g., application sharing data, web browsing data, and other collaboration data per the T.120 standard). Including Elliott’s collaboration data in the additional data exchanged by Knappe’s communication devices would allow different ways for the participants to collaborate thereby enhancing the effectiveness of the

collaboration among the participants. *See* Reasons to Combine Knappe and Elliott, *supra* ¶¶ 342-361.

409. As noted above, Knappe teaches that “[d]uring the multiparty communications sessions, participants are connected and stream media through the conference bridge 32,” where the media includes “audio, video, voice, data and other suitable types of real-time...traffic between source and destination endpoints.” Knappe (Ex. 1006) at 3:26-29, 5:31-33. Accordingly, it would have been obvious to a POSITA that in the combination, conference bridge 32 would receive the audio, video, and data (such as collaboration data per Elliott) from each participant and distribute the mixed data (including audio, video, and collaboration data) to each participant.

410. Thus, Knappe in combination with Elliott teaches that that call manager 30 and conference bridge 32 provide conference services, including receiving audio, video, and data (e.g., collaboration data such as web browsing data, application sharing data, and other collaboration data) from a first remote communications device 16, which discloses “*receiving first audio data, first video data, and first collaboration data at a server from a first remote client,*” as claimed.

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

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

412. **First**, as discussed at [19.1.1], Knappe in combination with Elliott teaches receiving audio, video, and collaboration data from a first remote communications device 16.

413. **Second**, as discussed in [10.3.2.] (*supra* ¶¶ 234-235), Elliott teaches transmitting audio, video, and collaboration data over Internet network 18 using a VPN tunnel. *E.g.* Elliott (Ex. 1007) at 2:66-3:1, 6:53-54, 4:6-12, 4:31-34; *supra* ¶¶ 234-235.

414. Based on the disclosures of Elliott discussed in [10.3.2], *supra* ¶¶ 234-235, a POSITA would have understood that Elliott utilizes VPN tunneling for communicating conferencing 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 utilize VPN tunneling when communicating audio, video, and collaboration data between the first remote communication devices 16 and the call manager 30/conference bridge 32, because it would provide Knappe with secure communications as the communications traverse the Internet network 14. *See also* Reasons to Combine Knappe and Elliott, *supra* ¶¶ 342-361.

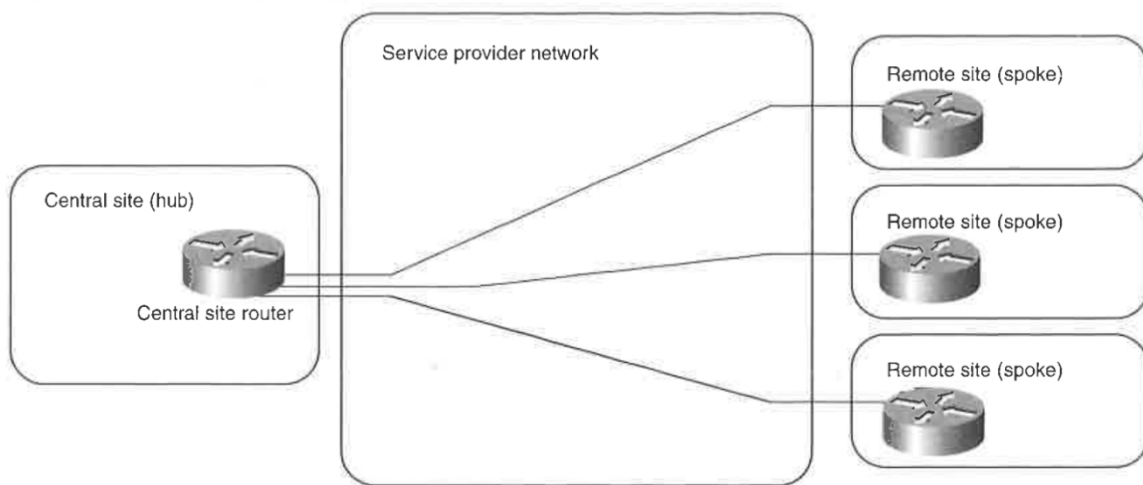
415. Further, the VPN Textbook teaches that the most common VPN topology is "hub-and-spoke":

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), similar to the setup in Figure 7-10.

VPN Textbook (Ex. 1010) at 129.

416. As shown in Figure 7-10, each remote site router (spoke) is connected to a central router (hub) through a VPN tunnel:

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



Id. at 130.

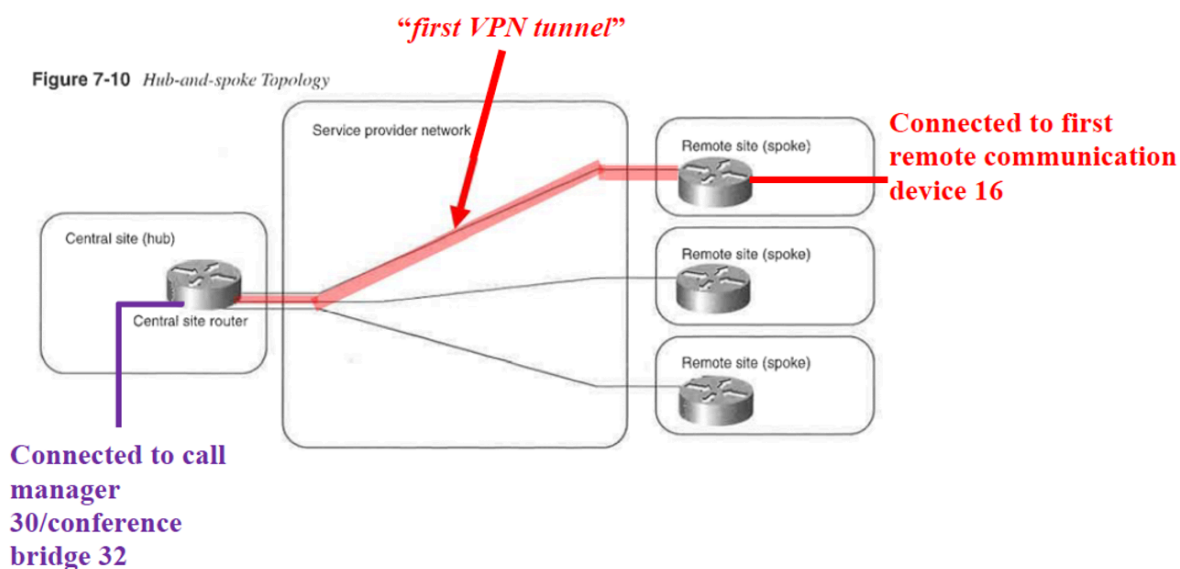
417. A POSITA would have understood that each remote site router (spoke) would have one or more network devices connected to it.

418. As explained above, Knappe's devices may be at different remote locations on the network. *See, e.g.*, discussion at [19.1.1]; 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 are 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* ¶¶ 362-370.

419. 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 4:30-31 (“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 first remote communications device 16 would be connected to a remote site (spoke) router, while the call manager 30/conference bridge 32 would be connected to the central site (hub) router, as illustrated below in annotated Figure 7-10:



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

420. As illustrated above, the remote site router (spoke) of the first remote communications device 16 and the central router (hub) of the call manager 30/conference bridge 32 are connected through a VPN tunnel. When the conference bridge 32 receives data from or transmits data to the first remote communications device 16, the data would be received/sent through the same VPN tunnel, which corresponds to *"a first VPN tunnel."*

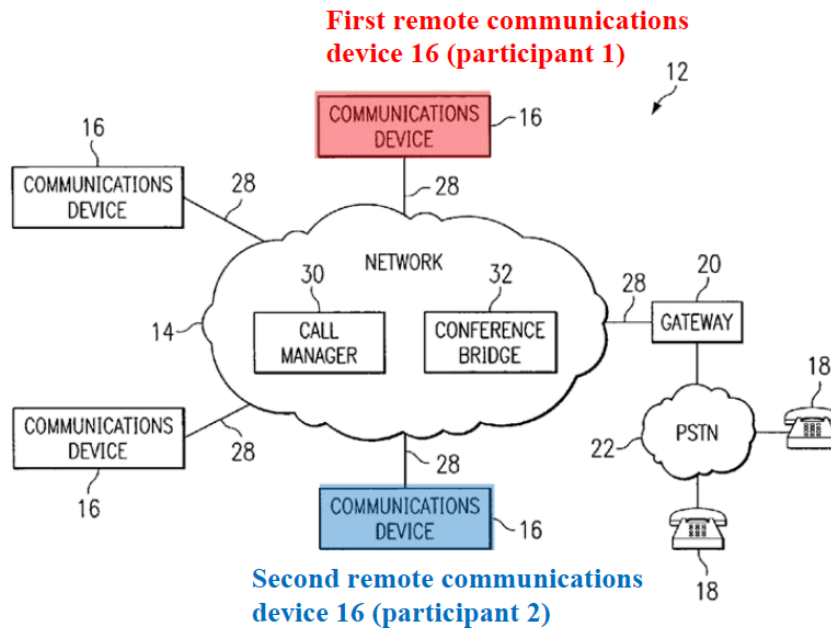
421. Thus, Knappe in combination with Elliott and the VPN Textbook teaches that audio, video, and collaboration data from the first remote communications device 16 is received through a VPN tunnel, which renders obvious receiving data from the remote client “*through a first Virtual Private Network (VPN) tunnel,*” as claimed.

[19.2.1] receiving second audio data at the server from a second remote client

422. The prior art renders obvious receiving second audio data at the server from a second remote client.

423. **First**, as discussed at [19.1.1], conference bridge 32, either alone or together with conference call manager 30, corresponds to the claimed “*server.*” Further, as discussed at [19.1.1], Knappe teaches that each of communication devices 16 is remote with respect to each other.

424. For explanation purposes, without limiting Knappe’s disclosure, I will proceed with the example where the participant 2 participates in the conference by communicating real-time audio, video and/or other information over the network 14 using a second remote communication device 16 (“*second remote client*”), as illustrated in annotated Figure 1, below:



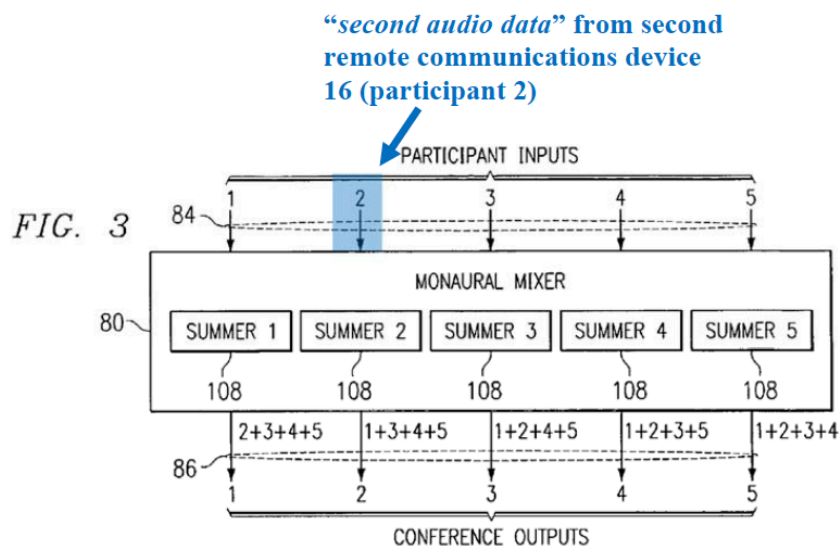
Knappe (Ex. 1006) at FIG. 1.

425. **Second**, it would have been obvious to a POSITA that the devices illustrated in Figure 1 communicate audio/video and other data over links 28 and network 14 to the conference bridge 32. *See. e.g., id.* at 3:45-49 (“The communication devices 16, standard analog telephones 18 and gateway 20 are connected to the network 14 and/or PSTN 22 through twisted pair, cable, fiber optic, radio frequency, infrared, microwave and/or any other suitable type of wireline or wireless links 28.”), 4:58-66 (“The conference bridge 32 provides conference call and other suitable **audio, video, and/or real-time multiparty communication** sessions between communication devices... In particular, the conference bridge 32 **receives media** from participating devices 16 and, using suitable signal processing techniques, mixes the media to produce conference

signals.”), 5:31-33 (“During the multiparty communications sessions, participants are connected and stream media through the conference bridge 32.”). Since audio and video are forms of media, it would have been obvious to a POSITA that the conference bridge 32 receives audio data from the second remote communications device used by participant 2.

426. **Third**, Knappe teaches that mixer 80 (which is a part of conference bridge 32, as discussed at [19.1.1]) receives the audio data from the participating communication devices 16. *E.g. id.* at 4:63-66, 6:59-7:11; *see supra* ¶¶ 400-402.

427. As shown below in Figure 3, Knappe’s mixer 80 receives input audio from participant 2, which corresponds to the second remote communications device 16 (“*second remote client*”):



Id. at FIG. 3 (annotated).

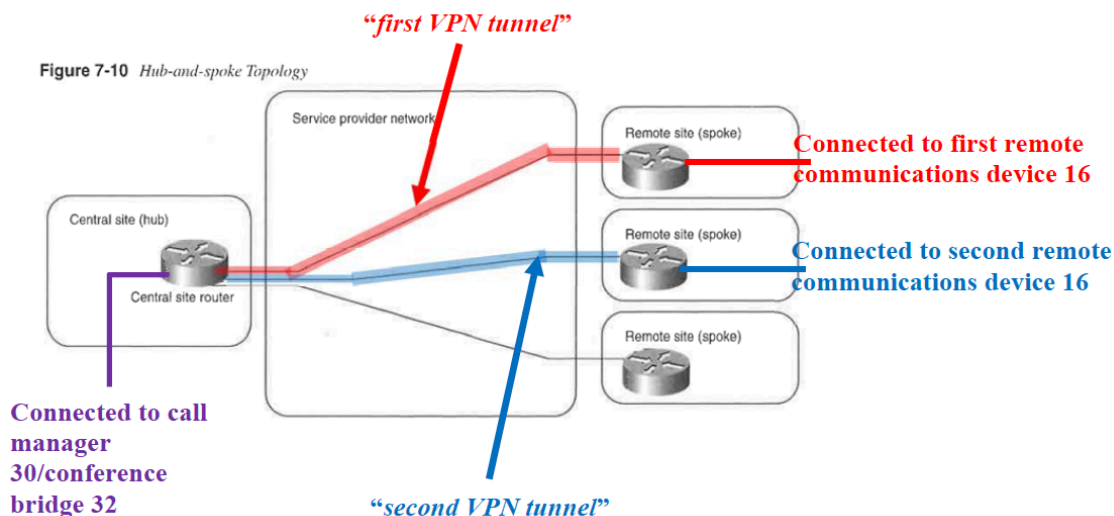
428. Thus, Knappe teaches that call manager 30 and conference bridge 32 provide conference services, including receiving audio data from a second remote communications device 16, which renders obvious “*receiving second audio data at the server from a second remote client,*” as claimed.

[19.2.2] through a second VPN tunnel;

429. The prior art renders obvious that the second audio data is received through a second VPN tunnel.

430. As discussed at [19.1.2], Knappe in combination with Elliott and the VPN Textbook teaches using a conventional hub-and-spoke VPN topology, where each remote site router (hub) is connected through a VPN tunnel to the central router (hub) to transmit and receive data.

431. It would have been obvious to a POSITA that when implementing the commonly used hub-and-spoke VPN topology for conferencing (per Elliott and the VPN Textbook), Knappe’s call manager 30/conference bridge 32 would be connected to the central site (hub) router that utilizes a first VPN tunnel for communications with a remote site (spoke) connected to the first communications device 16 and utilizes a second VPN tunnel for communications with a remote site (spoke) connected to the second communications device 16, as illustrated below:



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

432. As illustrated above, the remote site router (spoke) of the second remote communications device 16 and the central router (hub) of the call manager 30/conference bridge 32 are connected through a second VPN tunnel. When the conference bridge 32 receives data from or transmits data to the second remote communications device 16, the data would be sent through the same VPN tunnel, which corresponds to “a second VPN tunnel.”

433. It would have been obvious to a POSITA to utilize the conventional hub-and-spoke VPN topology, when setting up a second VPN tunnel between the second remote communication device 16 and call manager 30/conference bridge 32, because it would secure communications among the connected devices in a cost effective and relatively simple way. *See also* Reasons to Combine Knappe

and Elliott, *supra* ¶¶ 342-361; Reasons to Combine Knappe and the VPN Textbook, *supra* ¶¶ 362-370.

434. Thus, Knappe in combination with Elliott and the VPN Textbook teach that the second audio data from the second remote communications device 16 is received through a second VPN tunnel, which renders obvious that the second audio data is received “*through a second VPN tunnel*,” as claimed.

[19.3] receiving third audio data at the server from a Public Switched Telephone Network (PSTN) client through a PSTN gateway;

435. The prior art renders obvious receiving third audio data at the server from a Public Switched Telephone Network (PSTN) client through a PSTN gateway.

436. **First**, as discussed at [19.1.1], conference bridge 32, either alone or together with conference call manager 30, corresponds to the claimed “*server*.”

437. **Second**, Knappe teaches a public switched telephone network (PSTN) telephone 18 that generates audio data which is sent through gateway 20 to the conference bridge 32, which receives it as streaming media:

During the multiparty communications sessions, participants are connected and **stream media through the conference bridge 32**.

Knappe (Ex. 1006) at 5:31-33.

Referring to FIG. 1, communication system 12 includes a network 14 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 (PSTN) 22.

Id. at 3:41-45.

If the **originating device is a PSTN telephone 18**, the media streaming occurs between the gateway 20 and the destination device.

Id. at 5:23-26.

The gateway 20 provides conversion between analog and/or digital formats. The standard analog **telephones 18 communicate standard telephony signals through PSTN 22 to the gateway 20**. At the gateway 20, the signals are converted to IP packets in the VoIP format.

Id. at 4:18-22.

The communication devices 16 comprise IP or other digital telephones, personal and other suitable computers or computing devices, personal digital assistants (PDAs), cell or other mobile telephones or handset or any other device or set of devices such as the **telephone 18 and gateway 20 combination capable of communicating real-time audio, video and/or other information over the network 14**. The communication devices 16 also communicate control information with the network 14 to control call setup, teardown and processing as well as call services.

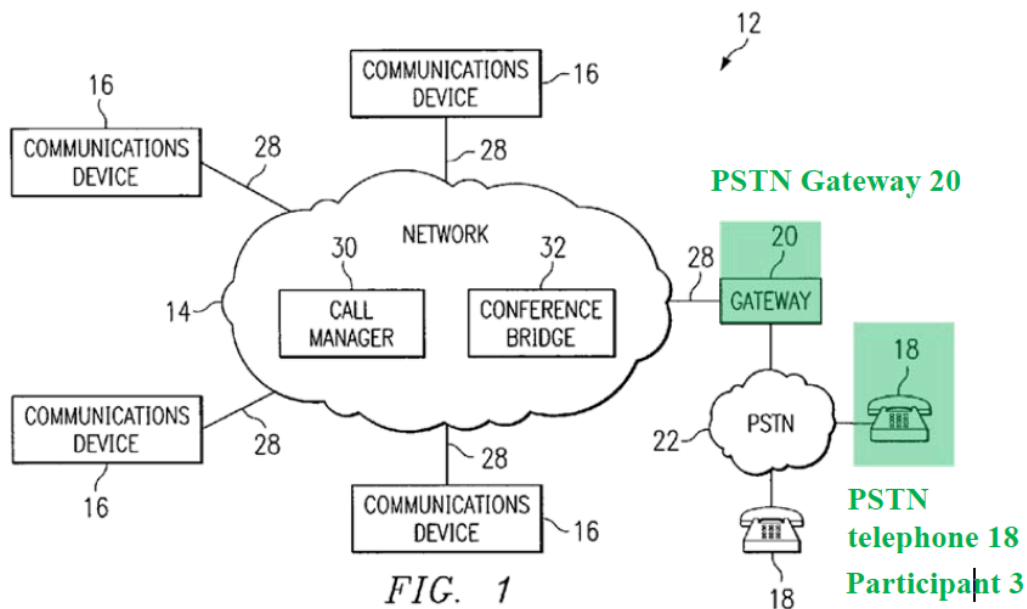
Id. at 3:62-4:4.

[C]onference bridge 32 receives media from participating devices 16 and, using suitable signal processing techniques, mixes the media to produce conference signals.

Id. at 4:63-66.

438. In view of the above disclosure, a POSITA would have understood that when the PSTN telephone 18 is the originating device, it sends data over gateway 20 which communicates the data to the conference bridge 32, which facilitates the multiparty communications.

439. As shown below in Figure 1, Knappe's system includes PSTN telephones 18 that communicate audio data through gateway 20:

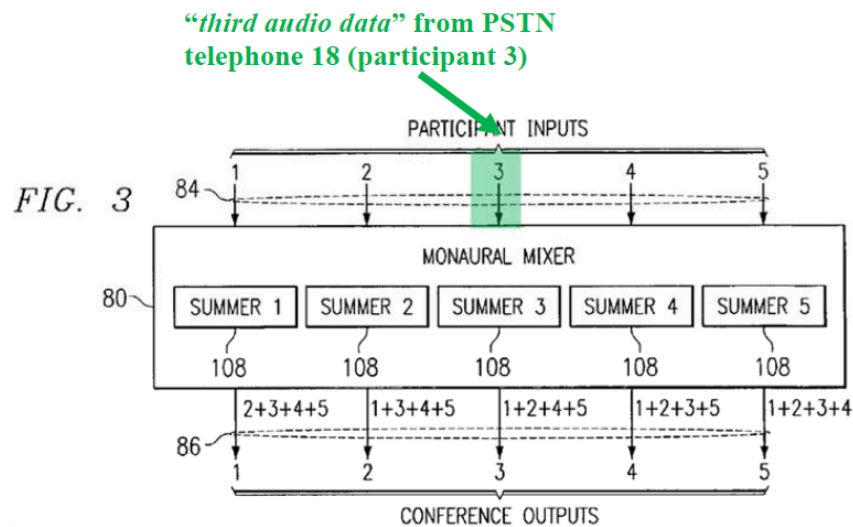


Id. at FIG. 1 (annotated).

440. It would have been obvious to a POSITA that the devices illustrated in Figure 1, e.g., telephone 18 and gateway 20, communicate audio data over link 28 over network 14 to the conference bridge 32. *Id.* at 6:59-7:11. I note that this is substantially the same arrangement described in the '612 patent. *See, e.g.,* '612 patent (Ex. 1001) at 6:37-7:11, FIGS. 4, 5 (PSTN client 413, PSTN Gateway 411).

441. **Third**, as discussed at [19.1.1], Knappe discloses that mixer 80 of conference bridge 32 receives input data from participants 1-5. For explanation purposes, and without limiting Knappe's disclosure, I have provided an annotated example of Figure 3 identifying which device from Figure 1 corresponds to which

of participants 1-5. In the figure below, participant 3 corresponds to the PSTN telephone 18 that provides audio data to input 3 of mixer 80 via gateway 20:



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

442. Thus, Knappe discloses that call manager 30 and conference bridge 32 provide conference services, including receiving audio data from a PSTN telephone 18 through gateway 20, which renders obvious “*receiving third audio data at the server from a Public Switched Telephone Network (PSTN) client through a PSTN gateway,*” as claimed.

[19.4] mixing the first audio data with the second audio data into a first mixed audio data;

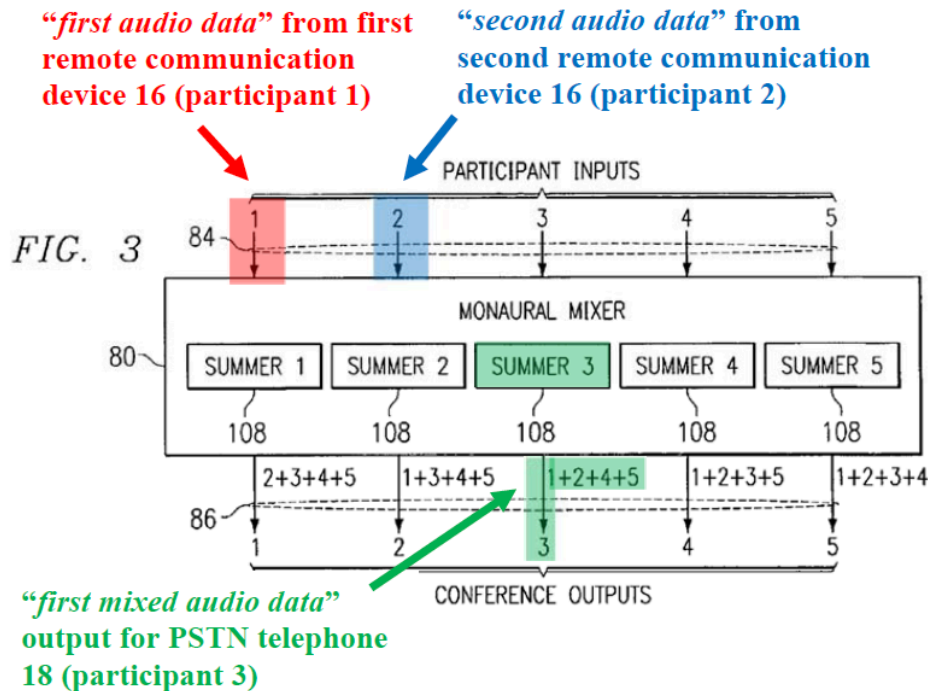
443. The prior art renders obvious mixing the first audio data with the second audio data into a first mixed audio data.

444. **First**, as discussed at [19.1.1], conference bridge 32, either alone or together with conference call manager 30, corresponds to the claimed “*server.*”

Further, as discussed at [19.1.1] and [19.2.1], Knappe teaches that the conference bridge 32 receives the first audio data from the first remote communication device 16 (participant 1) and also receives the second audio data from the second remote communication device 16 (participant 2).

445. **Second**, Knappe teaches that mixer 80 of the conference bridge 32 mixes the audio data from participant 1 and participant 2 into a mixed output stream for participant 3. *See e.g.* Knappe (Ex. 1006) at 6:59-7:11 (“output stream of participant 3 includes the audio inputs of participants 1-2 and 4-5); *see also e.g. supra* ¶¶ 243, 267, 323, 387, 401.

446. As discussed at [19.1.1], [19.2.1], and [19.3], participant 1 corresponds to the first remote communications device 16, participant 2 corresponds to the second remote communication device 16, and participant 3 corresponds to the PSTN telephone 18. As shown below in Figure 3, Knappe’s mixer 80 includes summer 3 that mixes the streams from the first remote communications device 16 (participant 1) and second remote communications device 16 (participant 2) and outputs a mixed data stream (“*a first mixed audio data*”) comprising signals from participants 1+2+4+5 for PSTN telephone 18 (participant 3):



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

447. Thus, Knappe discloses that mixer 80, including summer 3, mixes audio data from the first remote communication device 16 with audio data from the second communications device 16 into a first mixed audio data (1+2+4+5), which renders obvious *“mixing the first audio data with the second audio data into a first mixed audio data,”* as claimed.

[19.5] mixing the second audio data with the third audio data into a second mixed audio data

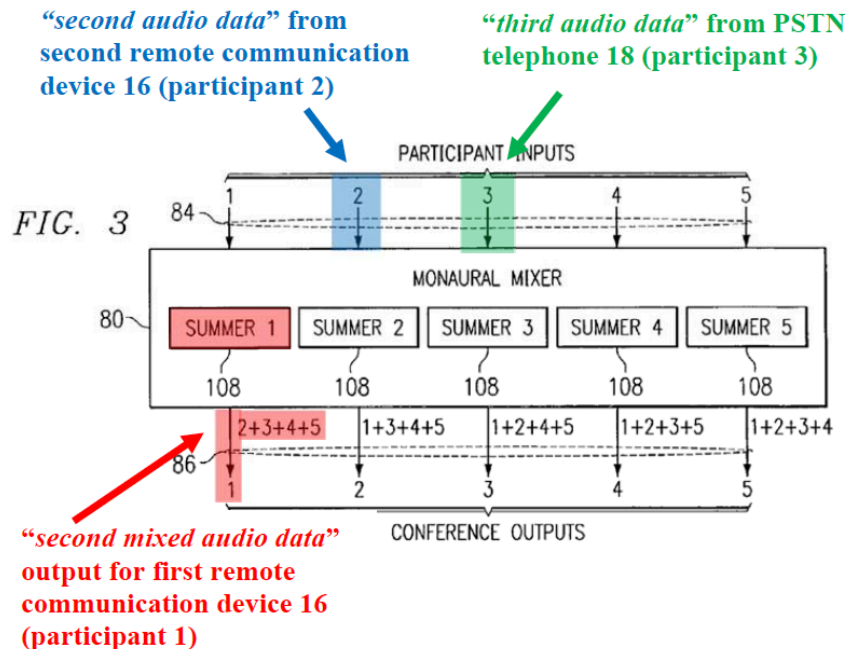
448. The prior art renders obvious mixing the second audio data with the third audio data into a second mixed audio data.

449. **First**, as discussed at [19.1.1], conference bridge 32, either alone or together with conference call manager 30, corresponds to the claimed *“server.”*

Further, as discussed at [19.2.1] and [19.3], Knappe teaches that the conference bridge 32 receives the second audio data from the second remote communication device 16 (participant 2) and also receives the third audio data from the PSTN telephone 18 (participant 3).

450. **Second**, as previously noted, Knappe teaches that mixer 80 of the conference bridge 32 mixes the audio data from participant 2 and participant 3 into a mixed output stream for participant 1. *Id.* at 6:59-7:11 (“output stream of participant 1 includes the audio inputs of participants 2-5”); *see supra* ¶¶ 243, 267, 323, 387, 401.

451. As discussed at [19.2.1], [19.3], and [19.1.1], participant 2 corresponds to the second communications device 16, participant 3 corresponds to the PSTN telephone 18, and participant 1 corresponds to the first remote communication device 16. As shown below in Figure 3, Knappe’s mixer 80 includes summer 1 that mixes the streams from the second remote communications device 16 (participant 2) and the PSTN telephone 18 (participant 3) and outputs mixed data stream (“*a second mixed audio data*”) comprising signals from participants 2+3+4+5 for participant 1:



Id. at FIG. 3 (annotated).

452. Thus, Knappe discloses that mixer 80, including summer 1, mixes audio data from the second remote communication device 16 with audio data from the PSTN telephone 18 into a second mixed audio data (2+3+4+5), which renders obvious “*mixing the second audio data with the third audio data into a second mixed audio data,*” as claimed.

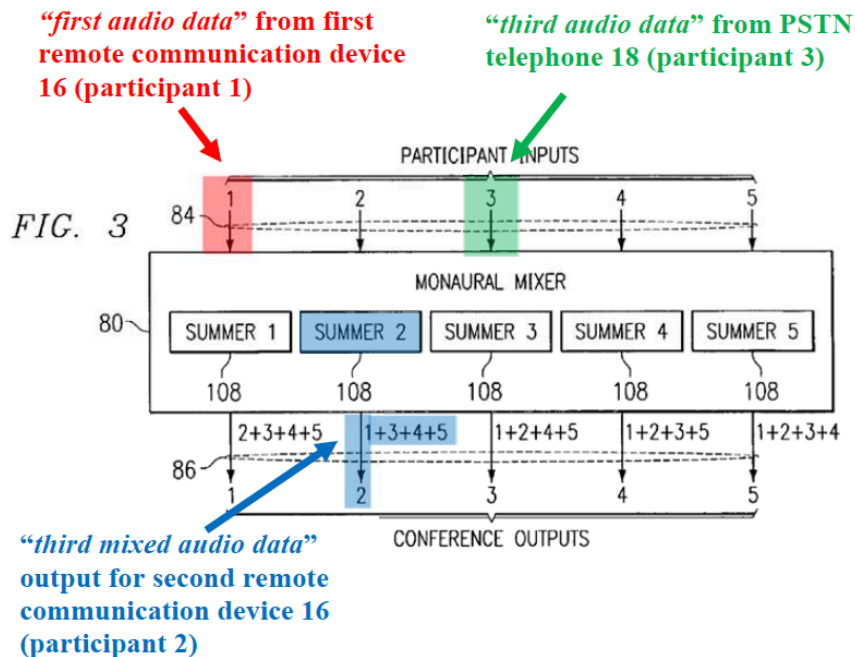
[19.6] mixing the first audio data with the third audio data into a third mixed audio data;

453. The prior art renders obvious mixing the first audio data with the third audio data into a third mixed audio data.

454. **Second**, Knappe teaches that mixer 80 of the conference bridge 32 mixes the audio data from participant 1 and participant 3 into a mixed output

stream for participant 2. *Id.* at 6:59-7:11 (“output stream of participant 2 includes the audio inputs of participants 1 and 3-5”); *see supra* ¶¶ 243, 267, 323, 387, 401.

455. As discussed at [19.1.1], [19.3], and [19.2.1], participant 1 corresponds to the first remote communications device 16, participant 2 corresponds to the second remote communication device 16, and participant 3 corresponds to the PSTN telephone 18. As shown below in Figure 3, Knappe’s mixer 80 includes summer 2 that mixes the streams from the first remote communications device 16 (participant 1) and the PSTN telephone 18 (participant 3) and outputs a mixed data stream (“a third mixed audio data”) comprising signals from participants 1+3+4+5 for participant 2:



Id. at FIG. 3 (annotated).

456. Thus, Knappe discloses that mixer 80, including summer 2, mixes audio data from the first remote communication device 16 with audio data from the PSTN telephone 18 into a third mixed audio data (1+3+4+5), which renders obvious “*mixing the first audio data with the third audio data into a third mixed audio data,*” as claimed.

[19.7] transmitting the first mixed audio data from the server to the PSTN client through the PSTN gateway;

457. The prior art renders obvious transmitting the first mixed audio data from the server to the PSTN client through the PSTN gateway.

458. **First**, as discussed at [19.1.1], conference bridge 32, either alone or together with conference call manager 30, corresponds to the claimed “*server.*” Further, as discussed at [19.3], Knappe teaches that the conference bridge 32 communicates to the PSTN telephone 18 through gateway 20. Further, as discussed at [19.4], Knappe teaches generating “*first mixed audio data*” as an output intended for participant 3 using the PSTN telephone.

459. **Second**, Knappe discloses that the conference bridge 32 transmits the mixed audio data to each participant:

FIG. 1 illustrates a communication system 12 in accordance with one embodiment of the present invention. In this embodiment, the communication system 12 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 endpoints.

Knappe (Ex. 1006) at 3:24-29

During the multiparty communications sessions, **participants are connected and stream media through the conference bridge 32.** The media is cross connected and mixed to produce conference output streams for each participant. The conference output stream for a participant includes the media of all other participants, a subset of other participants or other suitable mix dictated by the type of multiparty session and/or the participant.

Id. at 5:31-39.

During normal operation, each participant receives the audio input of each other participant. Thus, for example, the conference output stream of participant 1 includes the audio inputs of participants 2–5. Similarly, the conference output stream of participant 2 includes the audio inputs of participants 1 and 3–5. **The conference output stream of participant 3 includes the audio inputs of participants 1–2 and 4–5.**

Id. at 7:1-7

The call manager 30 manages calls in the network 14. A call is any **communication session between two or more parties.** The parties may be persons and/or equipment such as computers. The sessions may include real-time connections, connections having real-time characteristics, non real-time connections and/or a combination of connection types.

The call manager 30 is responsive to service requests from the communication devices 16, including the standard telephones 18 through the gateway 20. For example, the call manager 30 may provide voicemail, bridging, multicasting, call hold, conference call and other multiparty communications and/or other suitable services for the communication devices 16. The call manager 30 provides services by performing the services, controlling performance of the services, delegating performance of the services and/or by otherwise initiating the services.

The 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. In particular, the conference bridge 32 receives media from participating devices 16 and, using suitable signal processing techniques, mixes the media to produce conference signals. **During normal operation, each device 16 receives a conference signal that includes contributions from all other participating devices.**

Id. at 4:42-5:1; *see also id.* at 10:33-35 (“At step 212, the moderator and participant voice streams are transmitted.”), FIG. 8.

460. Further, Knappe teaches that the transmitted audio data is received at the gateway 20 and converted to standard telephony signals for delivery to the destination PSTN telephone 18, which receives the audio:

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.

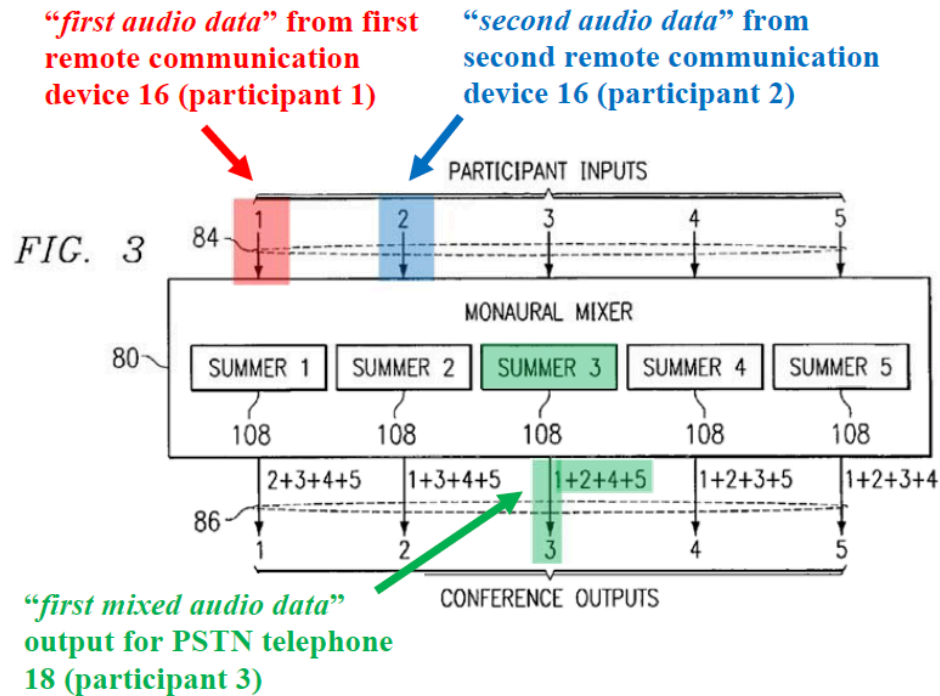
Id. at 4:21-25.

During normal operation, each participant receives the audio input of each other participant. Thus, for example...the conference output stream of participant 2 includes the audio inputs of participants 1 and 3–5.

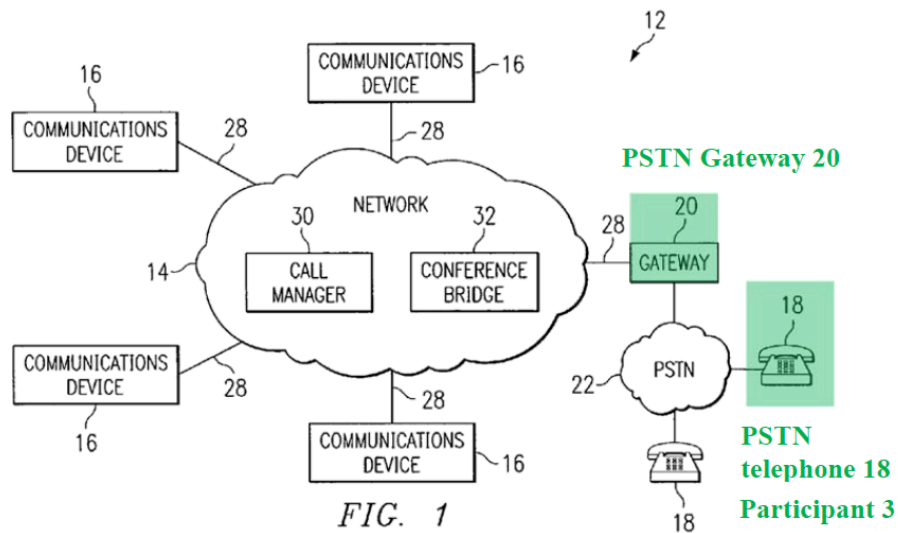
Id. at 7:1-7; *see also id.* at FIG. 1

461. As discussed at [19.3], participant 3 corresponds to the PSTN telephone 18 that receives communication through gateway 20. It would have been obvious to a POSITA, therefore, that the conference bridge 32 transmits the mixed audio output of summer 3 to the communication device of participant 3, which I previously identified in [19.3] as the PSTN telephone 18 which receives

audio data via gateway 20. As shown below, Knappe's mixer outputs mixed audio data (1+2+4+5) ("the first mixed audio data") to participant 3 (PSTN telephone 18):



Id. at FIG. 3 (annotated). As shown below, the PSTN telephone 18 receives communications from conference bridge 32 via PSTN gateway 20:



Id. at FIG. 1 (annotated).

462. Thus, Knappe discloses transmitting the first mixed audio data from the conference bridge 32 to the PSTN telephone 18 through gateway 20, which renders obvious “*transmitting the first mixed audio data from the server to the PSTN client through the PSTN gateway,*” as claimed.

[19.8] transmitting the second mixed audio data from the server to the first remote client through the first VPN tunnel;

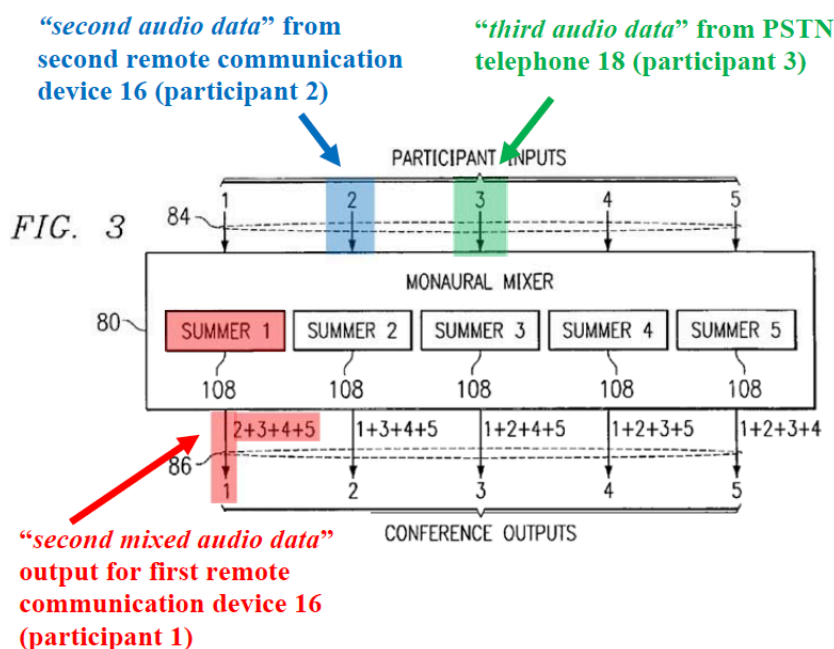
463. The prior art renders obvious transmitting the second mixed audio data from the server to the first remote client through the first VPN tunnel.

464. **First**, as discussed at [19.1.1], conference bridge 32, either alone or together with conference call manager 30, corresponds to the claimed “*server.*” Further, as discussed at [19.1.1], Knappe teaches a first remote communications

device 16. Further still, as discussed at [19.5], Knappe teaches generating second mixed audio data.

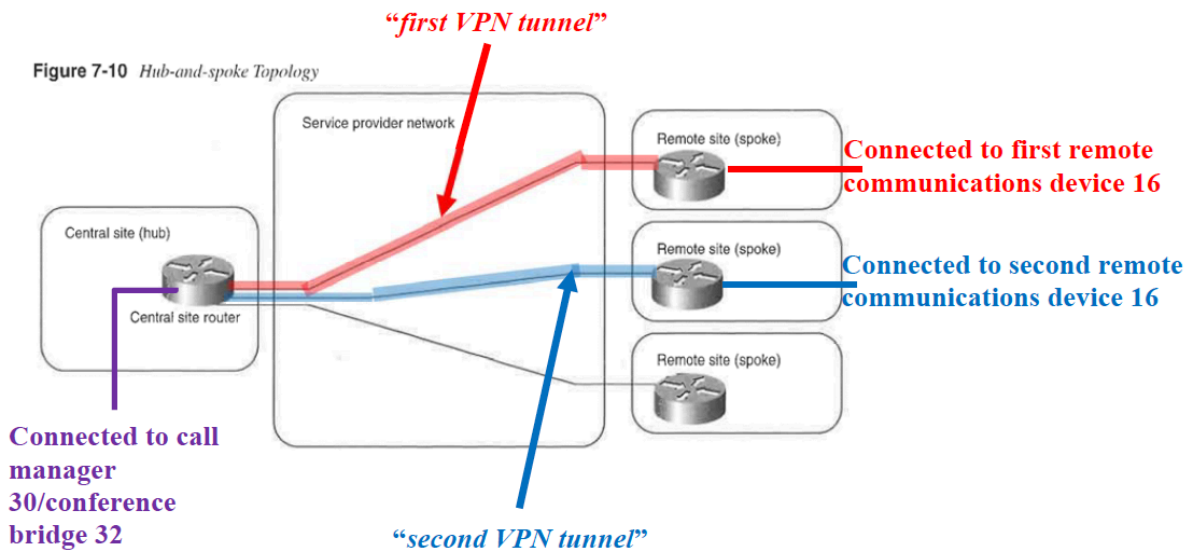
465. **Second**, for the same reasons discussed at [19.7], Knappe discloses transmitting the second mixed audio data to the first remote communications device 16, which receives the audio data. *See* Knappe (Ex. 1006) at 3:24-29, 5:31-39, 7:1-7, 4:42-51 (“each device 16 receives a conference signal that includes contributions from all other participating devices”); *see also id.* at 10:33-35, FIG. 8; *see also supra* ¶ 459.

466. As discussed at [19.3], participant 1 corresponds to the first remote communications device 16. As shown below, Knappe teaches outputting mixed audio data (2+3+4+5) (“the second mixed audio data”) to participant 1 (first remote communications device 16):



Id. at FIG. 3 (annotated).

467. **Third**, as discussed at [19.1.2], Knappe in combination with Elliott and the VPN Textbook teaches using a first VPN tunnel between the central site (hub) router of the conference bridge 32 and the remote site (spoke) router of the first remote communications device 16, as illustrated below:



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

468. With reference to the above figure, communications from the conference bridge 32 traverse the central site (hub) router, the first VPN tunnel, and the remote site (spoke) router to reach the first remote communications device 16. Accordingly, when the conference bridge 32 transmits the second mixed audio data to the first remote communications device 16, the data would be sent through the first VPN tunnel that was already established (see discussion at [19.1.2]), which corresponds to “*the first VPN tunnel.*”

469. It would have been obvious to a POSITA to utilize the VPN tunnel set up between the first remote communications device 16 and conference bridge 32 when the conference bridge 32 transmits the mixed audio data to the first remote communications device 16, because it would provide Knappe with secure communications and also because it would be a cost effective and relatively simple implementation. *See also* Reasons to Combine Knappe and Elliott, *supra* ¶¶ 342-361; Reasons to Combine Knappe and the VPN Textbook, *supra* ¶¶ 362-370.

470. Thus, Knappe in combination with Elliott and the VPN Textbook teaches that the conference bridge 32 transmits the second mixed audio data to the first remote communications device 16 through the first VPN tunnel, which renders obvious “*transmitting the second mixed audio data from the server to the first remote client through the first VPN tunnel,*” as claimed.

[19.9] transmitting the third mixed audio data from the server to the second remote client through the second VPN tunnel;

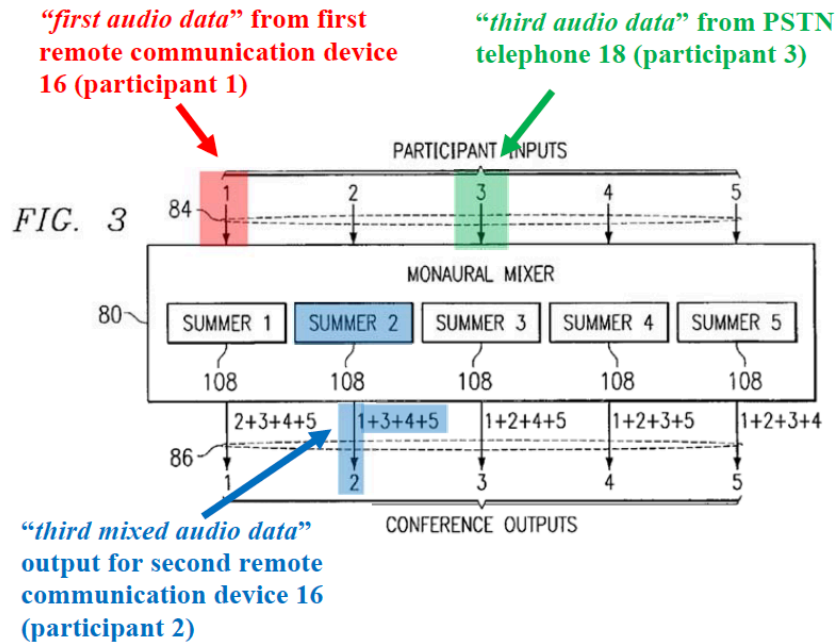
471. The prior art renders obvious transmitting the third mixed audio data from the server to the second remote client through the second VPN tunnel.

472. **First**, as discussed at [19.1.1], conference bridge 32, either alone or together with conference call manager 30, corresponds to the claimed “*server.*” Further, as discussed at [19.2.1], Knappe teaches a second remote communication device 16. Further still, as discussed at [19.6], Knappe teaches generating third

mixed audio data as an output intended for participant 2 using the second remote communications device 16.

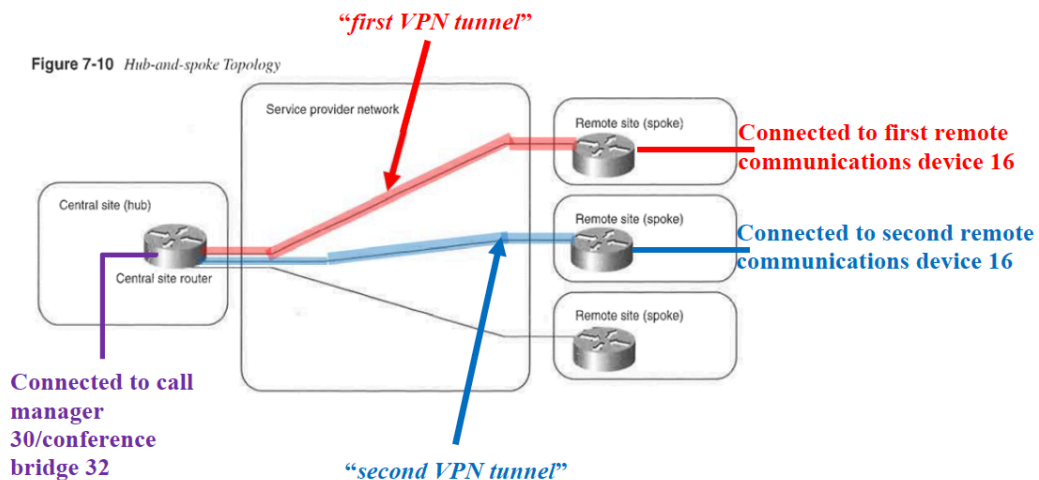
473. **Second**, for the reasons discussed in [19.7] and [19.8], Knappe discloses transmitting the third mixed audio data to the second remote communications device 16, which receives the audio data. *See* Knappe (Ex. 1006) at 3:24-29, 5:31-39, 7:1-7, 4:42-51 (“each device 16 receives a conference signal that includes contributions from all other participating devices”); *see also id.* at 10:33-35, FIG. 8; *see also supra* ¶¶ 459, 465.

474. As discussed at [19.2.1], participant 2 corresponds to the second remote communications device 16. As discussed at [19.6], Knappe teaches generating the third mixed audio data comprising audio data from the first communication device 16 (participant 1) and the PSTN telephone 18 (participant 3). As shown below, Knappe teaches outputting mixed audio data (1+3+4+5) (“*the third mixed audio data*”) to participant 2 (second remote communications device 16):



Id. at FIG. 3 (annotated).

475. **Third**, as discussed at [19.2.2], Knappe in combination with Elliott and the VPN Textbook teaches using a second VPN tunnel between the central site (hub) router of the conference bridge 32 and the remote site (spoke) router of the second remote communications device 16, as illustrated below:



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

476. With reference to the above figure, communications from the conference bridge 32 traverse the central site (hub) router, the second VPN tunnel, and the remote site (spoke) router to reach the second remote communications device 16. Accordingly, when the conference bridge 32 transmits the third mixed audio data to the second remote communications device 16, the data would be sent through the second VPN tunnel that was already established (see discussion at [19.2.2]), which corresponds to “*the second VPN tunnel*.”

477. It would have been obvious to a POSITA to utilize the VPN tunnel set up between the second remote communications device 16 and conference bridge 32 when the conference bridge 32 transmits the mixed audio data to the second remote communications device 16, because it would provide Knappe with secure communications and also because it would be a cost effective and relatively simple implementation. *See also* Reasons to Combine Knappe and Elliott, *supra* ¶¶ 342-361; Reasons to Combine Knappe and the VPN Textbook, *supra* ¶¶ 362-370.

478. Thus, Knappe in combination with Elliott and the VPN Textbook teaches that the conference bridge 32 transmits the third mixed audio data to the second remote communications device 16 through the second VPN tunnel, which renders obvious “*transmitting the third mixed audio data from the server to the second remote client through the second VPN tunnel*,” as claimed.

[19.10] transmitting the first video data and the first collaboration data from the sever [sic] to the second remote client.

479. The prior art renders obvious transmitting the first video data and the first collaboration data from the server to the second remote client.

480. **First**, as discussed at [19.1.1], conference bridge 32, either alone or together with conference call manager 30, corresponds to the claimed “*server*.” Further, as discussed at [19.1.1], Knappe in combination with Elliott teaches that the conference bridge 32 receives the first video data and the first collaboration data from the first remote communications device 16.

481. **Second**, Knappe teaches that conference bridge 32 transmits video data and collaboration data to the communications devices 16. Knappe (Ex. 1006) at 3:24-29 (“the communication system 12 is a distributed system **transmitting audio, video, voice, data** and other suitable types of real-time...traffic”), 4:42-5:1. FIG. 3; 4:9-17, 5:45-57; *see also supra* ¶¶ 459, 465.

482. It would have been obvious to a POSITA that Knappe’s “video, voice, data and other suitable types of real-time...traffic” corresponds to the “media” received by the conference bridge 32 and, after mixing, suitable signals are transmitted by the conference bridge 32 to the other conference participants. It would have further been obvious to a POSITA that the transmitted data include video and collaboration data because the “output stream for a participant includes the media of all other participants.” Knappe (Ex. 1006) at 5:35-39.

483. **Third**, as discussed at [10.4.1], [12.1] and [19.1.1], Elliott teaches transmitting collaboration data, such as “application sharing” data, “web browsing” data, and data per the “T.120 standard.” Elliott (Ex. 1007) at 1:40-44, 6:46-54, 5:49-61; *see also supra* ¶¶ 254, 289, 406.

484. As already noted, Knappe teaches that its communication devices 16 can “comprise...personal and other suitable computers or computing devices.” Knappe (Ex. 1006) at 3:62-64. A POSITA would have understood that when Knappe’s communication devices 16 are implemented as personal computers, Elliott’s teachings of web browsing and application sharing would be directly applicable. Accordingly, a POSITA would understand that Knappe’s “other information” would include collaboration data as Elliott teaches (e.g., application sharing data, web browsing data, and other collaboration data per the T.120 standard). Including Elliott’s collaboration data in the additional data exchange by Knappe’s communication devices would allow different ways for the participants to collaborate thereby enhancing the effectiveness of the collaboration among the participants. *See Reasons to Combine Knappe and Elliott, supra* ¶¶ 342-361.

485. As noted previously, Knappe teaches that “[d]uring the multiparty communications sessions, participants are connected and stream media through the conference bridge 32,” where the media includes “audio, video, voice, data and other suitable types of real-time...traffic between source and destination

endpoints.” Knappe (Ex. 1006) at 3:26-29, 5:31-33. Accordingly, it would have been obvious to a POSITA that in the combination, conference bridge 32 would receive the audio, video, and data (such as collaboration data per Elliott) from each participant and distribute the mixed data (including audio, video, and collaboration data) to each participant, including the second remote communications device 16.

486. Thus, Knappe in combination with Elliott teaches that the conference bridge transmits the video and collaboration data to the remote second communications device 16, which renders obvious “*transmitting the first video data and the first collaboration data from the sever to the second remote client,*” as claimed.

4. Claim 20

[20.P] The method of claim 19, further comprising:

487. See analysis for [19.P]-[19.10].

[20.1] receiving second video data at the server from the second remote client

488. The prior art renders obvious receiving second video data at the server from the second remote client.

489. **First**, as discussed at [19.1.1], conference bridge 32, either alone or together with conference call manager 30, corresponds to the claimed “*server.*” Further, as discussed at [19.2.1], Knappe teaches that conference bridge 32 receives audio data from the second remote communications device 16.

490. **Second**, Knappe renders obvious that the conference bridge 32 also receives video data from the second communication device 16 because Knappe teaches that each communications device 16 is enabled to transmit real-time video data:

During the multiparty communications sessions, participants are connected and **stream media through the conference bridge 32**.

Knappe (Ex. 1006) at 5:31-33.

The **communication devices 16** comprise IP or other digital telephones, personal and other suitable computers or computing devices, personal digital assistants (PDAs), cell or other mobile telephones or handset or any other device or set of devices...**capable of communicating real-time audio, video and/or other information over the network 14**.

Id. at 3:62-4:1.

FIG. 1 illustrates a communication system 12 in accordance with one embodiment of the present invention. In this embodiment, the communication system 12 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 endpoints.

Id. at 3:24-29; *see also id.* at 3:26-31.

491. POSITA would have understood Knappe's description of "stream[ing] **media**" to refer at least to its discussion of "communicating real-time audio, **video**, and/or other information" and "data."

492. Thus, Knappe teaches that conference bridge 32 receives second video data from the second remote communications device 16, which renders obvious

“receiving second video data at the server from the second remote client,” as claimed.

5. Claim 21

[21.P] The method of claim 19, further comprising:

493. See analysis for [19.P]-[19.10].

[21.1] wherein the PSTN client is a cellular phone

494. The prior art renders obvious receiving second video data at the server from the second remote client.

495. **First**, as discussed at [19.3], Knappe discloses that PSTN telephone 18 communicates conference data through gateway 20.

496. **Second**, as discussed at [13.1] Knappe teaches implementing a PSTN telephone 18 as a wireless device that connects to PSTN 22 over wireless link 28 and that one type of a wireless link is a cellular link used by a cellular telephone. See analysis at [13.1], *supra* ¶¶ 295-297 (including citations to Knappe (Ex. 1006) at 3:41-49, 3:62-4:1).

497. Knappe’s disclosure that PSTN telephone 18 communicates to PSTN 22 using “any...suitable type of...wireless link,” and its example of a cellular telephone, which a POSITA would have understood uses a cellular wireless link, renders obvious that PSTN telephone 18 is a “*cellular phone*,” as claimed.

498. Thus, Knappe teaches the PSTN telephone 18 uses a cellular link to wirelessly connect to PSTN 22, which renders obvious that “*the PSTN client is a cellular phone.*”

6. Claim 25

[25.P] An apparatus for conducting a secure multi-participant conference call including full duplex audio, the apparatus comprising:

499. The prior art renders obvious an apparatus for conducting a secure multi-participant conference call including full duplex audio.

500. As discussed at [19.0], Knappe in combination with Elliott and Hendricks teaches a method and a system for supporting multiple participants in a secure (VPN tunneling) conference call including full duplex audio, which renders obvious “[a]n apparatus for conducting a secure multi-participant conference call including full duplex audio,” as recited in the preamble. Further, as discussed below, the prior art teaches that the apparatus comprises the claimed elements.

[25.1.1] a server being configured to receive: first audio data, first video data, and first collaboration data from a first remote client

501. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “*server.*” As further discussed at [19.1.1], the prior art teaches that the conference bridge 32 receives first audio data, first video data, and first collaboration data from a first remote communications device 16.

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

502. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As further discussed at [19.1.2], the prior art teaches that the conference bridge 32 receives first audio data, first video data, and first collaboration data through a VPN tunnel.

[25.2.1] a server being configured to receive . . . second audio data from a second remote client

503. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As further discussed at [19.2.1], the prior art teaches that the conference bridge 32 receives second audio data from a second remote communications device 16.

[25.2.2] through a second VPN tunnel;

504. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As discussed at [19.2.2], the prior art teaches that the conference bridge 32 receives the second audio data through a second Virtual Private Network (VPN) tunnel.

[25.3] a server being configured to receive . . . third audio data from a Public Switched Telephone Network (PSTN) client through a PSTN gateway;

505. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As further discussed at [19.3], the prior art teaches that the conference bridge 32 receives third audio data from a PSTN telephone 18 through a PSTN gateway 20.

[25.4] the server being configured to mix the first audio data with the second audio data into a first mixed audio data;

506. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As discussed at [19.4], the prior art teaches that the conference bridge 32 mixes the first audio data from the first remote communication device 16 with the second audio data from the second communications device 16 into a first mixed audio data.

[25.5] [the server being configured to mix] . . . second audio data with the third audio data into a second mixed audio data;

507. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As discussed at [19.5], the prior art teaches that the conference bridge 32 mixes the second audio data from the second remote communication device 16 with the third audio data from the PSTN telephone 18 into a second mixed audio data.

[25.6] [the server being configured to mix] . . . the first audio data with the third audio data into a third mixed audio data;

508. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As discussed at [19.6], the prior art teaches that the conference bridge 32 mixes the first audio data from the first remote communication device 16 with the third audio data from the PSTN telephone 18 into a third mixed audio data.

[25.7] the server being configured to transmit the first mixed audio data to the PSTN client through the PSTN gateway;

509. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As discussed at [19.7], the prior art teaches that the conference bridge 32 transmits the first mixed audio data to the PSTN telephone 18 through the PSTN gateway 20.

[25.8] [the server being configured to transmit]. . . the second mixed audio data to the first remote client through the first VPN tunnel;

510. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As discussed at [19.8], the prior art teaches that the conference bridge 32 transmits the second mixed audio data to the first remote communications device 16 through the first VPN tunnel.

[25.9] [the server being configured to transmit]. . . the third mixed audio data to the second remote client through the second VPN tunnel;

511. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As discussed at [19.9], the prior art teaches that the conference bridge 32 transmits the third mixed audio data to the second remote communications device 16 through the second VPN tunnel.

[25.10] [the server being configured to transmit]. . . the first video data and the first collaboration data to the second remote client.

512. As discussed at [19.1.1], the prior art conference call manager 30 together with conference bridge 32 correspond to the claimed “server.” As discussed at [19.10], the prior art teaches that the conference bridge 32 transmits the first video data and the first collaboration data to the second remote communications device 16.

7. Claim 26

[26.P] The apparatus of claim 25:

513. See analysis for [25.P]-[25.10].

[26.1] wherein the server is configured to receive second video from the second remote client.

514. See analysis for [20.P]-[20.1].

8. Claim 27 is Obvious Over Knappe in Combination with Drell, Elliott, the VPN Textbook, and Hendricks

[26.P] The apparatus of claim 25:

515. See analysis for [25.P]-[25.10].

[26.1] wherein the PSTN client is a cellular phone.

516. See analysis for [21.1].

C. Claims 22-24 and 28-29 Are Obvious Over Knappe in Combination with Drell, Elliott, the VPN Textbook, and Hendricks

1. Reasons to Combine Prior Art

517. I previously provided a summary of Drell at *supra* ¶¶ 115-131. I also provided the reasons a POSITA would have been motivated to combine Drell with

Knappe, with Drell as the primary reference at *supra* ¶¶ 137-157. Regardless of which reference is the primary reference, the motivation to combine the two references remain the same. Here, I reiterate and further explain the reasons a POSITA would have been motivated to combine Knappe (as the primary reference) with Drell.

518. As previously explained, a POSITA would have been motivated to combine the teachings of Knappe and Drell for multiple reasons, including to produce the obvious, beneficial, and predictable result of implementing in Knappe's conferencing devices an audio codec using a standard or proprietary compression algorithm to reduce the data transmitted.

519. Again, Knappe and Drell are analogous prior art both pertaining to the same field of endeavor of multiparty conferencing. *See, e.g.*, Knappe (Ex. 1006) at Abstract ("A system and method for...receiving a plurality of participant voice streams from a plurality of respective conference participants."); *id.* at 1:8-9 ("The present invention relates generally to the field of multiparty communications."); Drell (Ex. 1008) at 1:16-18 ("The present invention relates generally to conferencing systems, and more particularly to a videoconferencing apparatus for use with multi-point conferences.")

520. Knappe teaches that "the communication devices 16 comprise...[a] codec (coder/decoder) [that] converts audio...IP packets received from the

network. Knappe (Ex. 1006) at 4:5-17. Regarding codecs, Drell teaches that the near communications endpoint 100 utilizes an “[a]udio codec [that] receives the audio data stream...[which is] decod[ed] in accordance with a standard...or proprietary audio compression algorithm.” Drell (Ex. 1008) at 6:18-45. It would have been obvious to a POSITA for Knappe’s system, including the communication devices 16, to utilize an audio codec that employs a compression algorithm for encoding/decoding received IP packet data, as Drell teaches, because it would beneficially reduce the amount of data transmitted over the network. *See, e.g.,* Caputo (Ex. 1021) at 12-13 (“[C]ompression algorithms can be used to **reduce the amount of voice data that is transmitted**...This reduction...allows for more voice calls to transit links with limited bandwidth.”). As noted before, I personally, before the earliest claimed priority date of the ’612 patent, worked on the NBX 100 system, which allowed for a system administrator to select an audio compression algorithm for reducing the amount of audio data transmitted over the network by 50%.

521. Indeed, the combination is simply the use of a known technique (implementing a codec using standard or proprietary algorithms that compress/decompress data, per Drell) in the devices of Knappe’s system, including conference bridge 32 and the communication devices 16, to yield predictable results (reduce the amount of voice data transmitted).

522. Moreover, the result would have been predictable and there would have been a reasonable expectation of success since Knappe already includes a codec (*see* Knappe (Ex. 1006) at 4:9) and Drell itself evidences success with a compression codec in the conferencing context. Further, any modification needed to the teachings of Knappe, such as adding software programming and/or hardware to accommodate the teachings of Drell would have been within the level of ordinary skill in the art. Indeed, that there would have been a reasonable expectation of success is also evidenced by Drell.

523. Lastly, I am relying on the teachings of the references and do not suggest that physical incorporation of elements from Drell into Knappe is required in the proposed combination.

2. Claim 22

[22.P] *The method of claim 19, further comprising:*

524. See analysis for [19.P]-[19.10].

[22.1] *compressing the second mixed audio data before the transmitting to the first remote client*

525. The prior art renders obvious compressing the second mixed audio data before the transmitting to the first remote client.

526. As discussed at [14.1] Drell teaches using a compression algorithm to compress mixed audio data before transmission to a remote endpoint. *See supra*

¶ 303. **Additionally**, as discussed at [14.1], Drell teaches that the communication

process may include a standard process such as H.323, which a POSITA understands includes compression protocols and codecs described in the standard. *See supra* ¶ 304.

527. It would have been obvious to a POSITA to utilize in Knappe an audio codec (using a compression algorithm or protocol) that compresses the mixed audio data before transmission to the first remote communications device 16, as Drell teaches, because this would provide Knappe's system the benefit of reducing the amount of data transmitted over the network. *See also* Reason to Combine Knappe and Drell, *supra* ¶¶ 517-523, *see also* ¶¶ 137-157.

528. Thus, Knappe in combination with Drell teaches using a compression algorithm and/or compression protocols for compressing the second mixed audio data before transmitting the second mixed audio data to the first remote communications device 16, which renders obvious “*compressing the second mixed audio data before the transmitting to the first remote client,*” as claimed.

3. Claim 23

[23.P] The method of claim 22, further comprising:

529. See analysis for [22.P]-[22.1].

[23.1] decompressing the compressed digital audio data at the first remote client.

530. The prior art renders obvious decompressing the digital audio data at the first remote client.

531. As a threshold matter, the phrase “*compressed digital audio data*” is not previously recited in the claim, which only discusses “*compressing the second mixed audio data*” (see element [22.1]). For analysis purposes only, I will assume that the claimed “*compressed digital audio data*” in [23.1] corresponds to the output of the “*compressing*” step discussed at [22.1].

532. **First**, as discussed at [22.1], Knappe and Drell teach that the second mixed audio data is compressed with a codec (using a standard or proprietary compression algorithm) and transmitted to the first remote communications device 16.

533. **Second**, Drell teaches that its compressed data is transmitted as a stream over Internet network 106 in packets:

[N]etwork 106 may comprise a **packet** switched network, such as the **Internet**. Although a single network 106 is shown, the invention contemplates the use of two or more networks (for example, the PSTN and the Internet) to connect conference endpoints utilizing different communication protocols.

Drell (Ex. 1008) at 3:33-38, FIG. 4.

As will be described in further detail below, MP conferencing application 234 is configured to instantiate a processing train for each remote conference endpoint 102 and 104 to which near conference endpoint 100 is connected. The processing trains process audio and video data streams received from remote conferencing endpoints 102 and 104. The processed audio and video data streams are combined with each other and with locally generated audio and video streams, and the **combined audio and video streams are thereafter distributed to remote conferencing endpoints 102 and 104.**

Id. at 4:21-31; *see also id.* at 5:48-63.

534. Further, Knappe teaches that the data communicated over the network is digitally encoded. Knappe (Ex. 1006) at 4:11-14 (“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.”). It would have been obvious to a POSITA that Drell is transmitting data over Internet network 106 that is in digital form because communications over the Internet utilize digital data encapsulated in packets. *See, e.g.,* Caputo (Ex. 1021) at 12-13. Further, it would have been obvious to a POSITA that Drell’s codec performs its encoding to generate digital signals since that is how codecs work. Thus, it would have been obvious that the compressed second mixed audio data (discussed at [22.1]) is “*digital*,” as claimed. I also note that the ’612 patent specification’s only reference to “digital” data is a description of how “digital IP is routed via Internet.” ’612 patent (Ex. 1001) at 6:18-20.

535. **Third**, Drell discloses that the conference endpoint utilizes a codec (using a standard or proprietary algorithm) to decompress received compressed audio data:

In the receive mode...Audio codec 408 receives the audio data stream from communication process 404 and applies redundancy reduction decoding in accordance with a standard (e.g., ITU Recommendation G.711) or proprietary audio compression algorithm.

Drell (Ex. 1008) at 5:66-6:23.

By dynamically generating a separate instance of a processing train for each remote endpoint session, a videoconferencing system embodying the invention may easily and flexibly accommodate endpoint sessions comprising a range of connection bandwidths and communication protocols.

Id. at 6:66-7:4.

536. It would have been obvious to a POSITA that Drell's remote conference endpoint, which receives compressed audio data during the conference, would utilize a codec to decompress the compressed communications for playback to the conference participant.

537. **Additionally**, as discussed at [14.1] and [22.1], Drell teaches that the communication process may include a standard process such as H.323, which a POSITA understands includes compression and decompression protocols and codecs described in the standard. *See supra* ¶ 304. It would have been obvious to a POSITA that in H.323 communications, the device receiving compressed audio data would utilize a codec to decompress the received compressed communications for playback to the conference participant.

538. It would have been obvious to a POSITA to utilize in Knappe an audio codec (using a compression algorithm or protocol) that decompresses compressed digital audio data received at the first remote communications device 16, as Drell teaches, because this would allow Knappe's system to beneficially

reduce the amount of data transmitted over the network. Further, it would have been obvious to a POSITA to decompress received compressed audio to facilitate playback to the conference participant at the first remote communications device 16, thereby facilitating conference participation. Without decompression, the audio data could not be played back to the conference participant. *See, e.g.,* Knappe (Ex. 1006) at 4:13-17 (“IP packets received from the network 14 are converted back into analog signals and played to the user. It will be understood that the communication devices 16 may otherwise suitably encode and decode signals transmitted over or received from the network 14.”); Drell (Ex. 1008) at 3:45-46 (“[A] loudspeaker [] present[s]...speech of the remote conference participants.”); *see also*, Reason to Combine Knappe and Drell, *supra* ¶¶ 517-523, *see also* ¶¶ 137-157.

539. Thus, Knappe in combination with Drell teaches decompressing the compressed digital audio data at the first remote communications device 16, which renders obvious “*decompressing the compressed digital audio data at the first*

540. *remote client*,” as claimed.

4. Claim 24

[24.P] The method of claim 22, further comprising:

541. See analysis for [22.P]-[22.1].

[24.1] wherein the compressing the second mixed digital audio data is compressed using a standard audio codec.

542. The prior art renders obvious the compressing the second mixed digital audio data is compressed using a standard audio codec.

543. As a threshold matter, the phrase “*second mixed digital audio data*” is not previously recited in the claim, which only discusses “*compressing the second mixed audio data*” (see element [22.1]). For analysis purposes only, I will assume that the claimed “*second mixed digital audio data*” in [24.1] corresponds to the input of the “*compressing*” step discussed at [22.1].

544. **First**, as discussed at [22.1], Knappe and Drell teach that the second mixed audio data is compressed with a codec and transmitted to the first remote communications device 16.

545. **Second**, it would have been obvious to a POSITA that the second mixed audio data is “digital” audio data for several reasons. First, Knappe discloses that the conference output streams produced by the mixer are “converted by a corresponding converter and buffered by a corresponding output buffer for transmission to the corresponding participant.” Knappe (Ex. 1006) at 6:38-42. A POSITA would have been aware that buffering digital audio data is relatively trivial and could be accomplished in various ways, such as using well-known data structures like a first-in, first-out data queue or circular data buffer. Buffering audio data in analog form, however, would be substantially more complicated. Thus, a POSITA would have understood that the mixed audio data output by

Knappe's mixer 80 is "*digital*" audio data. This understanding is further confirmed by Knappe's incorporation (at 8:40-43) of the contents of U.S. Patent No.

6,011,851. The '851 patent describes processing digital audio:

The spatial audio processor and context switching system 26 selectively switches incoming sound sources between desired foreground and background priorities. New audio applications may be subsequently launched with their associated audio paths assigned to any available incoming source stream for perceptual assignment to a new background or foreground location. In one implementation, **audio processing is performed on digitally sampled 16-bit linear audio samples, with the resultant output also in 16-bit linear form.**

U.S. 6,011,851 (Ex. 1045) at 4:10-18. Accordingly, a POSITA reading Knappe's disclosure would have understood that the mixed audio output generated for the first remote communication device 16 ("*second mixed audio data*") is "*digital*."

546. It would have been obvious to a POSITA that Drell describes mixing audio in a digital format because Drell describes performing audio mixing using software running on a computer. Specifically, Drell "utilizes software-based processing of video and audio data streams." Drell (Ex. 1008) at 6:63-64. Drell's "MP conferencing application" includes an "audio mixing module 308." *Id.* at 4:32-38. Because all, or substantially all, computers for executing software are digital computers, it would have been obvious to a POSITA that Drell's software-based "audio mixing module 308" processes audio data in "*digital*" form.

547. **Third**, Drell teaches that audio codec 408 (discussed at [22.1]) is “using a standard... audio compression algorithm.” *Id.* at 6:43-44. Further, the H.323 codecs (discussed at [22.1]) are also standard. *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... This standard also specifies lowest-common denominator compression protocols, such as G.711...G.723 and G.729.”). It would have been obvious to a POSITA to utilize a standard audio codec (using a standard compression algorithm or protocol) that compresses the mixed audio data, as Drell teaches, because this would provide Knappe’s system the benefit of reducing the amount of data transmitted over the network. Further, a POSITA would have recognized that by utilizing standard audio codecs to compress and decompress audio data throughout the network would beneficially allow Knappe to avoid investing time and money to develop proprietary compression algorithms. Furthermore, using standard audio codecs to compress and decompress in Knappe would beneficially facilitate interoperability among conferencing devices from different manufacturers. *See also*, Reason to Combine Knappe and Drell, *supra* ¶¶ 517-523, *see also* ¶¶ 137-157.

548. Thus, Knappe in combination with Drell teaches that the compressing the second mixed digital audio data is compressed (see discussion at [22.1]) with a

standard audio codec, which renders obvious “*wherein the compressing the second mixed digital audio data is compressed using a standard audio codec*” as claimed.

5. Claim 28

[28.P] The apparatus of claim 25:

549. See analysis for [25.P]-[25.10].

[28.1] *wherein the second mixed audio data is compressed digital audio data.*

550. The prior art renders obvious the second mixed audio data is compressed digital audio data.

551. **First**, as discussed at [19.5] and [19.8] (as well as [25.5] and [25.8]), the prior art teaches mixing and transmitting the second mixed audio data to the first remote communications device 16.

552. **Second**, as discussed at [14.1], [22.1] and [23.1], Drell teaches using a compression algorithm to compress mixed audio data before transmission to a remote endpoint. Drell (Ex. 1008) at 6:33-45, FIG. 4, 6:33-45, 6:66-7:4; *see supra* ¶ 303. **Additionally**, as discussed at [14.1], [22.1] and [23.1], Drell teaches that the communication process may include a standard process such as H.323, which a POSITA understands includes compression protocols and codecs described in the standard. *See supra* ¶ 304.

553. It would have been obvious to a POSITA to utilize in Knappe an audio codec (using a compression algorithm or protocol) that compresses the

mixed audio data before transmission to the first remote communications device 16, as Drell teaches, because this would provide Knappe's system the benefit of reducing the amount of data transmitted over the network. *See also*, Reason to Combine Knappe and Drell, *supra* ¶¶ 517-523, *see also id.* ¶¶ 137-157.

554. **Third**, as discussed in [23.1] and [24.1], Drell and Knappe teach mixing and compressing audio data in the form of “*digital*” audio data. *See e.g.* analysis at [23.1] *supra* ¶¶ 533-534. Thus, it would have been obvious that Drell's compressed second mixed audio data is “*digital*,” as claimed.

555. Thus, Knappe in combination with Drell teaches that the second mixed audio data is compressed with an audio codec before transmission to the first remote communications device 16, which renders obvious “*wherein the second mixed audio data is compressed digital audio data*,” as claimed.

6. Claim 29

[29.P] The apparatus of claim 85:

556. See analysis for [28.P]-[28.1].

[29.1] wherein the compressed audio data is compressed using a standard audio codec.

557. As discussed at [28.1], the prior art teaches that the second mixed audio data is compressed with an audio codec into compressed digital audio data. As discussed at [24.1], the prior art teaches that the audio codec is a standard audio codec.

558. Thus, Knappe in combination with Drell teaches that the compressed audio data is compressed with a standard audio codec before transmission to the first remote communications device 16, which renders obvious “*wherein the compressed audio data is compressed using a standard audio codec*” as claimed.

X. SECONDARY CONSIDERATIONS

559. 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 '612 patent.

560. I am not aware of any allegations made by the named inventor of the '612 patent or any assignee of the '612 patent that any secondary considerations tend to rebut the obviousness of any of the claims of the '612 patent.

561. 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.

XI. DECLARATION

562. 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.