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.

Case IPR2022-00378
U.S. Patent No. 9,094,525
(Claims 1-8)

PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,094,525

TABLE OF CONTENTS

I. INTRODUCTION 1

II. MANDATORY NOTICES UNDER 37 C.F.R § 42.8(a)(1) AND (B) 2

A. Real Party-In-Interest Under 37 C.F.R. § 42.8(b)(1) 2

B. Lead and Back-Up Counsel and Service Information Under 37 C.F.R. § 42.8(b)(3) and (b)(4) 2

C. Related matters Per 37 C.F.R. § 42.8(b)(2) 3

III. PAYMENT OF FEES – 37 C.F.R. § 42.103 3

IV. GROUNDS FOR STANDING UNDER 37 C.F.R. § 42.104(a) 3

V. CHALLENGE UNDER 37 C.F.R. § 42.104(B) AND RELIEF REQUESTED 3

VI. OVERVIEW OF THE ’525 PATENT 5

A. The Specification 5

B. The Challenged Claims. 7

C. Level of Ordinary Skill in the Art 8

D. Claim Construction 8

E. The ’525 Patent Priority Date 10

VII. THE BOARD’S DISCRETION UNDER 35 U.S.C. § 314(A) SHOULD NOT PRECLUDE INSTITUTION 10

VIII. IDENTIFICATION OF WHERE EACH ELEMENT OF EACH CLAIM IS FOUND IN THE PRIOR ART 14

A. Ground 1: Claims 1-4 Are Obvious Over Drell in Combination With Knappe and Voois 14

B. Ground 2: Claims 5-8 Are Obvious Over Drell in Combination with Knappe, Voois, Elliott, and the VPN Textbook 54

C. Ground 3: Claims 1-4 Are Obvious in Light of Drell in Combination with Knappe, Elliott, and the VPN Textbook 67

D. Ground 4: Claims 5-8 Are Obvious Over Drell in Combination with Knappe, Elliott, and the VPN Textbook 68

IX. CONCLUSION 68

TABLE OF AUTHORITIES

Cases

In re Etter,
756 F.2d 852 (Fed. Cir. 1985) (en banc) 23

In re Mouttet,
686 F.3d 1322 (Fed. Cir. 2012) 23, 26, 58, 61

Intel Corp. v. VLSI Tech. LLC,
IPR2020-00158, Paper 16 (May 20, 2020). 12

Phillips v. AWH Corp.,
415 F.3d 1303 (Fed. Cir. 2005) (*en banc*) 8, 10

Sand Revolution II, LLC v. Continental Intermodal Group-Trucking LLC,
IPR2019-01393, Paper 24, 7 (June 16, 2020) 12, 13

Sotera Wireless, Inc. v. Masimo Corp.,
IPR2020-01019, Paper 12 (PTAB Dec 1, 2020) 14

VMware, *Inc. v. Intellectual Ventures I LLC*,
IPR2020-00470, Paper 13 at 20 (PTAB Aug. 18, 2020) 14

Statutes

35 U.S.C. § 312 1

Rules

37 C.F.R. § 41.10(b) 2

37 C.F.R. § 42.104(a) 3

37 C.F.R. § 42.15(a) 3

37 C.F.R. § 42.6(c) 3

Regulations

83 Fed. Reg. 51341 (October 11, 2018) 8

TABLE OF EXHIBITS

Ex.	Description
1001	U.S. Patent No. 9,094,525 ("the '525 patent")
1002	Patent Assignments for the '525 patent
1003	Declaration of Dr. Henry H. Houh
1004	*Curriculum Vitae* for Dr. Henry H. Houh
1005	File History of Application No. 13/674,233 (the '233 application)
1006	U.S. Patent No. 7,180,997 ("Knappe")
1007	U.S. Patent No. 6,690,654 ("Elliott")
1008	U.S. Patent No. 7,089,285 ("Drell")
1010	Guichard *et al*., "MPLS and VPN Architectures, A practical guide to understanding, designing and deploying MPLS and MPLS-enabled VPNs," Cisco Press, 2001 ("VPN Textbook")
1011	Declaration of Ingrid Hsieh-Yee, Ph.D., under 37 C.F.R. § 1.68
1012	U.S. Patent No. 3,060,265
1013	U.S. Patent No. 4,139,731
1014	U.S. Patent No. 3,601,530
1015	U.S. Patent No. 4,796,293
1016	A Primer on the T.120 Series Standard
1017	T.120 Standard (1996)
1018	U.S. Patent No. 6,237,025
1019	U.S. Patent No. 6,782,413
1020	U.S. Patent No. 6,980,526

Ex.	Description
1021	Caputo, "Cisco Packetized Voice & Data Integration," McGraw-Hill, 2000
1022	De Moraes *et al.*, "The Internet Multicast from ITS: How it was Done and Implications for the Future," IEEE Communications Magazine, January 1995
1023	U.S. Patent Publication No. 2004/0261116 A1
1024	U.S. Patent No. 6,850,496
1025	U.S. Patent No. 6,553,515
1026	U.S. Patent No. 6,633,848
1027	Computer Desktop Encyclopedia
1036	U.S. Patent No. 6,215,515 ("Voois")
1046	Declaration of C.J. Alice Chuang

I. INTRODUCTION

Zoho Corporation and Zoho Corporation Pvt., Ltd. ("Petitioner") respectfully requests *inter partes* review ("IPR") under 35 U.S.C. §§ 311, 312, 314(a), and 37 C.F.R. § 42.100, for claims 1-8 ("Challenged Claims") of U.S. Patent No. 9,094,525, titled "Audio-Video Multi-Participant Conference Systems Using PSTN and Internet Networks" ("'525 patent," Ex_1001). Meetrix IP, LLC ("Patent Owner") is the patent owner of record. Ex_1002 (conveying the '525 patent family to Meetrix).

The '525 patent is directed to a multi-participant conference system and method enabling "audio and video telecommunications for collaboration over hybrid networks," where audio data from multiple participants are mixed together and transmitted to other participants. Ex_1001 at Abstract, 1:19-20. The hybrid network includes a Public Switched Telephone Network (PSTN) and the Internet. *Id.* at 4:56-5:3, FIG. 3. As described below and in the Declaration of Dr. Henry H. Houh (Ex_1003), these concepts were well-known by the claimed priority date for the '525 patent. The primary reference ("Drell," Ex_1008) describes a multi-point video conference system that receives audio and video signals from multiple participants, including a moderator and PSTN telephone clients. Ex_1008, Abstract. Other claimed concepts, such as using virtual private network (VPN) tunneling and voice over IP (VoIP) for conferencing communications, were also

well known in the prior art. The combination of these prior art teachings would have been obvious to one of ordinary skill in the art, and thus the claims of the '525 patent are invalid.

II. MANDATORY NOTICES UNDER 37 C.F.R § 42.8(A)(1) AND (B)

A. Real Party-In-Interest Under 37 C.F.R. § 42.8(b)(1)

The real parties-in-interest are Zoho Corporation and Zoho Corporation Pvt., Ltd.

B. Lead and Back-Up Counsel and Service Information Under 37 C.F.R. § 42.8(b)(3) and (b)(4)

Lead Counsel:
Hector J. Ribera (Reg. No. 54,397)
Marton Ribera Schumann & Chang LLP
548 Market St., Suite 36117
San Francisco, CA 94104
Email: hector@martonribera.com
Tel: (415) 360-2512

Back-up Counsel:
C.J. Alice Chuang (Reg. No. 65,914)
Marton Ribera Schumann & Chang LLP
548 Market St., Suite 36117
San Francisco, CA 94104
Email: cjalice@martonribera.com
Tel: (415) 360-2516

Petitioner consents to electronic service by delivering the documents to the email addresses of primary and backup counsel. A Power of Attorney is being filed concurrently under 37 C.F.R. § 41.10(b).

C. Related matters Per 37 C.F.R. § 42.8(b)(2)

To the best of Petitioner's knowledge, the '525 patent is involved in the following pending district court cases:

Case Name	Case Number	Court	Date Filed
Meetrix IP, LLC v. Zoho Corporation	6:21-cv-01288	W.D. Tex.	12/10/21
Meetrix IP, LLC v. Verizon Comms., Inc.	6:21-cv-01289	W.D. Tex.	12/10/21

III. PAYMENT OF FEES – 37 C.F.R. § 42.103

Petitioner pays the fee for this petition as required by 37 C.F.R. § 42.15(a) contemporaneously with the filing of this petition.

IV. GROUNDS FOR STANDING UNDER 37 C.F.R. § 42.104(A)

Petitioner certifies under 37 C.F.R. § 42.104(a) that the '525 patent is available for IPR and that Petitioner is not barred or estopped from requesting an IPR challenging the validity of the challenged claims on the identified grounds. Petitioner has not filed a civil action challenging the validity of any claim of the '525 patent.

V. CHALLENGE UNDER 37 C.F.R. § 42.104(B) AND RELIEF REQUESTED

Petitioner requests IPR and cancellation of claims 1-8 of the '525 patent under pre-AIA 35 U.S.C. § 103. Pursuant to 37 C.F.R. § 42.6(c), copies of the Exhibits are filed herewith.

Under 37 C.F.R. §§ 42.104(b)(1) and (2), these grounds are offered as reasons for cancelling claims 1-8 of the ’525 patent:

Ground	Claims	Statutory Basis	Reference(s)
1	1-4	§ 103	Drell, Knappe, and Voois
2	5-8	§ 103	Drell, Knappe, Voois, Elliott, and VPN Textbook
3	1-4	§ 103	Drell and Knappe
4	5-8	§ 103	Drell, Knappe, Elliott, and VPN Textbook

U.S. Patent No. 7,089,285 to Drell (“Drell” – Ex_1008) was filed on October 5, 2000.

U.S. Patent No. 7,180,997 to Knappe (“Knappe” – Ex_1006) was filed on September 6, 2022.

U.S. Patent No. 6,690,654 to Elliott, *et al*. (“Elliott” – Ex_1007) was filed on October 1, 1998.

U.S. Patent No. 6,215,515 to Voois, *et al*. (“Voois” – Ex_1036) was filed on June 10, 1998 and issued April 10, 2001.

“MPLS and VPN Architectures,” (“VPN Textbook” – Ex_1010) by Guichard *et al*., was published in 2001. *See* Declaration of Ingrid Hsieh-Yee, (Ex_1011), ¶¶16, 23, 95-111.

Drell, Knappe, and Elliott are prior art under 35 U.S.C. § 102(e). Voois and the VPN Textbook is prior art under 35 U.S.C. § 102(b).

VI. OVERVIEW OF THE '525 PATENT

The '525 patent was filed on November 12, 2012 as a continuation of a string of applications, the first of which was a provisional application filed March 10, 2003. Ex_1001.

A. The Specification

The '525 patent is directed to a "a multi-participant" conferencing system that "includes a PSTN client, at least one remote client and a first participant client." *Id*., Abstract.

To accomplish this, the specification describes use of a hybrid Internet and PSTN network:

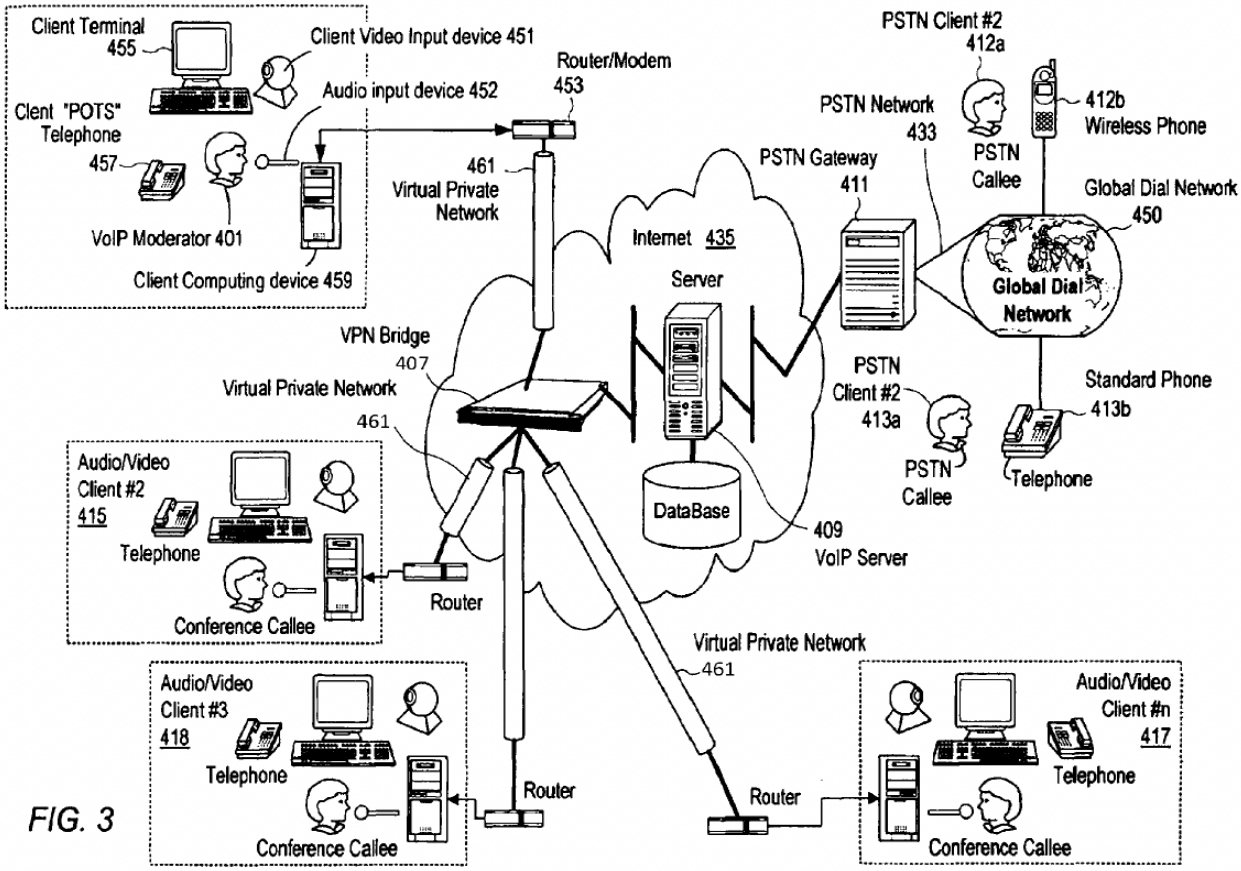


FIG. 3

Ex_1001, FIG. 3; Ex_1003 ¶¶50-51.

Specifically, Figure 3 depicts Internet-based clients, including moderator 401, and audio video clients 415, 417, and 418 "connected through routers or modems 453 preferably in a virtual private network configuration 461" with "the VPN bridge 407" located in the center of a "star topology," with connections to the VPN bridge serving as a "VPN tunnel." Ex_1001, 5:24-49, 6:37-59, FIGS. 4-5; Ex_1003 ¶52. The PSTN network includes "telephony clients using standard wired 413 or wireless telephone 412 systems," with these PSTN clients connected

to the Internet-based clients via PSTN gateway 411. Ex_1001, 5:16-18; Ex_1003 ¶53.

The specification also discloses performing encoding and decoding of audio data using codecs of the prior art H.323 standard and the transmission of collaboration data per the prior art T.120 standard. Ex_1001, 2:57-63, 3:12-14, FIG. 2; Ex_1003 ¶55.

B. The Challenged Claims.

Claims 1-8 are the Challenged Claims. Claim 1 is the independent claim. Independent claim 1 is directed to a system supporting a multi-participant conference that recites:

- a "first mixer" that mixes a PSTN client audio stream with a moderator audio-video data stream into a first mixed data stream;
- a "first transport output" that transmits the first mixed data stream to at least one remote client that receives the first mixed data stream, the at least one remote client communicatively coupled to the Internet, which generates a remote client audio-video data stream;
- a "second mixer" that mixes the moderator audio-video data stream with the remote client audio-video data stream into a second mixed data stream; and
- a second transport output that transmits a mixed audio data stream, corresponding to the second mixed data stream, to the PSTN client.

Ex_1001, 10:2-17. Claim 2 further adds the a "third mixer" that mixes the audio data from the PSTN client with the audio-video data from the remote client into a third mixed data stream and communicates it to the moderator. *Id.*, 10:18-22.

Claims 3-8 add further known limitations, such as having a VoIP decoder, audio decompressor, and using a VPN tunnel. *Id.*, 10:23-43; Ex_1003 ¶¶58-60.

C. Level of Ordinary Skill in the Art

The subject matter of the '525 patent generally relates to the technical field of audio/video conferencing systems. A Person of Ordinary Skill in the Art ("POSITA") in March 2003 would have had a working knowledge of the conferencing art. Ex_1003 ¶27. A POSITA would have had a bachelor's degree in computer science, computer engineering, or an equivalent, and three or more years of professional experience relating to conferencing systems in packet-based networks, or without said professional experience, further education relating to conferencing systems in packet-based networks. *Id.*¶¶25-27.

D. Claim Construction

During IPR claims are construed according to the "*Phillips* standard," as set forth in *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (*en banc*). *See* 83 Fed. Reg. 51341 (October 11, 2018).

1. "mixed data stream"

Claim 1 recites a "*first mixed data stream*" and a "*second mixed data stream*." Each "*mixed data stream*" is produced by mixing an input data stream from a different client with a "*moderator audio-video data stream*." Ex_1003 ¶68. This understanding is confirmed by other claim language in the '525 patent. *Id.* ¶69. For example, claim 1 refers to a "*mixed audio data stream, corresponding to*

the second mixed data stream." If the "*second mixed data stream*" did not include both audio and video data (e.g., if it included only audio data), then the separately recited "*mixed **audio** data stream*" would have the same meaning as the "*second mixed data stream*." *Id*. Claim 6 further confirms that each claimed "*mixed data stream*" includes at least audio and video data, requiring "*encod[ing] the second mixed data stream*" and, separately, "*encod[ing] the second mixed voice data*." *Id*. ¶70. Claim 9, which includes two separate "*mixing*" steps, also confirms this understanding. *Id.* ¶71. In the first mixing step, two "*audio data*" inputs and one "*video data*" input are mixed "*into mixed audio-video data*." In contrast, the second mixing step includes three "*audio data*" inputs that are mixed "*into mixed audio data*." This confirms the claims' understanding that mixing inputs that include both audio and video produces mixed data that also include both audio and video data. *Id*.

This interpretation is also consistent with the '525 patent specification, which refers to a "mix of audio and video data." Ex_1001, 1:51; Ex_1003 ¶72. Furthermore, claim 6 recites, for example, "*transmitting the second mixed data stream to the PSTN [public switched telephone network] client*." The '525 patent specification describes providing "audio, **video**, and data **connections** between clients conferencing on the Internet **and clients conferencing on standard telephone systems**." Ex_1001, 4:64-67; Ex_1003 ¶72. Accordingly, as would

have been understood by a POSITA in light of the '525 Patent specification, "*mixed data stream*" is a "data stream including at least audio data and video data." Ex_1003 ¶73.

This Petition analyzes the remaining claim terms consistent with ordinary and customary meaning as would be understood by a POSITA in light of the specification. *Phillips*, 415 F.3d at 1314-17; Ex_1003 ¶67.

E. The '525 Patent Priority Date

The '525 patent issued from U.S. Patent Application No. 13/674,233 (the '233 application), filed November 12, 2012. Ex_1001, 1. The '233 application is a continuation of a series of applications that claim the benefit of U.S. Provisional Patent Application No. 60/453,307, filed March 10, 2003. *Id.* For this Petition, Petitioner relies on March 10, 2003 as the priority date for the Challenged Claims.

VII. THE BOARD'S DISCRETION UNDER 35 U.S.C. § 314(A) SHOULD NOT PRECLUDE INSTITUTION

On June 21, 2022, the Director of the USPTO issued a memorandum setting forth interim procedures for determining institution of IPRs in view of the factors set forth in *Apple Inc. v. Fintiv, Inc.* *See* https://www.uspto.gov/sites/default/files/documents/interim_proc_discretionary_denials_aia_parallel_district_court_litigation_memo_20220621_.pdf ("Memorandum"). The Memorandum clarified (1) that the PTAB will not rely on the *Fintiv* factors to discretionarily deny institution where a petition presents

compelling evidence of patentability, and (2) consistent with *Sotera Wireless*, the PTAB will not discretionarily deny institution where a petitioner presents a stipulation not to pursue in a parallel proceeding the same grounds or any grounds that could have been raised before the PTAB. In light of both these factors, the Board should not exercise its discretion to deny institution here.

First, this Petition presents compelling evidence of unpatentability. As set forth below in this Petition and in the accompany declaration of Dr. Houh (Ex_1003), the subject matter of the '525 patent relates to well-known concepts of multi-party conferencing between participants in hybrid Internet and PSTN networks. By the '525 patent's claimed priority date, there were many available multi-party conferencing systems and a POSITA was well aware of how to use known communications technologies such as audio mixing, audio codecs, VPN tunneling, and IP packets to communicate conferencing data across the hybrid networks. *See e.g.* Ex_1003 ¶¶74-123. Thus, this is not a case in which Petitioner barely meets the IPR standards with obscure technical references. Here, there is compelling evidence that the claims of the '525 patent merely recite well-established multi-party conferencing technology.

Second, upon the filing of this Petition, Petitioner intends to file in the district court litigation a request for pre-institution stay of proceedings with a

stipulation that, if the IPR is instituted, Petition will not rely on any grounds that is raised or could have been in this Petition.

Finally, the remaining *Fintiv* factors weigh against denial:

Fintiv factor 1, likelihood of a stay. As noted, Petition is seeking a pre-institution stay and if needed, will renew the request for a stay upon institution. Even if Patent Owner suggests that district court litigation may not be stayed, the Board "will not attempt to predict how the district court in the related district court litigation will proceed because the court may determine whether or not to stay any individual case, including the related one, based on a variety of circumstances and facts beyond [its] control and to which the Board is not privy." *Sand Revolution II, LLC v. Continental Intermodal Group-Trucking LLC*, IPR2019-01393, Paper 24, 7 (June 16, 2020) (Informative); *see Intel Corp. v. VLSI Tech. LLC*, IPR2020-00158, Paper 16, *7 (May 20, 2020).

Fintiv factor 2, trial date proximity to Final Written Decision. No trial date has been set in the district court litigation. The district court litigation is still in the early stages, having been transferred between divisions in the W.D. Texas and stayed for several months pending re-assignment. Given the parties have yet to initiate discovery or conduct claim construction, it is likely a Final Written Decision would issue far in advance of any potential trial date in the district court litigation, particularly as Petitioner will be seeking a stay.

Fintiv factors 3 and 4, investment in the parallel proceedings and overlap in issues.

As of the filing this Petition, the district court remains in the early stages. Although the case was originally filed in December 2021, early procedural motions and an intra-district transfer and stay of the case means that there has been no fact discovery, no substantive motions, and no claim construction proceedings. In short, no substantive action has been taken in the district court and very little court and party investment has been made in that proceeding. *Sand Revolution*, IPR2019-01393, Paper 24 at 10-11 (two-page *Markman* Order did not demonstrate high level of investment of time and resources).

Critically, the invalidity grounds, arguments, and evidence presented in this Petition will be different than those presented in the district court. As noted above, Petitioner will stipulate in the Litigation that if the Board institutes IPR, it will not pursue invalidity on the same grounds at issue in this Petition or any grounds that could have been raised in this Petition. This mitigates concerns of duplicated efforts and conflicting decisions. *See Sotera Wireless, Inc. v. Masimo Corp.* , IPR2020-01019, Paper 12 (PTAB Dec 1, 2020); *see also VMware*, *Inc. v. Intellectual Ventures I LLC*, IPR2020-00470, Paper 13 at 20 (PTAB Aug. 18, 2020) (finding similar stipulation "mitigates any concerns of duplicative efforts between the district court and the Board, as well as concerns of potentially conflicting decisions"). As such, these factors weigh against discretionary denial.

VIII. IDENTIFICATION OF WHERE EACH ELEMENT OF EACH CLAIM IS FOUND IN THE PRIOR ART

Petitioner respectfully requests cancellation of claims 1-8 of the '525 patent.

A. Ground 1: Claims 1-4 Are Obvious Over Drell in Combination With Knappe and Voois

1. The Prior Art

a) Summary of Drell

Drell describes multi-point (MP) conferencing, specifically video conferencing among a plurality of endpoints, as shown in Figure 1:

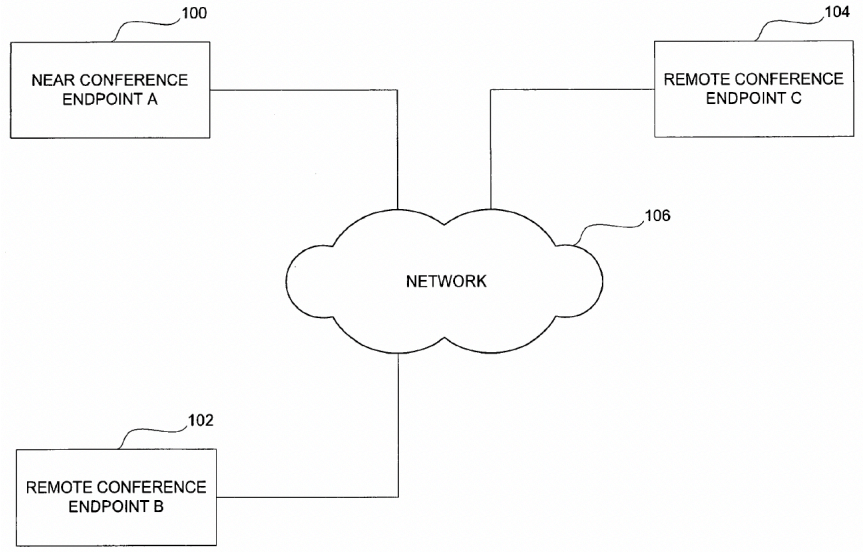


FIG. 1

Ex_1008, FIG. 1; Ex_1003 ¶129.

Specifically, Drell discloses a videoconferencing apparatus "to combine and distribute audio and video signals received from a plurality of remote conference endpoints[.]" Ex_1008, 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*., 3:20-23. Drell "contemplates the use of two or more networks (for example, the PSTN and the Internet) to connect conference endpoints utilizing different communication protocols." Ex_1008, 3:33-38; Ex_1003 ¶130.

b) Summary of Knappe

Knappe discloses a multiparty conferencing system over a hybrid 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." Ex_1006, 3:41-45; Ex_1003 ¶136.

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

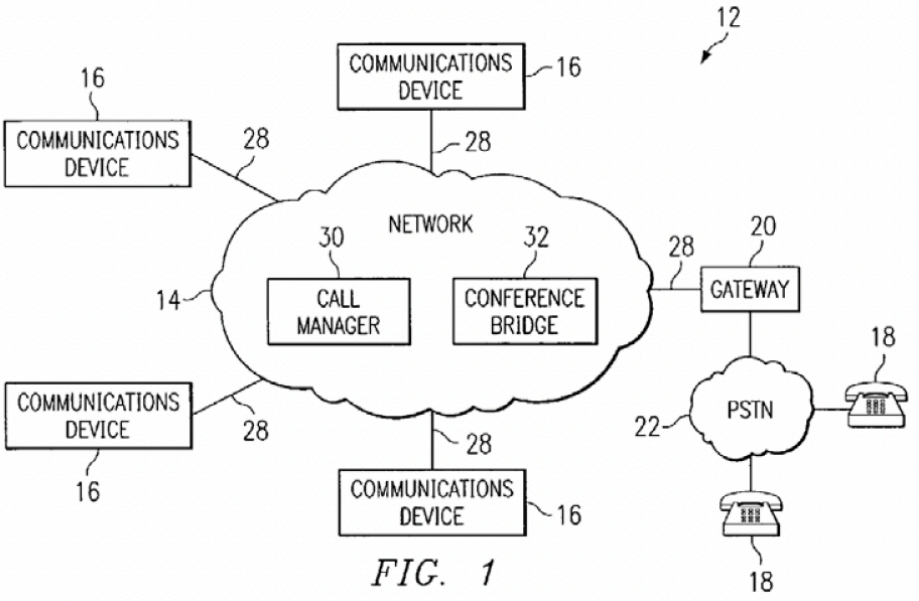
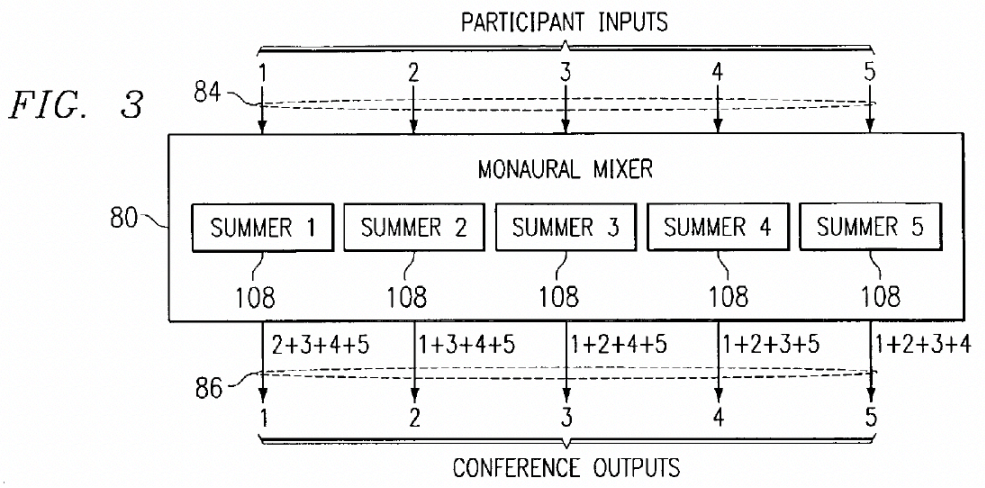


FIG. 1

Ex_1006, FIG. 1; Ex_1003 ¶137.

Knappe's "participants are connected and stream media through the conference bridge 32," which provides conference call and other suitable audio, video, and/or real-time multi-party communication sessions[.]" Ex_1006, 5:32-33, 4:58-60; Ex_1003 ¶141. Knappe's conference bridge 32 "combines the streams in summers 82 to generate conference output streams 86 for each participant to a conference call":

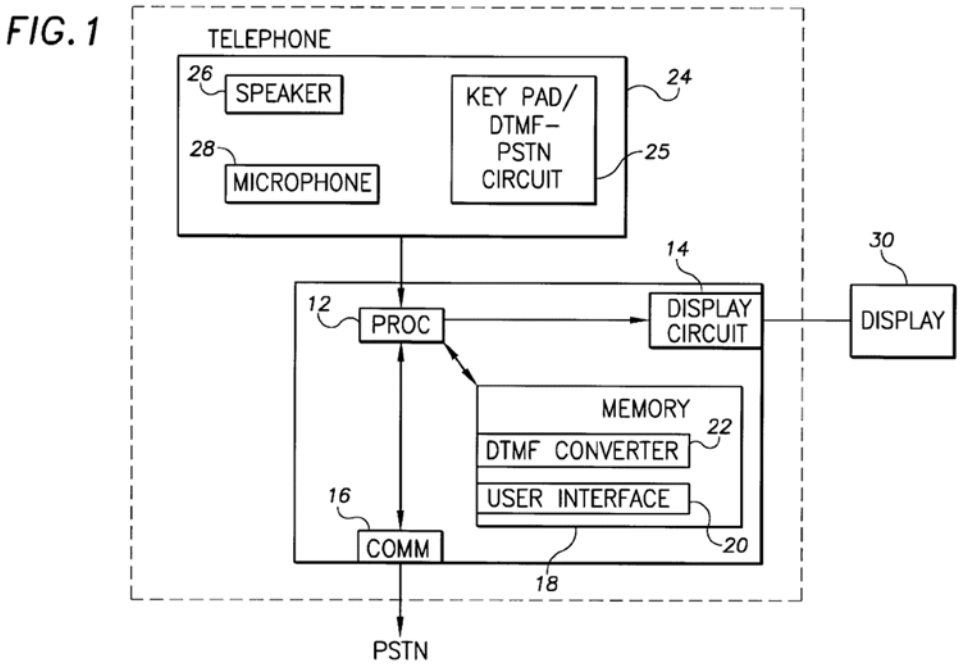


Ex_1006, FIG. 3; 6:59-62; Ex_1003 ¶¶142-145.

c) Summary of Voois

Voois describes "videoconferencing over a conventional communications channel." Ex_1036, 1:34-37; Ex_1003 ¶146. Specifically, Voois teaches videophones for communicating "video and audio data over a plain old telephone service (POTS) line." Ex_1036, Abstract; Ex_1003 ¶146.

In Figure 1, reproduced below, Voois shows an example of a videophone system "having a telephone for operating a videophone user interface" and a "[c]ommunication component 16 … to provide an interface with a communication network, such as a public switched telephone network (PSTN), the internet, local area networks, and/or wide area networks." Ex_1036, 3:62-4:4.



Id., FIG. 1; Ex_1036 ¶147.

2. Reasons to Combine Prior Art

a) Reasons to Combine Drell and Knappe

A POSITA when considering Drell would have also considered Knappe as they are analogous, both pertaining to the field of multiparty conferencing, including mixing data from Internet and PSTN conference endpoints. *E.g.* Ex_1008, Abstract, 3:33-37; Ex_1006, 3:41-51, 6:59-62; Ex_1003 ¶¶148-149.

(1) moderator

A POSITA of conferencing systems would have recognized that it would be desirable to have a leader or moderator to initiate and organize the conference. Ex_1003 ¶150. Consistent with this knowledge, Knappe teaches that at least "one participant may be given the priority of moderator" who could "control the flow and/or organization of the conference." Ex_1006, 8:52-56 & 1:53-55; Ex_1003 ¶¶150-151. Since Knappe teaches that at least one of the participants is a moderator, it would have been obvious to a POSITA for either the user of the remote conference endpoint 104 or near conference endpoint 100 of Drell to be a moderator because doing so would beneficially allow the user (acting as the moderator) to initiate, organize, control, and influence the conference. Ex_1003 ¶¶151-152. Furthermore, as there are a finite number of participants in a conference, it would have been obvious to try designating one of Drell's endpoint users as a moderator. *Id.* ¶152. The combination of Drell and Knappe is merely the ordinary use of a common technique (designating a conference endpoint as a moderator per Knappe) in the similar system of Drell to yield predictable results of initiating, organizing, controlling, and influencing the conference. *Id*.

(2) mixing

A POSITA also would have been motivated to combine these references because Drell teaches that its system "combine[s] and distribute[s] audio and video

signals received from a plurality of remote conference endpoints," while Knappe supplements and explains how to mix audio data from multiple conference endpoints. Ex_1006, Abstract; Ex_1003 ¶¶154-155. Implementing Knappe's mixers into Drell would result produce beneficial results. Ex_1003 ¶¶154-159. Specifically, Knappe teaches using a mixer 80 with a plurality of summers *(e.g.,* summers 1 to 5, one for each participant), where each mixes audio data from the other participants. Ex_1006, 6:59-7:15, FIGS. 1-3; Ex_1003 ¶155. Thus, 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. Ex_1003 ¶156. Such an implementation would complement Drell's teaching of generating separate instances for each endpoint. *See* Ex_1008, 2:19-21, 6:66-7:4 (describing MP conferencing application generating separate instances of a processing train); Ex_1003 ¶156; *see also id*. ¶137.

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. Ex_1003 ¶158. 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. *Id.* 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. Ex_1006, 7:1-11; Ex_1003 ¶158. 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 the exclusion of audio of the participant receiving the mixed audio, thereby avoiding undesirable audio effects. Ex_1003 ¶¶158-160.

(3) encapsulating VoIP packets

Drell teaches using codecs to encode mixed voice data before transmission over the Internet. Ex_1008, 6:33-45, 3:33-34; Ex_1003 ¶161. A POSITA would understand that Drell transmits audio data using IP packets because the Internet uses IP packets for communications. Ex_1003 ¶161.

Knappe supplements Drell and teaches that "[r]eal-time IP packets such as VoIP packets are encapsulated in real-time transport protocol (RTP) packets for transmission over the network 14." Ex_1006, 3:50-58; Ex_1003 ¶162. 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. Ex_1003 ¶162; Ex_1021 at 22 ("The

Real- Time Protocol (RTP)…was created to transport real-time traffic**,** such as video and audio, over the Internet.").

Furthermore, Drell teaches that its "communication process … may … comprise an … H.323 (packet-based) process." Ex_1008, 2:26-28. A POSITA would have known that H.323 standards were adopted to foster interoperability of equipment from multiple vendors. Ex_1003 ¶163. 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. *Id.* 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 use of both VoIP and RTP were known in the conferencing context. *Id*. ¶164.

(4) Gateway

Drell contemplates "use of two or more networks (for example, the PSTN and the Internet) to connect conference endpoints utilizing different communication protocols." Ex_1008, 3:20-37. A POSITA would have recognized that since Drell's Internet and PSTN use different protocols, it would be beneficial to use a device that can convert data between the two networks. Ex_1003 ¶165. To that end, Knappe teaches using a "gateway… [that] provides conversion between analog and/or digital formats." Ex_1006, 4:18-20, *see also id*. 4:20-25; Ex_1003 ¶165. 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. Ex_1003 ¶166. Doing so would 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* Ex_1008, 3:20-37; Ex_1003 ¶166. 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 both relate to similar conferencing systems and using gateways was conventional and known to be applied in the conferencing context. Ex_1003 ¶167; *see also id*. ¶153, 160, 164.

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. *Id.* ¶153. Petitioner's proposed combination permits, but does not require, physical incorporation of elements from Drell and Knappe. Ex_1003 ¶168; *see In re Mouttet*, 686 F.3d 1322, 1332 (Fed. Cir. 2012); *In re Etter* 756 F.2d 852, 859 (Fed. Cir. 1985) (en banc) (noting that criterion for obviousness is not whether references can be combined physically, but whether claimed invention is rendered obvious by the teachings of the prior art as a whole).

b) Reasons to Combine Drell and Voois

A POSITA when considering the teachings of Drell would have also considered the teachings of Voois, as they are analogous art pertaining to multiparty conferencing over PSTN. Ex_1008, 3:33-37; Ex_1036, 1:33-37, 3:62-4:4; Ex_1003 ¶169.

As discussed above, Drell in combination with Knappe teaches using a gateway between a PSTN and Internet network. As also discussed above, Drell supports both audio and video communications and Knappe's gateway can also communicate both audio and video. *See, e.g.* Ex_1006, 3:65-4:1 ("gateway…capable of communicating real-time audio, video and/or other information"). A POSITA looking to implement the teachings of Knappe and Drell, would have recognized that when the gateway is processing both audio and video signals, it would be beneficial to have a PSTN conference endpoint that can receive such signals. Ex_1003 ¶170.

Voois supplements the combination of Drell and Knappe by describing "a videophone" that communicates over "a public switched telephone network (PSTN)." Ex_1036, 3:62-4:4; Ex_1003 ¶171. Voois' videophone is capable of "transmitting (and receiving) audio and video data." Ex_1036, 4:32-39. A POSITA would have recognized that Voois's PSTN videophone teachings are well suited for Drell's system (as modified by Knappe to include a gateway capable of

communicating both audio and video) because it would allow for the user of the PSTN conference endpoint to participate in a conference with both audio and video. Ex_1003 ¶171. This would allow more information to and by the user of the PSTN conference endpoint. *Id*. Thus, it would have been obvious to a POSITA to implement Drell's PSTN conference endpoint as a videophone, per Voois, because it would beneficially allow for the user of the PSTN conference endpoint to exchange both audio and video data with the other conference participants, thereby improving the experience of the user and enhancing the effectiveness of collaboration. *Id*.

Such a combination simply combines prior art elements (videophone taught by Voois with the conferencing system of Drell as modified by Knappe) according to known methods (disclosed by Voois) to yield predictable results (allow for PSTN conference endpoint participants in Drell's system to share both audio and video data during a conference for collaboration purposes). *Id.* ¶172. Additionally, the combination of Drell and Voois is a simple substitution of one known element (PSTN videophone taught by Voois) for another (the PSTN conference endpoint of Drell). *Id*.

The results would have been predictable and there would have been a reasonable expectation of success in the combination given that both Drell and Voois describe similar conferencing systems. *Id*. ¶173. Petitioner's proposed

combination permits, but does not require, physical incorporation of elements from Drell and Voois. *Id*.; *see In re Mouttet*, 686 F.3d at 1332.

3. Claim 1

[1.P] A system for supporting a multi-participant conference call comprising

Drell discloses a multiparticipant conferencing system, specifically a "videoconferencing apparatus for use with multi-point conferences." Ex_1008, 1:16-18; Ex_1003 ¶175.

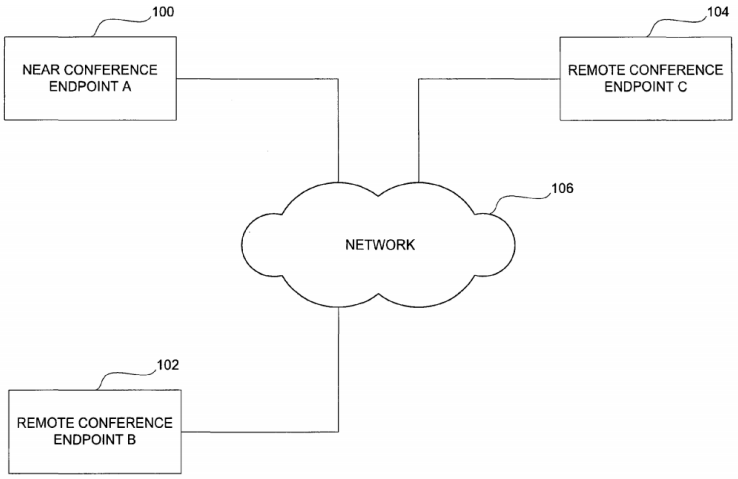


FIG. 1

Ex_1008, FIG. 1; Ex_1003 ¶175.

Drell discloses a 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 users. Ex_1003 ¶176. Specifically, Drell is directed to a "multi-point conferencing application" embodied in a near conference endpoint 100 that "is coupled to remote conference endpoints 102 and 104 via a network 106." Ex_1008, 3:4-8. Drell teaches that

"[r]emote 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." *Id.,* 3:8-11. Thus, a POSITA would have understood that Drell's system uses the Internet and PSTN to facilitate video/audio communications during a conference call between near and remote conference endpoints, including Internet and PSTN-based users. Ex_1003 ¶¶176-177.

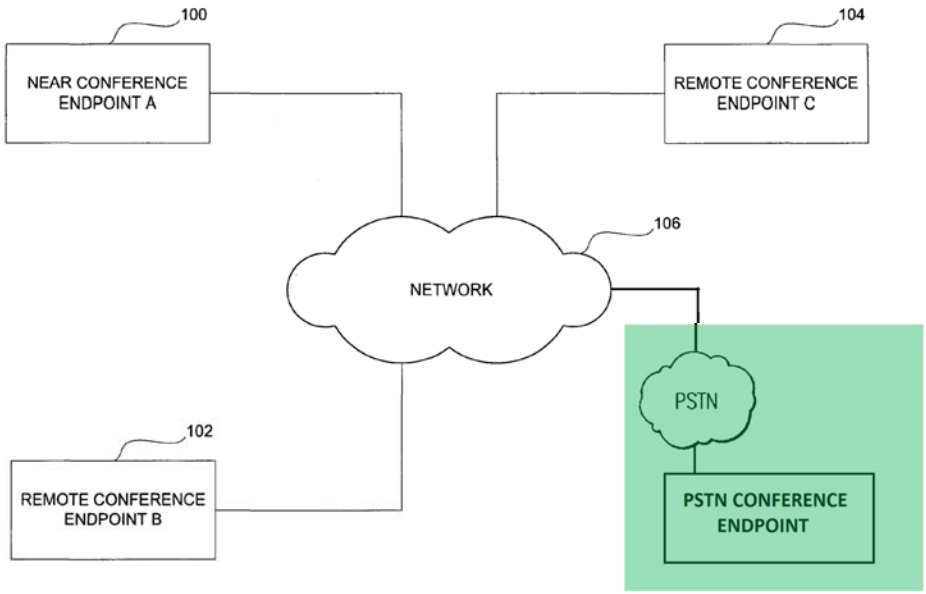
Drell further teaches a "call manager application" that "controls the establishment and termination of connections" among the conference endpoints. Ex_1008, 4:11-20. Such connections initiate a "multi-point conference." *Id*., 3:26-30; Ex_1003 ¶178. Thus, a POSITA would understand that Drell's multi-point conference is a "*conference call*" because it is established by a "call manager" and allows for audio (or audio and video) communication among the participants. Ex_1003 ¶¶178-179. Thus, Drell discloses "*[a] system for supporting a multi-participant conference call*." *Id*. ¶¶174-179.

[1.1] a first mixer that mixes Public Switched Telephone Network (PSTN) client audio data stream with a moderator audio-video data stream into a first mixed data stream

The prior art renders obvious this limitation. Ex_1003 ¶¶180-212.

First, Drell teaches "the transmission of audio and video data" over "two or more networks (for example, the PSTN and the Internet) [that] connect conference

endpoints." Ex_1008, 3:20-37; Ex_1003 ¶181. Thus, a POSITA would have understood that in Drell a remote PSTN conference endpoint (*i.e.* "*a PSTN client*") can connect to Drell's videoconferencing system via the PSTN, as illustrated below:

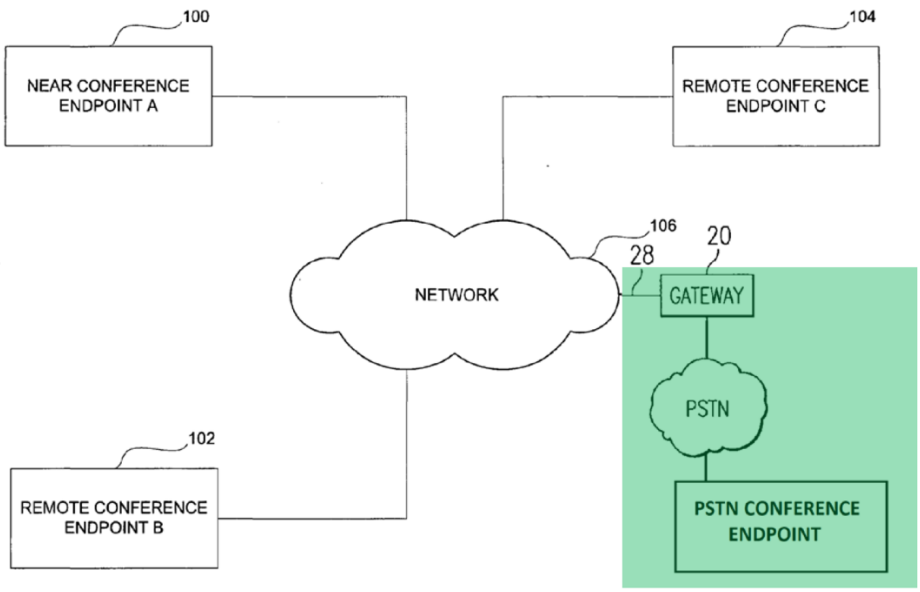


Id., FIG. 1 (annotated); Ex_1003 ¶¶181-182. Drell further discloses that the communication between conference endpoints (including from the PSTN endpoint) includes audio streams, and thus discloses a "*PSTN client audio data stream*." Ex_1008, 4:25-26 ("The processing trains process **audio** and video **data streams** received from remote conferencing endpoints 102 and 104."); Ex_1003 ¶¶181, 184.

Knappe supplements Drell and similarly describes a conferencing system with a PSTN telephone device as a participant where data to and from the PSTN

telephone is communicated through a gateway, noting that "the telephone 18 and gateway 20 combination [is] capable of communicating real-time audio, video and/or other information[.]" Ex_1006, 3:65-4:1; Ex_1003 ¶183. A POSITA would understand that a PSTN-connected telephone is a "*PSTN client*" as claimed because it receives conferencing services via the PSTN. *Id*. ¶183. Knappe further explains that audio from the conference participants is communicated as a data stream. Ex_1006, 5:46-47 (describing "multiparty communications sessions including real-time audio streams"), 2:1-6 ("receiving a plurality of participant voice streams from a plurality of conference participants"); Ex_1003 ¶¶183-184. It would have been obvious to a POSITA to use in Drell's system a gateway (as in Knappe) that converts standard audio data from PSTN conference endpoints into a streaming audio format suitable for transmission over the Internet. Ex_1003 ¶¶183-185; s*ee also id*. ¶¶ 148-168; Ex_1006, 4:18-25 (describing conversion of standard telephony signals into IP packets at the gateway).

As illustrated below, it would have been obvious to a POSITA that Drell's PSTN conference endpoint would transmit an audio stream through a gateway, like discussed in Knappe:



Ex_1008, FIG. 1 (annotated): Ex_1003 ¶185. Thus, the prior art teaches that near conference endpoint 100 receives an audio data stream from a PSTN conference endpoint through a PSTN gateway, rendering obvious "*a Public Switched Telephone Network (PSTN) client audio data stream*." Ex_1003 ¶186.

As Knappe teaches that at least one of the conference participants could be designated as a moderator (Ex_1006, 8:52-56), it would have been obvious to a POSITA for Drell's remote conference endpoint 104 to be a "moderator" to control the flow and organization of the conference. Ex_1003 ¶¶187-189; *see also id*. ¶¶148-168. Drell teaches that remote conference endpoint 104 can "transmit and receive both video (image) data and audio (speech) data." Ex_1008, 3:4-12; Ex_1003 ¶190. Drell's near conference endpoint 100 includes a multi-point (MP) "conferencing application" that processes "signal streams from/to each of the remote conference endpoints," and "separates the signal stream into audio and

video signals." Ex_1008, 2:17-33. Thus, Drell teaches receiving from remote conference endpoint 104 a signal stream including both audio and video data. Ex_1003 ¶¶191-193. Since it would have been obvious to a POSITA for remote conference endpoint 104 (which transmits audio and video data) to be used by a moderator (per Knappe), the combination of Drell and Knappe render obvious a "*moderator audio-video data stream*." *Id*. ¶194.

The combination of Drell and Knappe goes on to teach "*a first mixer*" that mixes the "*PSTN client audio data stream*" with the "*moderator audio-video data stream*" into a "*first mixed data stream*." Ex_1003 ¶¶195-212. As discussed in Section VI.D., "*first mixed data stream*" means a "*first data stream including at least audio data and video data*." Drell's MP conferencing application includes an "[a]udio mixing module 308 [] configured to combine audio data received from remote conference endpoints 102 and 104[.]" Ex_1008, 5:8-10; Ex_1003 ¶196. The audio mixing module will "mix the received audio input data streams" and produce multiple "output audio data streams." Ex_1008, 5:13-20. A POSITA would understand that when Drell's conference system includes a PSTN conference endpoint, the audio data received from the PSTN conference endpoint would be combined in the audio mixing module 308. Ex_1003 ¶¶196-197.

Drell's MP conferencing application also includes a video switching/continuous presence module 306 that "combines video data received

from remote conference endpoints 102 and 104." Ex_1008, 5:24-47; Ex_1003 ¶198. The video switching/continuous presence module may "select one of the video data input streams for output" and produces a "plurality of outputs." Ex_1008, 5:39-46.

Finally, Drell's MP conferencing application includes a "communication process 404" that receives mixed audio data from the audio mixing module and receives a video data stream output by the video switching/continuous presence module. *Id*, 6:33-45. The communication process 404 then "multiplexes the encoded audio and video data streams into a single audio/video data stream." *Id*., 6:46-48; Ex. 1003, ¶199. Drell also teaches that a separate "communication process 404" is instantiated for each remote conference endpoint. Ex_1008, 5:65-66; 4:21-23.

Thus, as shown below, Drell's audio mixing module 308, video switching/continuous presence module 306, and communication process 404 instantiated for a specific remote conference endpoint (*i.e.* the remote conference endpoint 102) collectively operate as a "*first mixer*" to mix the audio and video data streams received from multiple remote conference endpoints into a single audio and video data stream, *i.e.* the "*first mixed data stream*":

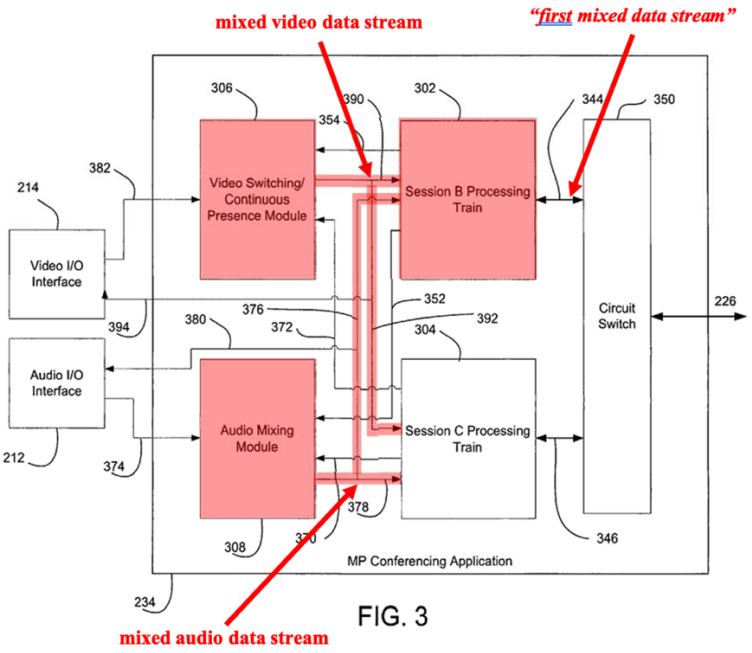


FIG. 3

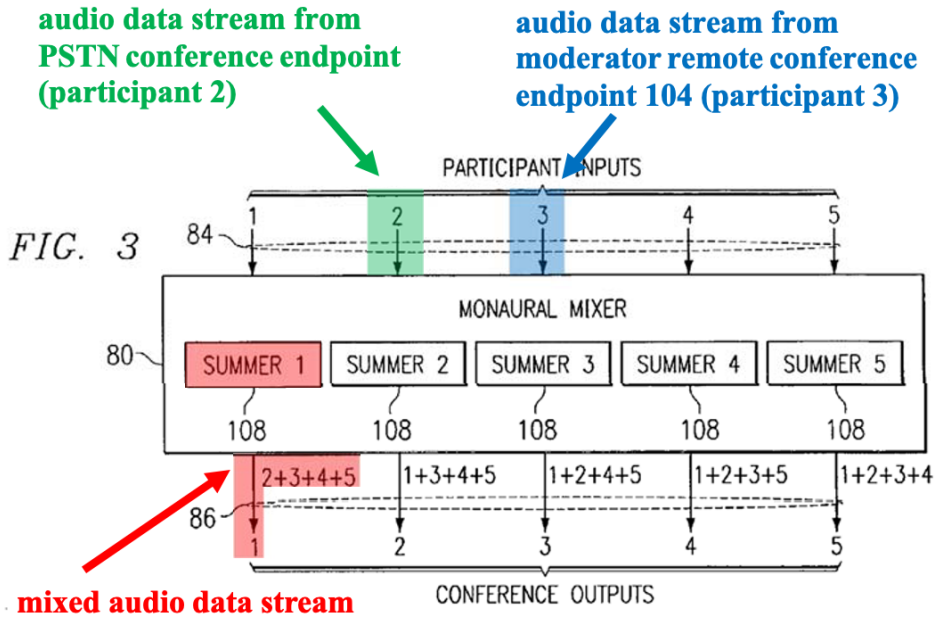
Ex_1008, FIG. 3 (annotated); Ex_1003 ¶¶200-202, 210-212.

Knappe supplements and explains how to mix audio data streams from multiple conference endpoints, including audio from PSTN conference endpoints. Ex_1006, FIGs. 1-3, 5:49-62, 6:59-7:15; Ex_1003 ¶204. Specifically, Figure 3 of Knappe shows "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." *Id.,* 6:59-7:3. The summers operate such that "each participant receives the audio input of each other participant," but not the participant's own audio. *Id*., 7:1-1

It would have been obvious to a POSITA to use in Drell's mixing module 308 a plurality of summers, as Knappe teaches. Ex_1003 ¶¶203-205. Using a separate summer for each participant would provide for separately mixing audio

data from the PSTN conference endpoint with audio data from the moderator remote conference endpoint 104, while providing higher prominence to the moderator audio and increasing the moderator's intelligibility. Ex_1003 ¶205. Additionally, providing a separate summer for each conference participant would reduce undesirable audio effects since each participant's mixed audio excludes its own audio data. *Id*. ¶205; *see also id*. ¶¶148-168. The use of a plurality of summers in Drell is consistent with Drell's desire to generate a separate instance for each remote endpoint and its description of generating a "plurality of output audio data streams." Ex_1008, 6:66-74, 5:13-17; Ex_1003 ¶206.

For explanation purposes, and without limiting Knappe's disclosure, below is annotated example of Figure 3 of Knappe, showing participant 2 as the PSTN conference endpoint and participant 3 as the moderator remote conference endpoint 104, with summer 1 outputting "mixed audio data stream" comprising audio from participants 2+3+4+5:



Ex. 1006, FIG. 3 (annotated); Ex_1003 ¶207.

A POSITA would understand that Drell's video switching/continuous presence module 306 would similarly include multiple separate video mixers to produce its "plurality of [video] outputs." Ex_1008, 5:43-44; Ex_1003 ¶208. For example, Knappe teaches that the "conference output stream for a participant includes the media of all **other** participants." Ex_1006, 5:35-37. Knappe's discussion of "media" includes "video" (*id*., 3:27), and thus Knappe suggests generating a mixed video stream output for a participant that includes the video streams "of all other participants. *Id*., 5:36-37: Ex_1003 ¶209.

Drell states that "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." Ex_1008,

5:39-42. Based on Knappe's teaching (and similar to the discussion of audio mixing above), it would have been obvious to a POSITA for Drell's video switching/continuous presence module 306 to include a separate video mixer so as to generate a separate "composite image" for each conference endpoint that shows only the video data streams from the *other* conference endpoints. Ex_1003 ¶210. By producing an individualized composite image in this way, the video display at each conference endpoint would show the video streams from the other conference endpoints without displaying its own locally produced video stream. *Id.* Since it is important for participants to see other's demonstrative gestures and behavioral cues, this approach would maximize the beneficial use of the available space on the videoconference display for displaying such gestures and cues. *Id*.

Thus, Drell in combination with Knappe renders obvious "*a first mixer that mixes a Public Switched Telephone Network (PSTN) client audio data stream with a moderator audio-video data stream into a first mixed data stream*." *Id*. ¶¶180-212.

[1.2.1] a first transport output that transmits the first mixed data stream to at least one remote client that receives the first mixed data stream

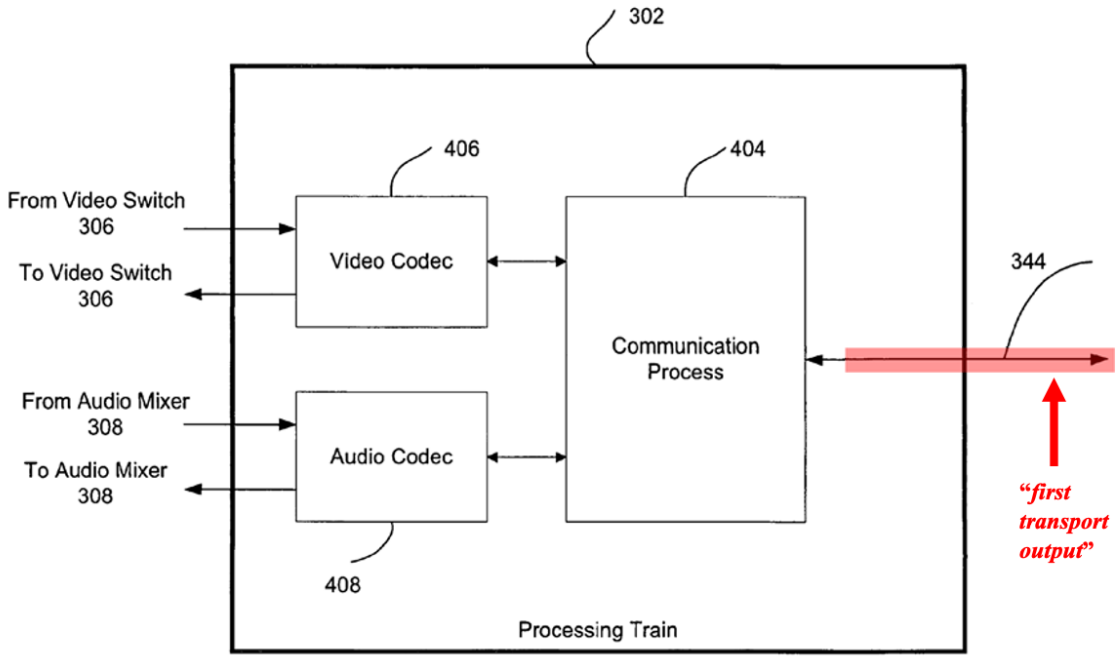
As discussed at [1.1], Drell and Knappe teach that communication process combines mixed audio and video data into a single audio/video data stream, *i.e.* "*first mixed data stream*." Ex_1008, 6:46-48. Drell further teaches that "combined audio and video streams are thereafter distributed to remote

conferencing endpoints 102 and 104." *Id*., 4:29-31. This includes "**transmitting** the combined audio and video signals to each of the plurality of remote conference endpoints." *Id*., 7:30-31; *see also id*. at 4:49-52; 6:33; 6:55-56. As discussed above in [1.1], Drell, alone or together with Knappe, renders obvious producing a separate mixed data stream for each remote conference endpoint. Ex_1003, ¶¶ 203-210, 215. Thus, it would have been obvious to a POSITA that distributing the combined audio and video streams would include "*transmit[ting]*" the "*first mixed data stream*" to a remote conferencing endpoint, such as remote conferencing endpoint 102. *Id.* ¶215.

It would have been obvious to a POSITA that the remote conferencing endpoint 102 then receives the "*first mixed data stream*" because receiving transmitted audio and video data is merely conventional videoconferencing device functionality. Ex_1008, 3:8-12 ("Remote conference endpoints 102 and 104...transmit and **receive both video (image) data and audio (speech) data.**"). Thus, the remote conferencing endpoint 102 is "*at least one remote client that receives the first mixed data stream*." Ex_1003 ¶215.

Drell teaches a separate communication process 404 (part of a "processing train") for each conference endpoint. Ex_1008, 4:21-24, 5:65-6:2; Ex_1003 ¶216. Drell's Figure 4 shows that the communication process 404 provides an output

(labeled 344) that conveys the "single audio/video data stream" (e.g., the "*first mixed data stream*"):



Ex_1008, FIG. 4 (annotated); Ex_1003 ¶217.

While Drell's Figure 1 shows that remote conference endpoint 102 is coupled to network 106 through an unlabeled connection, a POSITA would understand that network 106 would include a plurality of network communication devices, such as switches and routers, because that is how networks such as the Internet and PSTN are composed. Ex_1003 ¶¶217-218. It would have further been obvious to a POSITA that the line in Figure 1 that connects remote conference endpoint 102 to network 106 represents a connection from remote conference endpoint 102 to a router, switch, or other communication device within network 106. *Id.* ¶218. This router, switch, or other communication device would transmit

data to remote conference endpoint 102 that network 106 receives and that is destined for the remote conference endpoint 102. *Id.* The transmitted data would include the "*first mixed data stream*" transmitted by near conference endpoint 100 because Drell teaches that "combined audio and video signals" are transmitted to "each of the plurality of remote conference endpoints." *Id.*; Ex_1008, 7:30-31.

Accordingly, it would have been obvious to a POSITA that the output of network 106 leading toward remote conference endpoint 102 is a "*first transport output*." *Id.* ¶219. This mapping is illustrated below:

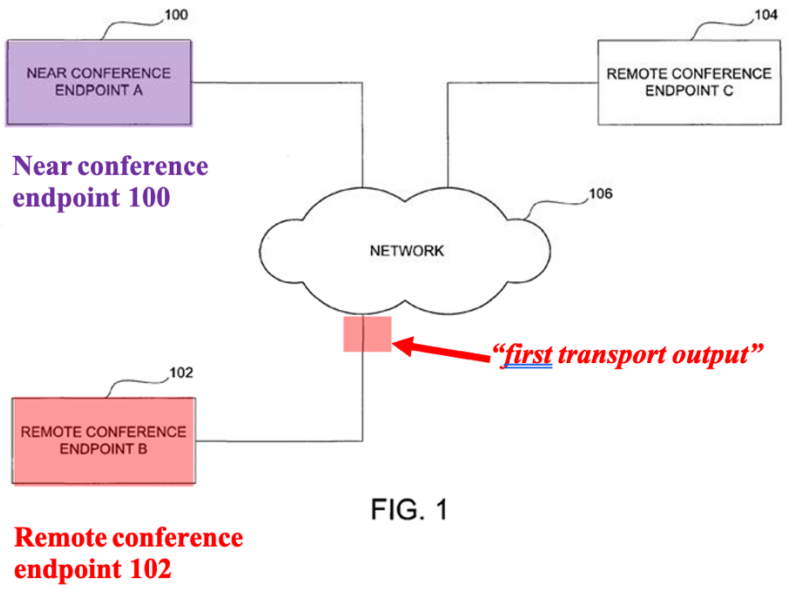


FIG. 1

Ex_1008, FIG. 1 (annotated); Ex_1003 ¶219. Thus, the prior art teaches that near conference endpoint 100 has a first transport output that transmits the first mixed data stream to remote conference endpoint 102 that receives the first mixed data stream, which renders obvious this limitation. Ex_1003 ¶220.

[1.2.2] the at least one remote client communicatively coupled to the Internet, which generates a remote client audio-video data stream;

As discussed at [1.2.1], Drell teaches a remote conference endpoint 102 that is "*at least one remote client*." Ex_1003 ¶221. Drell further shows remote conference endpoint 102 connected to network 106, which "may comprise a packet switched network, such as the **Internet**." Ex_1008, 3:33-34; Ex_1003 ¶223. Thus, a POSITA would understand that Drell's remote conference endpoint 102 is "*communicatively coupled to the Internet*," since it receives communications from the near conference endpoint over Internet network 106:

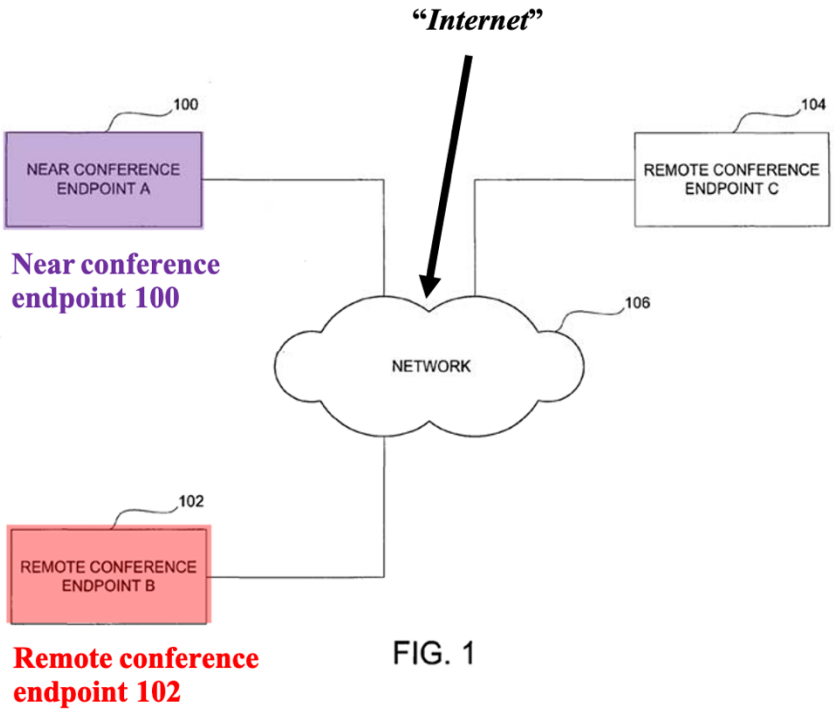


FIG. 1

Ex_1008, FIG. 1 (annotated), 3:4-19; Ex_1003 ¶¶223-224.

Drell teaches that the remote conference endpoint 102 is a "conventional videoconferencing device[] equipped to transmit…both video (image) data and audio (speech) data." Ex_1008, 3:10-12. Drell also describes other conference endpoints **"**conventional video camera 202 and microphone 204 are operative to generate video and audio signals representative of the images and speech of the near conference participant." *Id*., 3:39-45. A POSITA would understand that remote conference endpoint 102 could similarly include conventional equipment, such as a video camera and microphone, to generate video and audio signals of a conference participant because that is typical of videoconference endpoints. Ex_1003 ¶¶225-226.

Drell further describes processing "audio and video data streams received from remote conferencing endpoints 102 and 104," which arrive as a single "mixed data stream." Ex_1008, 4:21-31, 6:7-10; Ex_1003 ¶227. Because the audio and video information are received as a single stream, a POSITA would understand that remote conferencing endpoint 102 "*generates a remote client audio-video data stream*." *Id*. ¶227. Thus, Drell teaches that the remote conference endpoint 102 is communicatively coupled to the Internet network 106 and generates an audio-video data stream that is received by the near conference endpoint 100, which renders obvious this limitation. *Id.* ¶228.

[1.3] a second mixer that mixes the moderator audio-video data stream with the remote client audio-video data stream into a second mixed data stream; and

As discussed at [1.1], Drell and Knappe teach that the near conference endpoint 100 receives an audio-video data stream from the moderator remote conference endpoint 104. Ex_1003 ¶¶187-194, 230. Further, as discussed at [1.2.2], Drell teaches that the near conference endpoint 100 also receives a "*remote client audio-video data stream*" from the remote conference endpoint 102. *Id.* ¶¶223-227, 230.

As noted above, the phrase "second mixed data stream" is properly construed as a "*second data stream including at least audio data and video data*." *Id* ¶231. As discussed at [1.1], Drell discloses an audio mixing module 308, video switching/continuous presence module 306, and communication process 404 which collectively operate to mix audio and video data streams into a "*mixed data stream*." Ex_1003 ¶¶196-212, 232. Drell's mixing of the audio and video data streams received from remote conference endpoints 102 and 104 into a single audio and video data stream renders obvious a "*second mixed data stream*." Ex_1003 ¶¶232-233.

As discussed at [1.1], Drell teaches instantiating a separate communication process 404 for each remote conference endpoint. Ex_1008, 5:65-66, 4:21-23; Ex_1003 ¶¶ 203-211, 234. Because the "*second mixed data stream*" relates to the

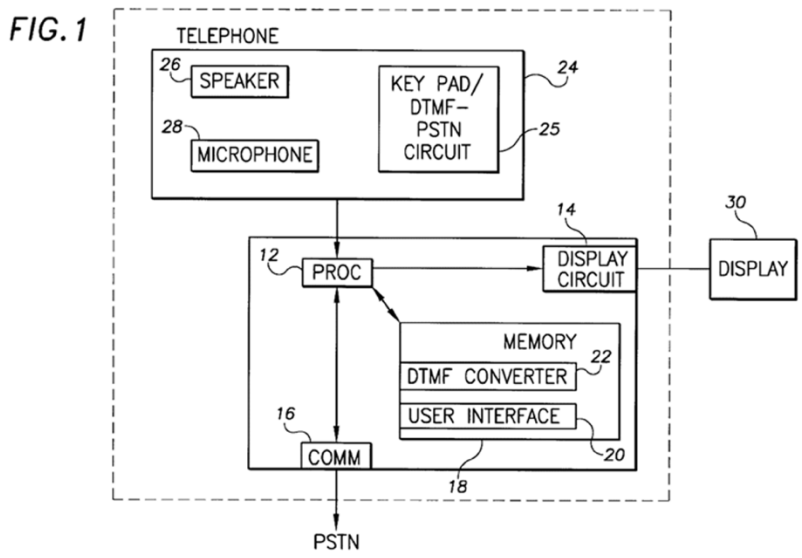
audio data stream transmitted to the "*PSTN client*" (*see infra*, [1.4]), it would have been obvious to a POSITA for the communication process 404 instantiated for the PSTN conference endpoint ("*PSTN client*," *see* [1.1]) to produce the "*second mixed data stream*." Ex_1003 ¶234. It would have been obvious to a POSITA, therefore, that Drell's audio mixing module 308, video switching/continuous presence module 306, and communication process 404 instantiated for the PSTN conference endpoint collectively teach the claimed "*second mixer*." *Id.*

Also, for reasons similar to those discussed at [1.1], it would have been obvious to a POSITA to implement audio mixing module 308 and video switching/continuous presence module 306 such that they separately mix audio and video for each conference endpoint. *Id.* ¶¶203-211, 235. Thus, it would have also been obvious to a POSITA to employ a separate audio mixer, a separate video mixer, and the communication process 404 instantiated for the PSTN conference endpoint, which collectively constitute the claimed "*second mixer*." *Id*. ¶235.

As discussed above regarding the "*second mixer*," claim 1 subsequently recites that audio corresponding to the "*second mixed data stream*" is transmitted to the "*PSTN client*." It would have been obvious to a POSITA to produce a "*second mixed data stream*" including both audio and video data for the "*PSTN client*" because it would have been obvious for the "*PSTN client*" to support both audio and video conferencing. *Id*. ¶236. Specifically, Drell explains that the near

conference endpoint 100 can "connect to conference endpoints" over "the PSTN." Ex_1008, 3:20-38. Knappe complements Drell by describing a "telephone 18 and gateway 20 combination" that connects to a videoconference system via the PSTN. Ex_1006, 3:66-67; Ex_1003 ¶237. The telephone and gateway are "capable of communicating real-time **audio, video** and/or other information." Ex_1006, 3:66-4:1. It would have been obvious to a POSITA, therefore, for the PSTN conference endpoint and gateway to support communicating both audio and video data. Ex_1003 ¶237. That the PSTN conference endpoint supports both audio and video is consistent with Claim 6, which recites "*transmitting the second mixed data stream to the PSTN client*." *Id*.

Voois supplements Drell and Knappe by describing a PSTN conference endpoint that is a videophone that "communicates video and audio data over a plain old telephone service (POTS) line." Ex_1036, Abstract. Voois' videophone "interface[s] with a communication network, such as a public switched telephone network (PSTN)[.]" *Id*., 3:62-4:4. The general organization of Voois' videophone, including its video display and connection to the PSTN, is shown in Figure 1:



Id., FIG. 1. It would have been obvious to a POSITA for Drell's PSTN conference endpoint to be a videophone, as taught by Voois, because it would beneficially allow for the user of the PSTN conference endpoint (implemented as a videophone) to exchange audio and video with the other conference participants, thereby improving the experience of the user. Ex_1003 ¶¶238-240.

Thus, with Drell's video conference system supporting a videoconferencing PSTN conference endpoint (as suggested by Knappe and Voois), it would have been obvious to a POSITA that Drell's MP conferencing application (and specifically, an instance of communication process 404) would produce a mixed audio-and-video data stream for the videoconferencing PSTN conference endpoint. *Id.* ¶241. This approach would have been obvious because it would allow for providing the same rich, multimedia conferencing experience to PSTN-based

participants as is provided to participants using other forms of videoconferencing equipment. *Id.* ¶¶242-243.

[1.4] a second transport output that transmits a mixed audio data stream, corresponding to the second mixed data stream, to the PSTN client.

As discussed at [1.3], Drell, Knappe, and Voois teach that communication process 404 combines mixed audio and video data into a "single audio/video data stream" corresponding to the claimed "*second mixed data stream*." Ex_1008, 6:46-48; Ex_1003 ¶¶233-234, 244. Further, as discussed at [1.1] and [1.3], Drell, Knappe and Voois teach a PSTN conference endpoint (implemented as a videophone) is the "*PSTN client*." Ex_1003 ¶¶181-183, 236-242, 244.

Drell teaches "**transmitting** the combined audio and video signals to each of the plurality of remote conference endpoints." Ex_1008, 7:30-31. As discussed above in [1.1], Drell, alone or together with Knappe, renders obvious producing a separate mixed data stream for each remote conference endpoint. Ex_1003 ¶¶203-211, 245. Thus, it would have been obvious to a POSITA to "*transmit[]*" the "*second mixed data stream*" to a remote conferencing endpoint, such as the videophone PSTN conference endpoint ("*PSTN client*") taught by Drell and Voois. *Id.* ¶245. As discussed in [1.3], the "*second mixed data stream*" includes the mixed audio data stream produced by Drell's audio mixing module 308. *Id*. ¶¶232-242, 245. Because the mixed audio data stream is part of the "*second mixed data stream*," the mixed audio data stream corresponds to it. *Id.* ¶245. In other

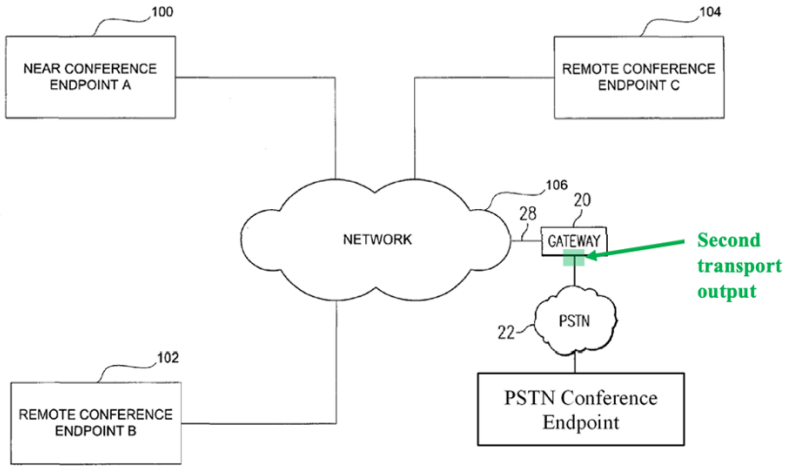
words, the "*mixed audio data stream*" is the audio portion of the "*second mixed data stream*." *Id.* Transmitting the "*second mixed data stream*" therefore includes transmitting a "*mixed audio data stream, corresponding to the second mixed data stream*." *Id*. ¶¶245-246. Thus, Drell in combination with Knappe and Voois teaches transmitting a mixed audio-video data stream to the PSTN conference endpoint (implemented as a videophone), which renders obvious to "*transmit[] a mixed audio data stream, corresponding to the second mixed data stream, to the PSTN client*." *Id*. ¶246.

Furthermore, as discussed at [1.2.1], the output of Drell's communication process 404 is a "*transport output*." *Id*. ¶¶213-220, 247. Drell further teaches instantiating a separate communication process 404 (part of a "processing train") for each conference endpoint. Ex_1008, 4:21-24; 5:65-6:2. Thus, the output of communication process 404 instantiated for the PSTN conference endpoint is a "*second transport output*." Ex_1003 ¶247.

Additionally, as discussed at [1.1], it would have been obvious to a POSITA to utilize a gateway, as Knappe teaches, in Drell's system. *Id*. ¶¶183, 248. Specifically, Knappe teaches utilizing a gateway positioned between an Internet network and a PSTN for converting data into a suitable format for each network. Ex_1006, 4:18-25, FIG. 1; Ex_1003 ¶248. Since Drell contemplates that its conference endpoints communicate over "PSTN and the Internet" networks

(Ex_1008, 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 Internet network communications to suitable signals for the PSTN, and vice-versa (Ex_1006, at 4:18-25). Ex_1003 ¶¶148-168, 249.

As shown below (combining portions of Drell's Figure 1 and Knappe's Figure 1), the gateway 20 has an output that couples it to a PSTN conference endpoint. Ex_1003 ¶250. It would have been obvious to a POSITA that the gateway would transmit at least the audio portion of the "*second mixed data stream*" to the PSTN conference endpoint ("*PSTN client*") so that the conference participant using the PSTN conference endpoint could hear the audio portion of the conference (that is, the speech of other conference participants).



Ex_1008, FIG. 1 (annotated); Ex_1003 ¶250.

Moreover, in the combination with Voois, it would have been obvious to a POSITA for the gateway (which is "capable of communicating real-time audio [and] video," Ex_1006, 3:65-41) to also transmit the video portion of the second mixed data stream to the PSTN conference endpoint (implemented as a videophone). Ex_1003 ¶251. Thus, the prior art teaches that a transport output of a PSTN gateway transmits mixed audio and video data (corresponding to the "*second mixed data stream*") to a PSTN conference endpoint, which renders obvious this element. *Id*. ¶252.

4. Claim 2

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

See [1.P]-[1.4].

[2.1.1] a third mixer that mixes the audio data from the PSTN client with the audio-video data stream from the remote client into a third mixed data stream and

As discussed at [1.1], Drell and Knappe teach that the near conference endpoint 100 receives audio data stream from the PSTN conference endpoint. Ex_1003 ¶¶184-194, 255. Further, as discussed at [1.2.2], Drell teaches that the near conference endpoint 100 also receives an audio-video data stream from the remote conference endpoint 102. *Id.* ¶¶223-228, 255.

The phrase "*third mixed data stream*" means a *third data stream including at least audio data and video data*. *Id*. ¶256. As discussed at [1.1], Drell discloses an audio mixing module 308, video switching/continuous presence module 306, and

communication process 404 which collectively operate to mix audio and video data streams into a "*mixed data stream*." *Id*. ¶¶196-212, 257. Drell's mixing of the data streams received from the PSTN conference endpoint and remote conference endpoint 102 into a single audio and video data stream for the moderator remote conference endpoint 104 renders obvious a "*third mixed data stream*." *Id*. ¶258.

As discussed at [1.1], Drell teaches instantiating a separate communication process 404 for each remote conference endpoint. Because the "*third mixed data stream*" is communicated to the moderator remote conference endpoint 104 (*see infra*, [2.1.2]), it would have been obvious to a POSITA for the communication process 404 instantiated for remote conference endpoint 104 (used by the "*moderator*," *see* [1.1]) to produce the "*third mixed data stream*." *Id*. ¶259. It would have been obvious to a POSITA, therefore, that Drell's audio mixing module 308, and video switching/continuous presence module 306, and the communication process 404 (of the processing train 304) instantiated for the moderator remote conference endpoint 104 collectively teach the claimed "*third mixer*." *Id*.

For reasons analogous to those discussed at [1.1], it would have been obvious to a POSITA to implement audio mixing module 308 and video switching/continuous presence module 306 such that they separately mix audio and video for each conference endpoint. *Id.* ¶260. Thus, it would have also been

obvious to a POSITA that the separate audio mixer, the separate video mixer, and the communication process 404 instantiated for remote conference endpoint 104 collectively constitute the claimed "*third mixer.*" *Id*. ¶¶260-261. Thus the prior art renders obvious this limitation. *Id*. ¶261.

[2.1.2] communicates the third mixed data stream to the moderator

As discussed at [1.2.1], Drell teaches "transmitting the combined audio and video signals to each of the plurality of remote conference endpoints." Ex_1008, 7:30-31, *see also* 4:29-31. Drell explains that each communication process 404 (part of the "*third mixer*," *see* [2.1.1]) is "for sending… audio and video data to… a single remote endpoint." *Id*., 8:21-23; *see also* 6:46-52; Ex_1003 ¶¶264-265. Thus, it would have been obvious to a POSITA that communication process 404 instantiated for the remote conferencing endpoint 104 "*communicates*" the "*third mixed data stream*" to the remote conferencing endpoint 104 used by the "*moderator*" (*see* [1.1]), which renders obvious this limitation. Ex_1003 ¶¶264-267.

5. Claim 3

[3.P] The system of claim 1 further comprising:

See [1.P]-[1.4]

[3.1] a VoIP decoder that receives IP packets including the audio data stream generated by the PSTN client, the VoIP decoder decodes the IP packets including the audio data from the PSTN client.

As discussed at [1.1], Drell and Knappe teach that the near conference endpoint 100 receives an audio data stream from the PSTN conference endpoint, via a gateway. Ex_1003 ¶¶184-186, 270. It would have been obvious to a POSITA that near conference endpoint 100 "*receives IP packets including the audio data stream generated by the PSTN client*" since the conference endpoint 100 is "connect[ed to other] conference endpoints" over the "PSTN and the Internet" because communications over the Internet use IP packets. *See, e.g.,* Ex_1006, 3:53-54; Ex_1003 ¶¶271-272.

As already discussed above, it would have been obvious to use a gateway as taught in Knappe. Knappe explains that "[a]t the gateway 20, the signals are converted to IP packets in the VoIP format" and "encapsulated in real-time transport protocol (RTP) packets" for transmission over the Internet. Ex_1008, 4:18-25, 3:50-61; Ex_1003 ¶273. Thus, it would have been obvious that in the combination with Knappe, the near conference endpoint 100 would "*receive[] IP packets including the audio data stream generated by the PSTN client*." Ex_1003 ¶¶275-276.

Drell also teaches that "[a]udio codec 408 receives the audio data stream from communication process 404 and…decod[es] in accordance with a standard… or proprietary audio compression algorithm" and produces a "decoded audio stream." Ex_1008, 6:18-23. Accordingly, Drell's disclosure of a communication

process and an audio codec renders obvious "*a VoIP decoder*." Ex_1003 ¶¶277-279. Drell also teaches that its communication process includes a "H.323 (packet-based) process." Ex_1008, 2:17-33. It would have been obvious to a POSITA that Drell's H.323 process includes a "*VoIP decoder*" that "*decodes*" because H.323 was a well- known standard that specifies encoding and decoding VoIP communication. Ex_1003 ¶¶280-281

Knappe also teaches a "codec (coder/decoder)" for processing "IP packets such as VoIP packets…encapsulated in real-time transport protocol (RTP) packets." Ex_1006, 4:9-14, 3:50-57. As Knappe's gateway converts data between PSTN and VoIP format), it would have been obvious to a POSITA to use a corresponding audio codec configured to decode received VoIP packets. Ex_1003 ¶¶282-283. Thus, Knappe's codec also renders obvious a "*VoIP decoder*." *Id*. ¶283.

Thus, Drell and Knappe teach that the near conference endpoint 100 includes a VoIP decoder that receives and decodes VoIP packets containing audio data from the PSTN conference endpoint, which renders obvious this limitation. *Id*. ¶¶284-285.

6. Claim 4

[4.P] The system of claim 1 further comprising:

See [1.P]-[1.4]

[4.1] at least one audio decompressor that decodes IP packets including the audio-video data stream generated by the remote client.

As discussed at [1.2.1], and [1.3], Drell and Knappe teach that the near conference endpoint 100 receives an audio-video data stream from the remote conference endpoint 102. Ex_1003 ¶288. Since the audio-video data stream is received over the Internet network 106, it would have been obvious to a POSITA that it is received as "*IP packets*." *Id*. ¶¶289-291.

And as discussed in [3.1], Drell teaches "an audio codec 408" that generates a "decoded audio data stream" using a standard or proprietary compression algorithm. Ex_1008, 6:18-23. A POSITA would have understood that decoding audio using an audio compression algorithm is performing decompression. Ex_1003 ¶¶292-293. Drell's audio codec that utilizes an audio compression algorithm to decompress IP packets containing audio- video data stream from remote conference endpoint 102 discloses an "*audio decompressor.*" *Id*. Also, it would have been obvious to a POSITA that Drell's H.323 process (as discussed in [3.1]) includes an "*audio decompressor*" because H.323 was a well-known standard that specified certain compression protocols. *Id*. ¶¶294-295.

Thus, the prior art teaches that near conference endpoint 100 includes audio decompression algorithm or protocol that decodes IP packets including the audio-video data stream generated by the remote conference endpoint 102, which renders obvious this limitation. *Id*. ¶296.

B. Ground 2: Claims 5-8 Are Obvious Over Drell in Combination with Knappe, Voois, Elliott, and the VPN Textbook

1. The Prior Art

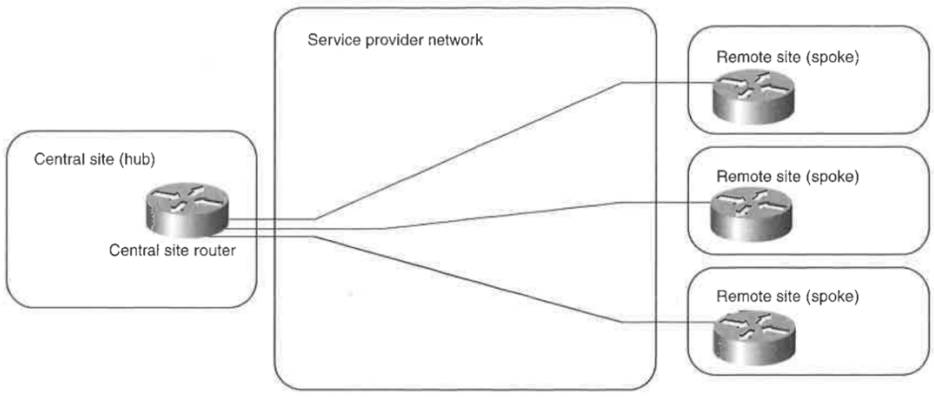
a) Summary of Elliott

Elliott teaches conferencing features such as "collaborative Web browsing, audio conferencing, video conferencing, and application sharing." Ex_1007, 1:40-44. Elliott employs a "virtual private network (VPN)" and "IP-tunneling" for the conference transmissions. *Id*., 2:66-3:1, 4:29-34, 6:53-54; Ex. 1003, ¶¶297-298.

b) Summary of VPN Textbook

The VPN Textbook describes the commonly encountered "hub-and-spoke topology, where a number of remote offices (spokes) are connected to a central site (hub)." Ex. 1010 at 129. The hub-and-spoke topology (below) is comonly referred to as a "star" topology. Ex_1003 ¶300.

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



Ex_1010 at 130; Ex_1003 ¶¶301-302.

2. Reasons to Combine Prior Art

a) Reasons to Combine Drell and Elliott

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 VPN tunneling within Drell's system. Ex_1003 ¶303.

A POSITA, when considering the teachings of Drell would have also considered the teachings of Elliott because they are analogous art pertaining to multiparty conferencing between computer and PSTN participants. Ex_1008, Abstract, 3:33-38; Ex. 1007, Abstract, 5:35-39, 6:1-1, 8:19-39, FIG. 1; Ex_1003 ¶¶304-305. Drell 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." Ex_1008, 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. Ex_1003 ¶311. It was also known that communications over a public network could be secured using a VPN. *Id.* 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. Ex_1008, 1:20-25; Ex_1003 ¶¶311-312.

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." Ex_1007, 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. Ex_1003 ¶313. Other prior art artisans were using similar combinations prior to the priority date of the '525 patent. *See id*. 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. *Id*. 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. *Id*. ¶314.

The results of combining Elliott's VPN tunneling with Drell's system would have been predictable and there would have been a reasonable expectation of

success in the combination given that both references are directed to conferencing systems. Ex_1003 ¶316. This is especially so since VPN tunneling was conventional. *Id.* 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. *Id*. Petitioner's proposed combinations permit but do not require physical incorporation of elements from Elliott into Drell. *See In re Mouttet*, 686 F.3d 1322; Ex. 1003 ¶315.

b) Reasons to Combine Drell and VPN Textbook

A POSITA would have been motivated to combine the teachings of Drell and the VPN Textbook for multiple reasons, including to produce the obvious, beneficial, and predictable result of providing a hub-and-spoke VPN topology when implementing VPN tunneling in Drell (per Elliott), and to provide a cost effective and less complex VPN implementation. Ex_1003 ¶318.

A POSITA would have been familiar with conventional VPN tunneling implementations suggested by Elliott, and that such VPN implementations were well documented in readily available textbooks. *Id.* ¶319. Accordingly, a POSITA considering the combination of Drell and Elliott, which employs VPN tunneling, would have found it obvious to consider other references, such as the VPN Textbook, that provide more details and background information on VPN techniques. *Id.*

The VPN Textbook descibes various VPN topologies. *Id.* ¶320. A POSITA looking to implement Drell's system with VPN tunneling would have needed to choose a topology, and the VPN Textbook explains that the most common topology is the hub-and-spoke topology. Ex_1010 at 129 ("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)"). The hub-and-spoke VPN topology would have been an obvious choice for the POSITA since it is the most commonly used. Ex_1003 ¶321. This would be simply combining prior art elements (the VPN topology teachings of the VPN Textbook and the system of Drell) according to known methods (as taught by the VPN Textbook) to yield predictable results (securing the traffic exchanged in Drell's system during conferencing). *Id*. In the combination, Drell's system would utilize corresponding hardware and software (e.g., routers configured in a hub-and-spoke topology running VPN software) that would perform the routing and VPN tunneling functions described in the VPN textbook. *Id.* Such products were widely commercially available in the prior art time period. As confirmed by Dr. Houh, he personally "purchased VPN hardware routers for a work project during the 1998-1999 time frame." *Id*.

Using a hub-and-spoke VPN topology, as taught by the VPN Textbook in the system of Drell (as modified by Elliott) is simply the use of a known technique

(utilizing hub-and-spoke VPN topology) in Drell's system to yield predictable results (provide VPN tunnels to Drell's system for improved security). *Id.* ¶322. A POSITA would have been readily able to predict the result of such a combination, namely, that Drell's system would utilize a hub-and-spoke VPN topology to allow the conferencing devices to communicate with each other over secure VPN tunnels. *Id.* Because hub-and-spoke VPN topology was conventional and the most commonly used topology, making such a combination would have been within the POSITA's skill level. *Id*.

Additionally, a POSITA would have been motivated to combine the teachings of Drell (as modified by Elliott) with the teachings of the VPN Textbook because it would beneficially constrain cost and limit complexity. *Id.* ¶323. The VPN Textbook explains that the hub-and-spoke VPN topology serves to constrain cost and complexity. Ex_1010 at 129 ("[T]here 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.").

Moreover, the result would have been predictable and there would have been a reasonable expectation of success since VPN technology was conventional and known to be applied in the conferencing context. *Id.* ¶324. Any modification needed to the teachings of Drell, such as adding software programming and/or hardware, to accommodate the teachings of the VPN Textbook would have been

within the level of ordinary skill in the art, as demonstrated by the examples of other artisans combining conferencing and VPN technologies. *Id*. Indeed, Dr. Houh "personally, before the earliest claimed priority date of the '525 patent, configured, deployed and utilized VPN technology for carrying VoIP traffic using the NBX 100 system, which was also capable of performing conference calls." *Id*.

Petitioner's proposed combination permits but does not require physical incorporation of elements from the VPN Textbook into Drell's system. *See In re Mouttet*, 686 F.3d at 1322. Ex_1003 ¶325.

3. Claim 5

[5.P] The system of claim 1 further comprising:

See [1.P]-[1.4].

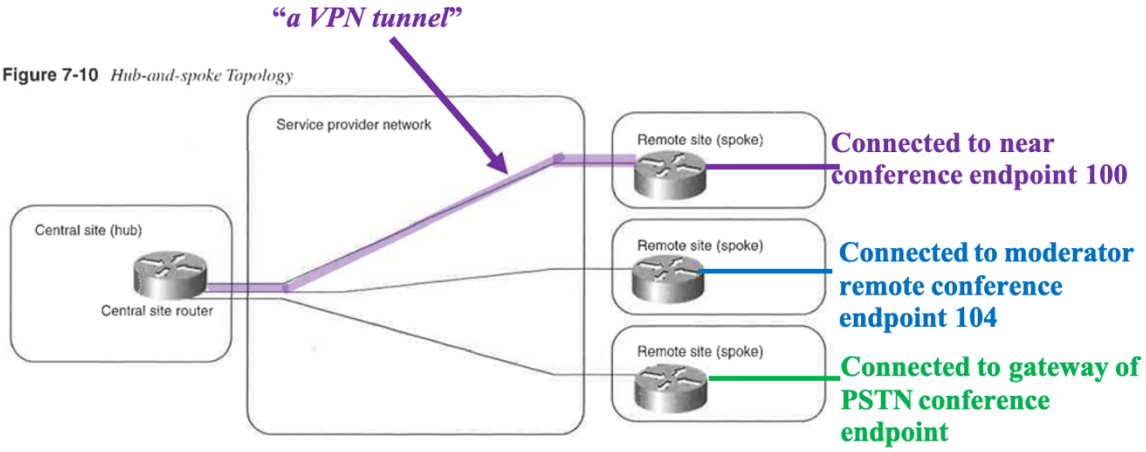
[5.1] wherein the PSTN client audio data stream and the moderator audio-video data stream are transported across a virtual private network tunnel

As discussed at [1.1], Drell and Knappe teach that the near conference endpoint 100 receives an audio data stream from the PSTN conference endpoint through a gateway and an audio-video data stream from the moderator remote conference endpoint 104. Ex_1003 ¶¶183-185, 328.

Elliott teaches "multi-media collaboration" including "voice, fax, data, video, etc." transmitted over the Internet using a "virtual private network (VPN)…which uses IP or IP-tunneling." Ex_1007, 2:66-3:13, 4:6-12, 4:31-34. It would have been

obvious to a POSITA to utilize VPN tunneling when communicating audio and video data to and from the PSTN conference endpoint's gateway and the moderator remote conference endpoint 104, because it would provide Drell with secure communications as the communications traverse Internet network 106. Ex_1003 ¶¶303-317, 331.

The VPN Textbook further teaches that 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)." Ex_1010 at 129. When implementing VPN tunneling in Drell, per Elliott, it would have been obvious to a POSITA to use a hub-and-spoke configuration because it was the "most commonly encountered topology" often chosen for "cost or complexity reasons." *Id.*; Ex_1003 ¶¶318-325, 334. As illustrated below in annotated Figure 7-10 of the VPN Textbook, it would have been obvious to connect Drell's conference endpoints to the routers providing a VPN tunnel:



Ex_1010 at 130 (annotated); Ex_1003 ¶¶334-335.

As illustrated above, each conference endpoint is connected to a separate site router. Ex_1003 ¶¶335-336. When the near conference endpoint 100 transmits data to or receives data from the moderator remote conference endpoint 104 and the PSTN conference endpoint's gateway, the data would be received/sent through the same VPN tunnel, which corresponds to "*a VPN tunnel.*" *Id*. ¶336.

Thus, the prior art teaches that near conference point 100 receives the PSTN conference endpoint audio data stream and the moderator remote conference endpoint 104 audio-video data stream via a VPN tunnel, which renders obvious this limitation. *Id*. ¶337.

4. Claim 6

[6.P] The system of claim 5 further comprising:

See [5.P]-[5.1].

[6.1] a VoIP encoder that encodes the second mixed data stream and encodes the second mixed voice data before transmitting the second mixed data stream to the PSTN client

Neither claim 1 nor 5, from which claim 6 depends, recites the phrase a "*second mixed voice data*." For analysis purposes only, this Petition will assume that the term "*second mixed voice data*" recited in claim 6 corresponds to the "*mixed audio data stream*" recited in [1.4]. Ex_1003 ¶340.

As discussed at [1.2.1]-[1.3], Drell and Knappe teach mixing the audio-video data stream from the moderator remote conference endpoint 104 with the

audio-video data stream from the remote conference endpoint 102 into a "*second mixed data stream*." *Id.* ¶¶230-242, 341. Further, as discussed at [1.4], Drell, Knappe, and Voois teaches encoding and transmitting the second mixed data stream, including the mixed audio data stream, to the PSTN conference endpoint. *Id.* ¶¶243-252, 341.

Drell further explains that the "communication process…compris[es] an H.323 (packet-based) process." Ex_1008, 2:26-28. A POSITA would understand that Drell's H.323 process includes a "*VoIP encoder*" for encoding the audio data and the mixed data transmitted to the PSTN conference endpoint (as claimed) because H.323 was a well-known standard for encoding and decoding VoIP communication. Ex_1003 ¶347. Thus, a Drell's H.323-based communication process 404 renders obvious the claimed "*VoIP encoder*." *Id*.

Drell also teaches use of "audio codec 408" that "encodes the audio data stream" and "video codec 406" that "encode[s] video data" before transmission over the Internet. Ex_1008, 6:33-45, 5:8-25. Knappe supplements Drell and teaches "[r]eal-time IP packets such as VoIP packets are encapsulated in real-time transport protocol (RTP) packets for transmission" over the Internet. Ex_1006, 3:50-58. It would have been obvious to a POSITA to encapsulate the mixed audio and video data, *i.e.* "*the second mixed data stream*" 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. Ex_1003 ¶¶342-345. Thus, Drell's audio codec and communication process that encodes data for transmission over the Internet using RTP packets (per Knappe) teaches a "*VoIP encoder*" and renders obvious this limitation. *Id*. ¶¶346-349.

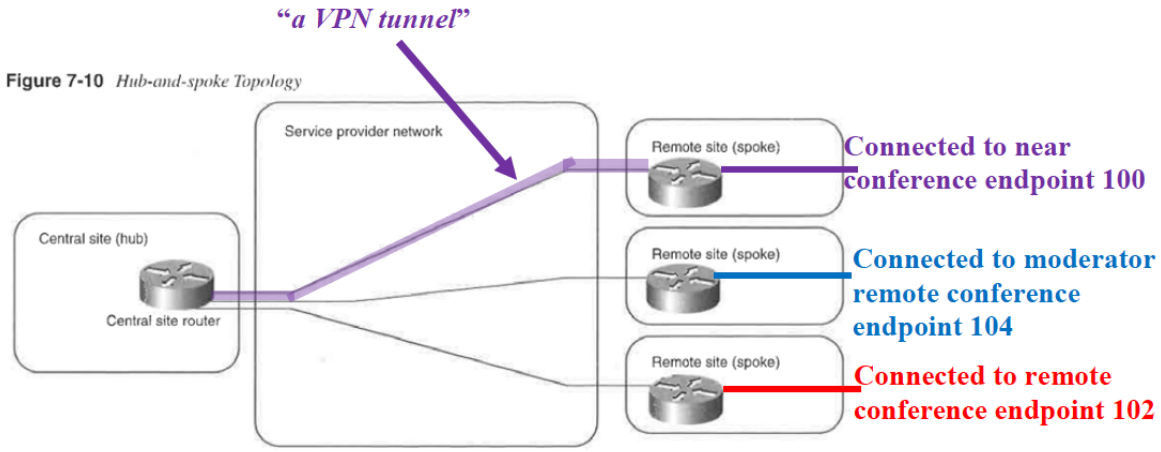
5. Claim 7

[7.P] The system of claim 1 further comprising:

See [1.P]-[1.4].

[7.1] wherein the moderator audio-video data stream and the remote client audio-video data stream are transported across a virtual private network tunnel

As discussed at [5.1], the prior art renders obvious using hub-and-spoke VPN tunneling to securely connect conference endpoints. Ex_1003 ¶¶330-337, 352-360. As shown in the figure below, it would have been obvious to a POSITA that Drell's near conference endpoint 100 (which performs audio and video mixing, as discussed *inter alia* in [1.1] and [1.3]) receives data transmitted by the moderator remote conference endpoint 104 and remote conference endpoint 102 via a "*VPN tunnel*":



Id. at 130 (annotated); Ex_1003 ¶¶356-361.

Thus, the prior art teaches that near conference endpoint 100 receives audio and video data streams from moderator remote conference 104 and the remote conference endpoint 102 via a VPN tunnel, which renders obvious this limitation. Ex_1003 ¶¶351-362.

6. Claim 8

[8.P] The system of claim 7 further comprising:

See [7.P]-[7.1].

[8.1] an encoder that receives the first mixed data stream and encodes the first mixed voice data before transmitting the first mixed data stream to the remote client.

Neither of claim 1 nor 7, from which claim 8 depends, recites the phrase a "*first mixed voice data*." For analysis purposes only, this Petition will assume that the term "*first mixed voice data*" recited in claim 8 corresponds to a mixed audio portion of the "*first mixed data stream*" recited in [1.1]. Ex_1003 ¶365.

As discussed in [6.1], Drell teaches an "H.323 (packet-based) process" that renders obvious an "*encoder.*" *See also id*. ¶372. As discussed in [1.1], Drell's H.323 process is part of "communication process 404" that "multiplexes the encoded audio and video data streams into a single audio/video data stream," *i.e.* the "*first mixed data stream.*" *See* Ex. 1003 ¶199; *see also* ¶¶181-212. For reasons analogous that discussed at [6.1], it would have been obvious to a POSITA that this single data stream (including the audio (voice) portion thereof) would then be encapsulated using real-time transport protocol (RTP), as taught by Knappe, because RTP "was created to transport real-time traffic, such as video and audio, over the Internet. *See id*. ¶370. Thus, Drell and Knappe teach a communication process ("*encoder*") that receives and encapsulates a mixed audio and video data stream to be transmitted to remote conference endpoint 102, which renders obvious this limitation. *Id*. ¶¶330, 360-371.

Additionally, as discussed in [7.1], the prior art renders obvious transmitting a mixed audio and video data stream to remote conference endpoint 102 via a VPN tunnel. *Id*. ¶375. It would have been obvious to a POSITA that the VPN transmission process would include receiving and encrypting this data stream (including the audio (voice) portion thereof) because encryption is generally used to provide security in a VPN. *Id.* This encryption step would be performed, for

example, by the remote site router described in the VPN Textbook. *Id*.; Ex_1010 at 119 (describing "encryption gear"), Fig. 7-3.

Thus, the prior art teaches a VPN router ("*encoder*") that receives and encrypts a mixed audio and video data stream to be transmitted to remote conference endpoint 102, which also renders obvious this limitation. Ex_1003 ¶¶366-376.

C. Ground 3: Claims 1-4 Are Obvious in Light of Drell in Combination with Knappe, Elliott, and the VPN Textbook

To the extent that Patent Owner argues that the claimed "second mixed data stream" does not require both audio and video data, then Drell and Knappe render obvious claims 1-4 for substantially the same reasons discussed above in Section VIII.A. Ex_1003 ¶377. Specifically, it would have been obvious to a POSITA that the PSTN conference endpoint suggested by Drell would be, for example, a "standard analog telephone[]" as taught by Knappe. Knappe (Ex. 1006) at 3:41-45; Ex_1003 ¶377. Since a standard telephone supports only audio communication, it would have been obvious to a POSITA for Drell's communication process 404 instantiated for a PSTN conference endpoint to transmit only the audio portion of the conference. Ex_1003 ¶377. Drell's audio mixing module 308 and communication process 404 instantiated for the PSTN conference client collectively teach a "second mixer," and their mixed audio output is a "second mixed data stream." *Id.* As discussed above, Knappe's gateway

converts from "VoIP format" to "standard telephony signals." Ex_1006, 4:18-25; Ex_1003 ¶377. Thus, the gateway has an output ("*second transport output*") that provides standard telephony signals ("*mixed audio data stream, corresponding to the second mixed data stream*") to the PSTN conference endpoint. Ex_1003 ¶377.

D. Ground 4: Claims 5-8 Are Obvious Over Drell in Combination with Knappe, Elliott, and the VPN Textbook

To the extent that Patent Owner argues that the claimed "*second mixed data stream*" does not require both audio and video data, then Drell, Knappe, Elliott and the VPN Textbook (without Voois) render obvious claims 5-8 for the same reasons discussed above in Section VIII.B. *Id.* ¶378.

IX. CONCLUSION

For the reasons given above, Petitioner has demonstrated a reasonable likelihood of success in establishing the unpatentability of claims 1-8 of the '525 patent and respectfully requests institution of IPR for each of the grounds presented herein and cancellation of the challenged claims.

Respectfully submitted,

Date: December 16, 2022

By: /Hector Ribera/

Hector J. Ribera (Reg. No. 54,397)
548 Market St., Suite 36117, San Francisco, CA 94104
Email: hector@martonribera.com
Tel: (415) 360-2511

Attorneys for Petitioner

CERTIFICATE OF WORD COUNT PURSUANT TO 37 C.F.R. § 42.24(D)

The undersigned hereby certifies that the word count for the above Petition for *inter partes* review is 12,367, which is less than the 14,000 allowed under 37 CFR § 42.24(a)(1)(i).

Date: December 16, 2021

By: /Hector Ribera/

Hector J. Ribera (Reg. No. 54,397)
Marton Ribera Schumann & Chang LLP
548 Market St., Suite 36117, San Francisco, CA 94104Email: hector@martonribera.com
Tel: (415) 360-2511

Attorneys for Petitioner

CERTIFICATE OF SERVICE PURSUANT TO 37 C.F.R. § 42.6 AND 37 C.F.R. § 42.105

Pursuant to 37 C.F.R. § 42.6(e)(4) and 42.105(b), the undersigned certifies that on December 16, 2022, a complete and entire copy of this Petition for *inter partes* review and all supporting exhibits were provided via FedEx First, to the Patent Owner by serving the correspondence address of record as follows:

ATTN: THOMAS MARTIN
AMEDEO FERRARO
WESLEY MEINERDING
MARTIN & FERRARO, LLP
1557 LAKE O'PINES STREET, NE
HARTVILLE, OH 44632

Date: December 16, 2022

By: /Hector Ribera/
Hector J. Ribera (Reg. No. 54,397)
Marton Ribera Schumann & Chang LLP
548 Market St., Suite 36117, San Francisco, CA 94104Email: hector@martonribera.com
Tel: (415) 360-2512

Attorneys for Petitioner