



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 603 02 561 T2** 2006.06.14

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 491 044 B1**

(21) Deutsches Aktenzeichen: **603 02 561.7**

(86) PCT-Aktenzeichen: **PCT/IB03/01468**

(96) Europäisches Aktenzeichen: **03 712 557.2**

(87) PCT-Veröffentlichungs-Nr.: **WO 03/081892**

(86) PCT-Anmeldetag: **27.03.2003**

(87) Veröffentlichungstag
der PCT-Anmeldung: **02.10.2003**

(97) Erstveröffentlichung durch das EPA: **29.12.2004**

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **30.11.2005**

(47) Veröffentlichungstag im Patentblatt: **14.06.2006**

(51) Int Cl.⁸: **H04N 7/14 (2006.01)**
H04N 7/15 (2006.01)

(30) Unionspriorität:

114402 27.03.2002 US

(73) Patentinhaber:

**Marconi Intellectual Property (Ringfence) Inc.,
Warrendale, Pennsylvania, US**

(74) Vertreter:

**Benedum Haseltine Lake Partners, 81669
München**

(84) Benannte Vertragsstaaten:

**AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB,
GR, HU, IE, IT, LI, LU, MC, NL, PT, RO, SE, SI, SK,
TR**

(72) Erfinder:

**CAMPBELL, Craig, Robert, Cranberry Township,
US; DRURY, Markham, David, Pittsburgh, US;
HUBER, R., Richard, Harmony, US; STRAIT, P.,
John, Pittsburgh, US; HILL, D., Peter, Monroeville,
US; ROSEN, Brian, Mars, US**

(54) Bezeichnung: **TELEKOMMUNIKATIONS SYSTEM**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

[0001] Die Erfindung betrifft allgemein das Gebiet der Kommunikation und insbesondere Vorrichtungen und Verfahren im Zusammenhang mit der Video- und Audiokommunikation.

[0002] Diese Patentschrift steht im Zusammenhang mit dem gleichzeitig eingereichten US-Entwurfspatent, laufende Nr. 29/158,247 mit dem Titel "Video-phone" von Marconi Communications Inc. Die Patentschrift steht auch im Zusammenhang mit der gleichzeitig eingereichten US-Patentschrift, laufende Nr. 10/115,323 mit dem Titel "Method and Apparatus for Displaying Images in Combination With Taking Images", von Marconi Communications Inc.

[0003] Die Erfindung betrifft ein Videophon. Die Erfindung betrifft insbesondere ein Videophon mit den Eigenschaften: hohe Auflösung, hohe Bildrate, Voll-duplex, geringe Verzögerung, Auslöschung von Video-, Audio- und Stereoechos, selbsttätige Mehrparteien-Konferenzdienste, Anwesenheitsdienste, Dienste für gemeinsam benutzte Oberflächen und TV-Verteilung.

[0004] Geographisch verstreute Teams sind in nahezu jedem großen Unternehmen (Marconi, Ford, Boeing, IBM), in jedem Unternehmen, das ein anderes Unternehmen übernimmt und in zahlreichen jungen Firmen der Normalfall. Reisen sind teuer, wenig ergiebig und unerwünscht (sie bieten keinen romantischen Anreiz mehr). Ressourcen (Personal) sind teuer, begrenzt und nur schwer zu verpflanzen.

[0005] Es besteht Bedarf für eine wesentlich verbesserte Kommunikation zwischen den Mitgliedern geographisch verstreuter Teams, und zwar durch aufgerüstete Kommunikationssysteme, die den Teammitgliedern zur Verfügung stehen, damit die Kommunikationsmöglichkeiten besser nachgebildet werden, die die Mitglieder an einem gemeinsamen Standort hätten. Das derzeit gängige Telefon ist hinsichtlich der Audiokommunikation zu einer natürlichen Erweiterung des täglichen Lebens geworden und erfüllt zahlreiche Bedürfnisse der Teammitglieder bzw. Benutzer für die gegenseitige Kommunikation. Ein aufgerüstetes Kommunikationssystem muss für die Benutzer so zuverlässig und selbstverständlich sein wie das Telefon, und es muss im Wesentlichen ein Telefon sein, dem jedoch eine Videodimension zugefügt ist, die dem Benutzer ebenfalls selbstverständlich ist. Die Qualität der Audio- und Videosignale einschließlich der Verzögerung derartiger Kommunikationssysteme muss so gut sein, dass sie die Benutzer ermutigt, ungezwungen zu kommunizieren. Der Bildschirm, die Lautsprecher, die Mikrofone und möglicherweise tausende von Kilometern zwischen den Benutzern sollten sich nicht bemerkbar machen.

[0006] Die Kommunikationsumgebung muss jedem Benutzer die Fähigkeit verleihen, "sofort zu wissen", wer an einer Konferenz teilnimmt, womit die anderen Benutzer gerade beschäftigt sind, und wann und wie er mit ihnen sprechen kann. Die Benutzer sollten improvisierte "Küchenklatsch"-Diskussionen führen und sogar über die Entfernung hinweg gemeinsam essen können, ohne sich um die dahinter stehenden Mechanismen Gedanken machen zu müssen.

[0007] Alle Handlungen und Merkmale sollten selbstverständlich sein und nicht geübt werden müssen. Zur Benutzerzielgruppe gehören Computerexperten und Neulinge ebenso wie Personen, die das Telefon intensiv oder auch sehr wenig nutzen. Es sollte nur in sehr geringem Umfang oder überhaupt nicht erforderlich sein, dass der Benutzer ein "Computernarr" ist. Informationen müssen leicht zugänglich sein, beispielsweise Namen, Telefonnummern, Pläne, Orte, die Uhrzeit an entfernten Orten, Anwesenheitsstatus, ein Protokoll der geführten Gespräche usw. Die sofortige Erreichbarkeit von Telefonnummern und die Verbindung, um sie wählen zu lassen, soll dieses Kommunikationssystem zur bevorzugten Echtzeitquelle für derartige Informationen machen. Es sollte Verbindungen zu vorhandenen Kontaktinformationsquellen aufweisen, beispielsweise Exchange, PDA (PDA = Personal Digital Assistant) usw., und bei geeigneter Nutzung dem Benutzer so wenig Handlungen abverlangen wie möglich ist, ohne die Benutzerhoheit zu beeinträchtigen.

[0008] Geringe Verzögerungszeiten sind ein Schlüsselfaktor für eine natürliche Kommunikation; jedermann, der bereits einmal eine Satellitenverbindung bei einem Telefongespräch verwendet hat, weiß dies nur zu gut. Bis heute gibt es gewichtige technische Probleme mit der Verzögerung in der Videokommunikation. Dabei führen die Gesamtverzögerung aus der Videocodierung, üblichen Netzverzögerungen und der Videodecodierung zu unannehmbaren Ergebnissen. Die RTP-Mechanismen waren nicht in der Lage, eine annehmbar geringe Verzerrungsrate zusammen mit geringer Verzögerung zu liefern. Die Folge ist, dass die derzeitige Videokonferenztechnik nicht die Transparenz erreicht, die für eine wirklich effektive Kommunikation erforderlich ist. Es gibt noch keine akzeptablen Präsentationen für Mehrbenutzerkonferenzen, die auf dem Bildschirm erzeugt und für einen Desktop gedacht sind. Vorhandene Telekonferenzsysteme kranken an einer mangelnden Kommunikationswirksamkeit bei wachsender Teilnehmerzahl. Sind mehr als zwei Personen an der Konferenz beteiligt, so neigen Steuerung, Verwaltung, Schnittstellen und Leistungsfähigkeit des Systems dazu, schlechter zu werden, und die Transparenz des Mediums geht verloren.

[0009] Die Erfindung ist dafür gedacht, ein Massenprodukt zu werden, das die meisten gängigen For-

men von Desktop-Telefonen ersetzen und eine Liste mit verbesserten Merkmalen anbieten kann, die Lösungen für die genannten Probleme liefern. Die Erfindung fügt der derzeit gängigen Kommunikationsumgebung Sichtbarkeit in Echtzeit und weitere Hilfen hinzu und führt schließlich zu einem unverfälschten Eindruck der Tele-Anwesenheit. Anders ausgedrückt kann man sich im Gegensatz zu derzeitigen Hilfsmitteln mit einer oder mehreren Personen an entfernten Orten in einer Weise austauschen, die viel näher an einem tatsächlichen körperlichen Treffen liegt. Die Erfindung stellt die folgenden Eigenschaften bereit:

- Sie besitzt das Potential, sowohl die Leistungsfähigkeit als auch die Wirksamkeit der Kommunikation in einer geographisch verstreuten Organisation durchschlagend zu verbessern.
- Die Leistungsfähigkeit verbessert sich durch den Einschluss der visuellen Seite der Kommunikation, da wir ca. 70 Prozent des Kommunikationsinhalts visuell erfassen, und zwar unabhängig von der eigentlichen Audiokommunikation. Die Leistungsfähigkeit verbessert sich, weil die unmittelbare ereignisgesteuerte Kommunikation (ad hoc) anstelle einer geplanten Kommunikation vereinfacht wird, da die erstgenannte Kategorie etwa 70 Prozent der lokalen Kommunikation ausmacht, die nun auf Fernkommunikation erweitert werden kann.
- Die Leistungsfähigkeit verbessert sich, weil die Benutzer dieses Werkzeug – wie ihr derzeitiges Telefon – an ihrem Schreibtisch kontinuierlich für den sofortigen Einsatz zur Verfügung haben. Im Gegensatz dazu finden derzeit die meisten Videokonferenzen in besonders dafür eingerichteten Räumen statt und müssen im voraus gebucht werden.
- Die Leistungsfähigkeit verbessert sich durch das Verteilen von "Anwesenheitsinformation", die sicherstellt, dass mehr Kommunikationsversuche erfolgreich abgeschlossen werden.
- Die Leistungsfähigkeit verbessert sich, weil dem Benutzer gleichzeitig getrennte Computer- und Kommunikationsplattformen zur Verfügung stehen.
- Die Leistungsfähigkeit verbessert sich, weil die Zeit verringert wird, die eine einschlägige Besprechungsgruppe zum Lösen einer Problemstellung benötigt, weil die Entscheidungszyklen wesentlich beschleunigt werden, und weil die Notwendigkeit zu reisen stark abnimmt.
- Die selteneren Reisen setzen nicht nur mehr produktive Zeit frei, sondern sparen auch in großem Umfang Flug- und Hotelkosten ein.
- Die Wirksamkeit verbessert sich, weil sichergestellt wird, dass die meisten beteiligten Personen in eine Entscheidung einbezogen werden, und nicht nur die gerade "greifbaren". Die Wirksamkeit wächst auch aufgrund der verbesserten persönlichen Beziehungen, die sich entwickeln können, wenn zum einfachen Telefonkontakt ein hochwer-

tiger Sichtkontakt hinzukommt.

- Es wird auch möglich, an einer bestimmten Stelle verfügbares Know-How überall in der Organisation sofort "anzuzapfen", wodurch derartiges Fachwissen sehr viel produktiver eingesetzt wird.
- Es könnte ein Transportmittel für eine wesentlich verbesserte Kommunikation zwischen Management und Belegschaft sein, und insbesondere von den führenden Unternehmensmanagern zur gesamten Firma, und das Hervorbringen von Teamgeist und Verpflichtung unterstützen.
- Wie bei jedem wirklich neuen Werkzeug werden kluge Köpfe Wege finden, es zum Nutzen der Organisation einzusetzen, an die man derzeit überhaupt nicht denkt.

[0010] Die Erfindung stellt eine Erweiterung des Telefonsystems dar, mit dem die Anwender gut vertraut sind. D. h., es handelt sich um eine Kommunikationsplattform und nicht um eine PC-Anwendung. Sofortiger Zugriff, Zuverlässigkeit und eine einfache intuitive Funktionalität sind einige seiner wesentlichen Merkmale.

[0011] Im Folgenden werden einige derzeit vorhandene Produkte für die Zusammenarbeit per Video vorgestellt.

[0012] Microsoft NetMeeting. Jedermann, der versucht hat, damit zu arbeiten, kennt dessen Grenzen. Es ist unerträglich langsam. Die Videobilder sind nahezu nicht verwendbar. Die Audioqualität ist gering. Die Benutzerschnittstelle ist außer für einfache Gespräche nur schwer verständlich.

[0013] Polycom, ViaVideo™. ViaVideo™ ist eine H.323-Vorrichtung, die eine Kamera und etwas Audioverarbeitung in einem Monitor-Aufsetzgehäuse mit einem USB-Anschluss für einen PC enthält. Das Produkt erzeugt Videosignale in H.261-Qualität mit bis zu 384 K, wobei die Ausgabe auf dem PC-Bildschirm in einem Fenster erfolgt, und eine Bildschirmemulation einer Infrarot-Fernbedienung als Benutzerschnittstelle. Die Videoqualität ist annehmbar, wobei die Kamera relativ schlecht ist. Die Audioqualität ist annehmbar. Eine Echounterdrückung ist verfügbar. Der Gesamteindruck liefert jedoch nicht die erforderliche Transparenz, die für eine wirksame Kommunikation nötig ist. Diese Videoeinrichtung liefert jedoch ein Niveau der Videoqualität zu einem bisher unerreichten Preis, nämlich für ungefähr 600 Dollar. Es gibt keine Anwesenheit, keine Telefonintegration, kein CTI (CTI = Computer Telephony Integration). Die Fenstergröße auf dem PC-Bildschirm schränkt die sonstigen Handlungsmöglichkeiten ein.

[0014] Webex ist ein Web-gestütztes Konferenzprodukt, das hauptsächlich als Dienst angeboten wird. Das Hauptproblem bei Webex ist die Bandbreite, da das derzeitige Produkt auf dem Internet aufsetzt (ca-

pital I). Die derzeitige Implementierung ist Server-gestützt, wodurch die Handlungsspielräume weiter eingeschränkt sind. Webex weist wesentlich einfachere Optionen für gemeinsame Oberflächen auf als Net-Meeting und bietet eine geringe Audioqualität, kein Video, kein CTI, keine Anwesenheit usw.

[0015] Teraglobal setzt auf der Apple-Plattform auf. Es wird nicht als Videophon sondern als vollständige Anwendungsplattform angeboten, in der die Kommunikation nur eine der Fähigkeiten darstellt. Die Technologie ist sehr eng mit dieser Plattform und dem zugrunde liegenden PowerPC-Prozessor "DigitalDNA" von Motorola verwoben, auf dem es läuft. Das gesamte Produkt ist völlig eigenständig und weist keine Gemeinsamkeiten mit irgendwelchen anderen Elementen auf. Es musste darauf zurückgegriffen werden, eine eigene E-Mail, Kalenderfunktionen und Scheduling-Programme zu entwickeln, um die benötigte Integration zu erzielen. Es weist eine gewisse Qualität auf (wenn man in Betracht zieht, dass die Bandbreite auf 1,5 Mbit/sec begrenzt ist). Seine Prüfwerkzeuge und Konferenzkontrollmöglichkeiten und ähnliche Fähigkeiten heben es in den Bereich der Großkonferenzen. Für eine Zusammenarbeit in Gruppen sind derartige Fähigkeiten nicht erforderlich. Müssen derartig viele Teilnehmer unterstützt werden, so nimmt die Vertrautheit des Benutzers im Vergleich zu einer geringeren Teilnehmerzahl beträchtlich ab. Die übliche Konferenzansicht ist beispielsweise eine einzige große Darstellung des Sprechenden, die halbautomatisch "weitergereicht" wird, sowie Miniaturbilder der anderen Teilnehmer, die jeweils nach einigen Sekunden aktualisiert werden.

[0016] Die Erfindung besteht aus mehreren End-Terminals (Geräten), einem Satz Servern, die Merkmale liefern, die nicht in die Geräte eingebaut sind, und aus einem Satz Gateways, die die Produkte mit vorhandenen Ressourcen und externen PSTN-Diensten (PSTN, PSTN = Public Switched Telecommunications Network, öffentliches Telefonnetz) verbinden. Die Grundfunktionen, die die Erfindung bietet, sind:

- Telefondienste, wobei Video in allen "on-net"-Gesprächen verfügbar ist, Audio und Video in sehr guter Qualität.
- Konferenzdienste für mehrere Teilnehmer, Audio und Video, ad hoc oder vorgeplant, vollständige Selbstbedienung, volle Integration in die Telefondienste.
- Anwesenheitsdienste, die verschiedene Werkzeuge zum Ermitteln der Erreichbarkeit für die Zusammenarbeit bieten.
- Dienste für gemeinsame Oberflächen wie elektronische Präsentationstafel, gemeinsame Anwendungen, gemeinsame Dokumente, Präsentationsverteilung über verbundene Anwendungen auf einem zugeordneten PC.
- Zusätzliche nützliche Dienste, beispielsweise

das Verteilen von TV-Videos.

[0017] Die Erfindung ist das bekannte Telefon mit gewaltigen neuen Funktionen, nicht der Computer, der versucht, das Telefon nachzuahmen. Dies erlaubt es, den Computer gleichzeitig für die Dinge zu nutzen, für die er geeignet ist, und bietet eine flexible und zugleich anwendungsspezifische Kommunikationsvorrichtung. Die Benutzerschnittstelle und den Hardwareentwurf kann man auf diese Anwendung abstimmen und eine sofort betriebsbereite und hochgradig zuverlässige Kommunikationseinrichtung vergleichbar zum derzeitigen Telefon bereitstellen. PCs werden diese Eigenschaften niemals haben. Dieser Ansatz erlaubt auch eine Kontrolle über die Betriebsumstände der Vorrichtung. Die Supportprobleme im Zusammenhang mit PC-Hardware- und Softwarekonfigurationsfragen treten nicht mehr auf.

ZUSAMMENFASSUNG DER ERFINDUNG

[0018] Gemäß einem ersten Aspekt der Erfindung wird ein Telekommunikationssystem bereitgestellt, umfassend: ein Netz; und einen ersten Knoten, einen zweiten Knoten und mindestens einen dritten Knoten, die miteinander über das Netz kommunizieren, dadurch gekennzeichnet, dass der erste Knoten dafür eingerichtet ist, einen ersten Videostrom einer Szene am ersten Knoten, einen zweiten Videostrom der Szene am ersten Knoten und einen Audiostrom der Szene am ersten Knoten an den zweiten Knoten und den dritten Knoten zu senden, und dass der zweite Knoten und der dritte Knoten dafür eingerichtet sind, den Audiostrom und entweder den ersten Videostrom oder den zweiten Videostrom wiederzugeben.

[0019] Gemäß einem zweiten Aspekt der Erfindung wird ein Videophon bereitgestellt, umfassend: eine Bildgebervorrichtung, die ein Videobild einer Szene aufnimmt und gekennzeichnet ist durch Vorrichtungen, die einen ersten Videostrom der Szene und einen zweiten Videostrom der Szene aus dem Videobild erzeugen und den ersten Videostrom und den zweiten Videostrom gleichzeitig an ein Netz senden, wobei die Erzeugungsvorrichtungen mit der Bildgebervorrichtung kommunizieren.

[0020] Gemäß einem dritten Aspekt der Erfindung wird ein Verfahren für einen Videoanruf bereitgestellt, umfassend den Schritt: Aufnehmen eines Videobilds einer Szene, gekennzeichnet durch die Schritte: Erzeugen eines ersten Videostroms der Szene und eines zweiten Videostroms der Szene aus dem Videobild; und gleichzeitiges Senden des ersten Videostroms und des zweiten Videostroms an das Netz.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0021] In den beiliegenden Zeichnungen sind die

bevorzugte Ausführungsform der Erfindung und bevorzugte Verfahren zum Verwirklichen der Erfindung erläutert. Es zeigt:

[0022] [Fig. 1](#) eine skizzenhafte Darstellung eines Systems der Erfindung;

[0023] [Fig. 2](#) eine skizzenhafte Darstellung eines Netzes der Erfindung;

[0024] [Fig. 3](#) eine skizzenhafte Darstellung eines Videophons, das an einen PC und an ein Netz angeschlossen ist;

[0025] [Fig. 4](#) eine skizzenhafte Darstellung des Systems der Erfindung;

[0026] [Fig. 5a](#) und [Fig. 5b](#) skizzenhafte Darstellungen der Vorderansicht und der Seitenansicht des Videophons;

[0027] [Fig. 6](#) eine skizzenhafte Darstellung einer Anschlussstafel des Videophons;

[0028] [Fig. 7](#) eine skizzenhafte Darstellung einer Anordnung des Videophons mit mehreren Bildschirmen;

[0029] [Fig. 8](#) ein Blockdiagramm des Videophons;

[0030] [Fig. 9](#) ein Blockdiagramm der Videophonarchitektur;

[0031] [Fig. 10](#) eine skizzenhafte Darstellung des Systems;

[0032] [Fig. 11](#) eine skizzenhafte Darstellung des Systems;

[0033] [Fig. 12](#) ein Blockdiagramm eines Anwesenheitssensors;

[0034] [Fig. 13](#) eine Seite eines Berührungsbildschirms des Videophons;

[0035] [Fig. 14](#) ein Blockdiagramm einer Vorrichtung der Erfindung;

[0036] [Fig. 15](#) eine skizzenhafte Darstellung eines Systems der Erfindung;

[0037] [Fig. 16](#) eine skizzenhafte Darstellung eines weiteren Systems der Erfindung; und

[0038] [Fig. 17](#) eine skizzenhafte Darstellung einer Vorrichtung der Erfindung.

[0039] Es wird nun Bezug auf die Zeichnungen und insbesondere auf [Fig. 1](#), [Fig. 8](#) und [Fig. 10](#) genommen, die ein Videophon **15** zeigen. In den Zeichnun-

gen bezeichnen gleiche Bezugszeichen innerhalb der diversen Abbildungen ähnliche oder identische Teile. Das Videophon **15** enthält eine Bildgebervorrichtung **30** zum Aufnehmen eines Videobilds einer Szene. Das Videophon **15** umfasst eine Vorrichtung zum Erzeugen eines ersten Videostroms der Szene und eines zweiten Videostroms der Szene aus dem Videobild und zum gleichzeitigen Senden des ersten Videostroms und des zweiten Videostroms auf ein Netz **40**. Die Erzeugungsvorrichtung kommuniziert mit der Bildgebervorrichtung **30**. Die Erzeugungsvorrichtung kann bevorzugt mehrere Videostrome der Szene erzeugen, wobei sich jeder Videostrom bevorzugt von den anderen unterscheidet. Die Erzeugungsvorrichtung enthält bevorzugt einen Codierer **36** und ein Field Programmable Gate Array **38**, das mit der Bildgebervorrichtung **30** kommuniziert sowie eine Netzchnittstelle **42**, die mit dem Field Programmable Gate Array **38** kommuniziert.

[0040] Die Erfindung betrifft ein Verfahren zum Ausführen eines Videotelefonats. Das Verfahren umfasst die Schritte des Aufnehmens eines Videobilds einer Szene. Es gibt den Schritt des Erzeugens eines ersten Videostroms der Szene und eines zweiten Videostroms der Szene aus dem Videobild. Es gibt auch den Schritt des gleichzeitigen Sendens des ersten Videostroms und des zweiten Videostroms auf einem Netz **40**.

[0041] Die Erfindung betrifft ein Telekommunikationssystem **10**. Das System **10** umfasst ein Netz **40**. Das System **10** enthält einen ersten Knoten **80**, einen zweiten Knoten **82** und mindestens einen dritten Knoten **84**, die untereinander über das Netz **40** kommunizieren. Der erste Knoten **80** sendet einen ersten Videostrom einer Szene am ersten Knoten **80**, einen zweiten Videostrom der Szene am ersten Knoten **80** und einen Audiostrom der Szene am ersten Knoten **80** an den zweiten Knoten und den dritten Knoten **84**. Der zweite Knoten und der dritte Knoten **84** geben den Audiostrom und entweder den ersten Videostrom oder den zweiten Videostrom wieder.

[0042] Der erste Videostrom weist bevorzugt eine Bildrate auf, die größer ist als 25 Bilder je Sekunde, und der zweite Videostrom hat bevorzugt eine Bildrate, die kleiner als 25 Bilder je Sekunde ist. Der erste Videostrom hat bevorzugt eine Bitrate, die größer oder gleich 1,5 Mbit je Sekunde ist, und der zweite Videostrom hat bevorzugt eine Bitrate, die kleiner ist als 1,5 Mbit je Sekunde. Der zweite Knoten und der dritte Knoten **84** besitzen bevorzugt einen Darstellungsbildschirm. Wird am zweiten Knoten oder am dritten Knoten **84** der erste Videostrom dargestellt, so wird der erste Videostrom als großes Bild auf einer Fläche dargestellt, die 20 oder mehr Prozent des Bildschirms einnimmt. Wird am zweiten Knoten oder am dritten Knoten **84** der zweite Videostrom dargestellt, so wird der zweite Videostrom als kleines Bild auf einer Flä-

che dargestellt, die weniger als 20 Prozent des Bildschirms einnimmt. Bevorzugt enthält das System **10** einen vierten Knoten, der mit dem ersten Knoten, dem zweiten Knoten und dem dritten Knoten **84** über das Netz **40** kommuniziert und der einen Fernseh-Videostrom eines Fernsehkanals an den ersten Knoten, den zweiten Knoten und den dritten Knoten **84** sendet. Der erste Knoten, der zweite Knoten und der dritte Knoten **84** können den Fernseh-Videostrom auf dem Bildschirm zusammen mit dem ersten Videostrom darstellen.

[0043] Der zweite Knoten **82** sendet bevorzugt einen ersten Videostrom einer Szene am zweiten Knoten **82**, einen zweiten Videostrom der Szene am zweiten Knoten **82** und einen Audiostrom der Szene an den dritten Knoten **84**. Der dritte Knoten **84** weist einen Anzeigecontroller **52** auf, der dazu dient, das Bild zu kontrollieren, das auf dem Bildschirm erscheint, und der jeden ersten Videostrom von jedem Knoten zusammen neben den anderen Strömen auf dem Bildschirm eines dritten Knotens **84** wiedergibt. Das Netz **40** ist bevorzugt ein Ethernet oder ein ATM-Netz **40**. Der erste und der zweite Videostrom und die Audioströme des ersten Knotens und des zweiten Knotens **82** werden für ein Videokonferenzgespräch bevorzugt über das Netz **40** gesendet, wobei für das Videokonferenzgespräch keine Konferenzbrücke oder MCU (MCU = Multipoint Control Unit) verwendet wird. Jeder Knoten verwendet für das Videokonferenzgespräch bevorzugt ATM-Point zu Multipoint.

[0044] Der dritte Knoten **84** legt bevorzugt vorab fest, ob der erste oder der zweite Videostrom vom ersten Knoten oder vom zweiten Knoten **82** dargestellt wird. Der dritte Knoten **84** wählt bevorzugt die Darstellung des ersten Videostroms vom ersten Knoten oder vom zweiten Knoten **82**, falls ein Benutzer in der jeweiligen Szene am ersten Knoten oder zweiten Knoten **82** gerade spricht, oder der dritte Knoten **84** hat vorbestimmt, den ersten Videostrom des ersten Knotens oder des zweiten Knotens **82** darzustellen. Der erste Videostrom des ersten Knotens und des zweiten Knotens **82** liegt bevorzugt in einem gewünschten Format vor, beispielsweise dem Format MPEG-2, wenn er über das Netz **40** gesendet wird. Der erste und der zweite Knoten **82** verwenden bevorzugt Continuous P dazu, den ersten Videostrom vom ersten Knoten und vom zweiten Knoten **82** in das Format MPEG-2 umzusetzen. Der erste und der zweite Knoten **82** beschneiden bevorzugt die ersten Videostrome der Szene am ersten Knoten bzw. am zweiten Knoten **82**.

[0045] Der erste und der zweite Knoten **82** beschneiden bevorzugt den ersten Videostrom ihrer jeweiligen Szenen dadurch, dass sie einen Abschnitt des ersten Videostroms entfernen, der zu einem Ort der entsprechenden Szene gehört, an dem sich kein

Benutzer befindet. Der erste und der zweite Knoten **82** senden bevorzugt nur einen Audiostrom der Szene des ersten Knotens bzw. des zweiten Knotens **82**, falls ein Geräusch über einem vorbestimmten Grenzwert in der jeweiligen Szene auftritt. Der erste Knoten **80** weist bevorzugt einen Anwesenheitssensor auf, der feststellt, ob ein Benutzer in der Szene am ersten Knoten **80** vorhanden ist, und der eine Anwesenheitskennung erzeugt, die angibt, ob sich der Benutzer am ersten Knoten **80** aufhält. Der erste Knoten **80** sendet die Anwesenheitskennung an den zweiten Knoten und den dritten Knoten **84**. Der erste Knoten **80** erzeugt bevorzugt ein Warnsignal, das jeden Benutzer in der Szene am ersten Knoten **80** daran erinnert, dass die Anwesenheitskennung nach Ablauf einer vorbestimmten Zeitspanne erstellt wird.

[0046] Der erste Knoten **80** enthält bevorzugt eine Bildgebervorrichtung, die ein Videobild der ersten Szene aufnimmt und den ersten Videostrom erzeugt. Das System **10** enthält bevorzugt einen Codierer **36**, der mit der Bildgebervorrichtung kommuniziert und den ersten Videostrom ohne Bildpufferung in das Format MPEG-2 komprimiert und codiert. Der erste Knoten **80** enthält ein Field Programmable Gate Array **38**, das mit dem Codierer **36** kommuniziert, den ersten Videostrom in Pakete unterteilt und auch den ersten Videostrom von der Bildgebervorrichtung empfängt und den zweiten Videostrom des ersten Knotens **80** erzeugt und den zweiten Videostrom in Pakete unterteilt. Der erste Knoten **80** enthält bevorzugt eine Netzchnittstelle **42**, die mit dem Field Programmable Gate Array **38** und dem Netz **40** kommuniziert und den ersten Videostrom des ersten Knotens **80** an das Netz **40** überträgt und den ersten Videostrom vom zweiten Knoten **82** empfängt und ihn an das Field Programmable Gate Array **38** sendet.

[0047] Der erste Knoten **80** enthält bevorzugt Mikrofonvorrichtungen, die Schall von der ersten Szene aufnehmen und den Audiostrom des ersten Knotens **80** erzeugen. Der erste Knoten **80** umfasst bevorzugt eine Lautsprechervorrichtung, die mit der Netzchnittstelle **42** kommuniziert und den Audiostrom vom zweiten Knoten **82** wiedergibt. Der erste Knoten **80** enthält bevorzugt einen DSP **62**, der den Audiostrom des ersten Knotens **80** in Pakete zerlegt und den Audiostrom an das Field Programmable Gate Array **38** liefert, das den Audiostrom des ersten Knotens **80** an die Netzchnittstelle **42** überträgt, die den Audiostrom des ersten Knotens **80** an das Netz **40** überträgt und den Audiostrom des zweiten Knotens **82** vom Field Programmable Gate Array **38** empfängt. Der erste Knoten **80** enthält bevorzugt eine Audioschnittstelle **60**, die den Audiostrom des ersten Knotens **80** von den Mikrofonvorrichtungen empfängt, ihn digitalisiert und an den DSP **62** liefert, und den Audiostrom vom zweiten Knoten **82**, den sie vom DSP **62** empfängt, in eine analoge Form umsetzt, die die Lautsprechervorrichtung wiedergeben kann.

[0048] Bevorzugt versieht die Netzschnittstelle **42** Pakete des Audiostroms und Videostroms des ersten Knotens **80** mit einer Zeitangabe, bevor sie an das Netz **40** gesendet werden, und sie richtet die Pakete des Videostroms und des Audiostroms des zweiten Knotens **82**, die der erste Knoten **80** empfängt, anhand der Zeitangaben aus, so dass beim Wiedergeben des Videostroms und des Audiostroms des zweiten Knotens **82** am ersten Knoten **80** der zugehörige Schall mit den Bild der Szene des zweiten Knotens **82** wiedergegeben wird. Das System **10** enthält bevorzugt einen Empfangsspeicher, in dem der erste Videostrom vom zweiten Knoten **82** empfangen und gespeichert wird, und einen Hauptcontroller **50**, der mit der Netzschnittstelle **42**, dem Codierer **36**, dem Field Programmable Gate Array **38** und dem DSP **62** verbunden ist und diese Geräte steuert, wobei der Hauptcontroller **50** der Netzschnittstelle **42** befiehlt, den ersten Videostrom des zweiten Knotens **82** zu wählen und ihn an den Empfangsspeicher zu senden, und der Hauptcontroller **50** den ersten Videostrom des zweiten Knotens **82**, der im Empfangsspeicher abgelegt ist, decodiert und expandiert und ihn an den Anzeigecontroller **52** sendet.

[0049] Bevorzugt enthält der erste Knoten **80** einen LCD-Controller, der an den Anzeigecontroller **52** angeschlossen ist, und der Anzeigebildschirm enthält einen Flachbildschirm, der mit dem LCD-Controller verbunden ist, wobei der LCD-Controller den ersten Videostrom des zweiten Knotens **82** vom Anzeigecontroller **52** empfängt und den ersten Videostrom des zweiten Knotens **82** für die Darstellung auf dem Flachbildschirm aufbereitet. Das System **10** enthält bevorzugt einen Berührbildschirm, auf dem Tasten mit zugeordneten Funktionen erscheinen, und einen Berührbildschirm-Controller, der mit dem Berührbildschirm und dem Hauptcontroller **50** verbunden ist und Informationen bezüglich der Tasten an den Hauptcontroller **50** überträgt, die ein Benutzer auf dem Berührbildschirm berührt hat. Der erste Knoten **80** enthält bevorzugt einen Decodierer **34**, der den ersten Videostrom von der Bildgebervorrichtung der Szene des ersten Knotens **80** in digitale Form umsetzt und ihn an den Codierer **36** und das Field Programmable Gate Array **38** liefert, wobei der Decodierer **34** mit dem Field Programmable Gate Array **38** und dem Codierer **36** verbunden ist, und der Decodierer **34** auch den Fernseh-Videostrom und andere analoge Videoströme empfängt und sie in digitale Form umsetzt.

[0050] Die Kameravorrichtung enthält bevorzugt eine analoge Videokamera, die mit dem Decodierer **34** kommuniziert, eine digitale Videokamera **47**, die mit dem Codierer **36** und dem Field Programmable Gate Array **38** kommuniziert, oder eine Fire-Wire-Kamera, die mit dem Field Programmable Gate Array **38** kommuniziert, wobei das Field Programmable Gate Array **38** jeglichen Videostrom, den es von der Fi-

re-Wire-Kamera empfängt, an den Codierer **36** liefert. Der DSP **62** liefert bevorzugt eine Stereoecho-Auslöschung für den Audiostrom der Szene des ersten Knotens **80**, und der Codierer **36** liefert für den ersten Videostrom des ersten Knotens **80** eine Auflösung von mindestens 720 × 640 Pixel, und die Mikrofonvorrichtungen und die Lautsprechervorrichtung arbeiten im Vollduplexbetrieb. Der DSP **62** überwacht bevorzugt die Mikrofonvorrichtungen hinsichtlich des Geräuschpegels, den die Lautsprechervorrichtung erzeugt, und stellt die Lautsprechervorrichtung so ein, dass ein gewünschter Geräuschpegel beibehalten wird. Bevorzugt erkennt der Hauptcontroller **50** einen Befehl des Benutzers und ruft automatisch einen weiteren Benutzer zu einer Videokonferenz an, und verwendet den Befehl dazu, den anderen Benutzern anzukündigen, dass der weitere Benutzer für die Videokonferenz angefordert wird.

[0051] Der erste Knoten **80** besitzt bevorzugt eine Videomail, falls der erste Knoten **80** einen Videoanruf nicht annehmen kann. Bevorzugt umfasst das System **10** einen Server **66**, der mit dem Netz **40** und dem ersten Knoten **80** kommuniziert, der den Videoanruf für den ersten Knoten **80** empfängt, falls der erste Knoten **80** den Videoanruf nicht empfangen kann, der eine Videonachricht speichert, die zu dem Videoanruf gehört, den der erste Knoten **80** nicht empfangen kann, und der eine Videoanruf-Wartenachricht an den ersten Knoten **80** sendet, die angibt, dass eine Videonachricht darauf wartet, dem ersten Knoten **80** gezeigt zu werden. Der erste Knoten **80** umfasst bevorzugt mehrere Wiedergabebildschirme, die mit dem Anzeigecontroller **52** verbunden sind, wobei die Bilder der ersten Videoströme von unterschiedlichen Knoten auf den Bildschirmen nebeneinander so dargestellt werden, als wären die zahlreichen Bildschirme ein fortlaufender Bildschirm. Der erste Knoten **80** umfasst bevorzugt eine zweite Bildgebervorrichtung, die den ersten Videostrom des ersten Knotens **80** erzeugt, der die Szene am ersten Knoten **80** unter einem anderen Blickwinkel sieht als die Bildgebervorrichtung. Der Hauptcontroller **50** sendet bevorzugt den zweiten Videostrom des ersten Knotens **80** vom Field Programmable Gate Array **38** an den Anzeigecontroller **52**, damit er auf dem Bildschirm dargestellt wird und der Benutzer des ersten Knotens **80** die Szene am ersten Knoten **80** auf dem Anzeigebildschirm sehen kann. Der erste Knoten **80**, der zweite Knoten **82** und der dritte Knoten **84** enthalten bevorzugt ein Videophon **15**. Der erste Knoten **80**, der zweite Knoten **82** und der dritte Knoten **84** können einen PC **68** enthalten, der mit dem Videophon **15** kommuniziert.

[0052] Bei Betrieb der bevorzugten Ausführungsform, siehe [Fig. 8](#), [Fig. 9](#) und [Fig. 10](#), wandelt eine Bildgebervorrichtung **30**, beispielsweise eine herkömmliche analoge Kamera **32**, die von Sony mit

S-Video geliefert wird, die Bilder einer Szene von der Bildgebervorrichtung **30** in elektrische Signale um, die über ein Kabel an einen Videodecodierer **34** gesendet werden, beispielsweise einen Decodierer SAA7114 NTSC/PAL von Phillips. Der Videodecodierer **34** setzt die elektrischen Signale in digitale Signale um und sendet sie als Pixelstrom der Szene aus, beispielsweise im Format BT 656. Der vom Videodecodierer **34** ausgesendete Pixelstrom wird in einen ersten Strom und in einen zweiten Strom unterteilt, der mit dem ersten Strom identisch ist. Ein Codierer **36**, bevorzugt ein Codierer eNV 420 von IBM, empfängt den ersten Pixelstrom, bearbeitet den ersten Strom und erzeugt einen Datenstrom im Format MPEG-2. Verglichen mit den Daten, die die Kamera erzeugt, ist der Datenstrom, den der Videodecodierer **36** hervorbringt, um etwa den Faktor **50** komprimiert. Der MPEG-2-Strom ist ein codierter digitaler Strom, der keiner Bildpufferung unterzogen wird, bevor er anschließend in Pakete zerlegt wird. Damit hält man jegliche Verzögerung so klein wie möglich. Das Field Programmable Gate Array **38** (FPGA), an das der MPEG-2-Strom geliefert wird, paketierte zusammen mit Software den codierten MPEG-2-Digitalstrom mit Hilfe von RTP, und überträgt ihn an ein Netz **40**, beispielsweise ein Ethernet 802.p oder ATM mit 155 Megabit pro Sekunde. Dazu wird eine Netzchnittstelle **42** über eine PLX 9054 PCI-Schnittstelle **44** verwendet. Auf Wunsch kann der Decodierer **34** einen Videostrom empfangen, der zu einem VCR (VCR = Videorekorder) oder einem Fernsehprogramm gehört, beispielsweise CNN oder einen Spielfilm, und ihn dem Anzeigecontroller **52** für die Wiedergabe direkt zuleiten. Ein im FPGA **38** angeordneter Decodiercontroller **46**, der mit dem Decodierer **34** verbunden ist, steuert den Betrieb des Decodierers **34**.

[0053] Wird wahlweise eine Digitalkamera **47** verwendet, so liegt der entstehende Strom, den die Kamera erzeugt, bereits im digitalen Format vor, und man braucht keinen Decodierer **34** vorzusehen. Der digitale Strom von der Digitalkamera **47**, der das Format BT 656 hat, wird direkt von der Kamera in einen ersten und einen zweiten Strom aufgespalten und braucht dazu keinen Videodecodierer **34** zu durchlaufen.

[0054] Gemäß einer weiteren Alternative kann eine Fire-Wire-Kamera **48**, beispielsweise eine Fire-Wire-Kamera **48** mit einer 1394-Schnittstelle, dazu dienen, dem FPGA **38** direkt ein digitales Signal zu liefern. Die Fire-Wire-Kamera **48** bietet den Vorteil, dass, wenn der Datenstrom in einer mehr als sehr kurzen Entfernung vom FPGA **38** erzeugt werden muss, die Digitalsignale beispielsweise mit einem Kabel von der Fire-Wire-Kamera **48** über diese längere Entfernung übertragen werden können. Das FPGA **38** liefert die Digitalsignale von der Fire-Wire-Kamera **48** für die beschriebene Verarbeitung an den Codierer **36** und erzeugt auch, siehe unten, einen Strom

mit geringer Bildrate.

[0055] Aus dem zweiten Strom, der dem FPGA **38** geliefert wird, erzeugt das FPGA **38** und die Software einen Strom mit geringer Bildrate, beispielsweise einen Motion-JPEG-Strom, der verglichen mit dem ersten Strom weniger Bandbreite benötigt. Das FPGA **38** und ein Hauptcontroller **50** mit Software codieren, komprimieren und pakettieren diesen Strom mit geringer Bildrate und liefern ihn der PCI-Schnittstelle **44**, die ihn ihrerseits über eine Netzchnittstellenkarte **56** an die Netzchnittstelle **42** leitet, damit er über das Netz **40** übertragen wird. Der codierte digitale MPEG-2-Strom und der Strom mit geringer Bildrate sind zwei im Wesentlichen identische, jedoch voneinander unabhängige Datenströme. Der Strom mit geringer Bildrate ist jedoch verglichen mit dem MPEG-2-Datenstrom auf einen kleineren Maßstab skaliert und liefert bezogen auf den MPEG-2-Strom ein kleineres Bild der gleichen Szene und benötigt weniger Ressourcen des Netzes **40**.

[0056] Auf dem Netz **40** wird jeder Digitalstrom an ein gewünschtes Empfängervideophon **15** oder an Empfängervideophone **15** übertragen, falls eine Konferenz mit mehr als zwei Parteien ansteht. Die Daten werden mit SIP geroutet. Die Netzchnittstellenkarte **56** des Empfängervideophons **15** empfängt die zum ersten und zum zweiten Datenstrom gehörenden Pakete und liefert die Daten aus den Paketen und den Videostrom (den ersten oder zweiten), den der Hauptcontroller gewählt hat, an einen Empfangsspeicher. Ein Hauptcontroller **50** des Empfängervideophons **15** mit Software decodiert und expandiert den gewählten empfangenen Datenstrom und überträgt ihn an einen Anzeigecontroller **52**. Der Anzeigecontroller **52** stellt die wiederhergestellten Bilder auf einer digitalen VGA-Flachbildschirmanzeige mit Hilfe von Standard-Skalierungshardware dar. Der Anwender am Empfängervideophon **15** kann an einem Berührungsbildschirm **74** wählen, welchen der beiden Datenströme er sehen will. Auf Wunsch kann er beide wählen, so dass ein großes und ein kleines Bild der Szene dargestellt werden, obwohl die Wiedergabe beider Ströme auf dem Empfängervideophon **15** normalerweise nicht stattfindet. Die Protokolle für die Wiedergabe werden im Weiteren besprochen. Da der Benutzer wählen kann, ob er die große Ansicht der Szene oder die kleine Ansicht der Szene sehen will, hat der Benutzer die Möglichkeit, die Ressourcen des Systems **10** so zuzuweisen, dass die für ihn momentan wichtigeren Personen in einem größeren und klareren Bild zu sehen sind. Die Personen, die der Anwender noch sehen will, obwohl sie für ihn im Augenblick nicht so bedeutsam sind, kann er nach wie vor sehen.

[0057] Der Anzeigecontroller **52** bewirkt, dass jeder einzelne Videostrom, falls mehr als ein Videostrom vorhanden ist (wenn ein Konferenzgespräch stattfindet), nebeneinander auf dem Bildschirm **54** er-

scheint. Die nebeneinander auf der Anzeige **54** dargestellten Bilder werden beschnitten und nicht verkleinert, so dass die Abmessungen der Objekte in der Szene selbst nicht verändert werden. Es werden nur die Außenbereiche zu beiden Seiten der Szene, die zu jedem Datenstrom gehören, entfernt. Auf Wunsch können die Bilder von Strömen, die zu kleineren Szenenbildern gehören, nebeneinander in der unteren rechten Ecke des Anzeigebildschirms **54** dargestellt werden. Der Anzeigecontroller **52** liefert digitales Standardvideo für den LCD-Controller **72**, siehe [Fig. 9](#). Der Anzeigecontroller **52**, der von ATI oder Nvidia hergestellt wird, ist ein Standard-VGA-Controller. Der LCD-Controller **72** nimmt das digitale Standardvideo vom Anzeigecontroller **52** und passt das Bild an die besondere verwendete Anzeige an, beispielsweise eine Anzeige von Phillips oder Fujitsu.

[0058] Zum Verbessern der Bildbeschneidung wird, anstatt einfach Teile des Bilds abzutrennen, die an der Außenkante beginnen und sich zur Mitte zu bewegen, der Abschnitt des Bilds entfernt, der keine bedeutsame Information zeigt. Erscheint die gerade sprechende Person in der linken oder rechten Seite des Bilds, so wünscht man die linke Seite zu beschneiden, falls sich die Person auf der rechten Bildseite befindet, bzw. die rechte Seite, falls sich die Person auf der linken Bildseite befindet, und nicht einfach von jeder Außenkante nach innen zu schneiden, denn dies könnte dazu führen, dass ein Teil der Person verloren geht. Mit Hilfe der Objektverfolgung wird das Bild untersucht, das aufgebaut wird, und geprüft, wo Veränderungen im Bild auftreten, um eine Person zu erkennen, die im Bild vorkommt. Dabei wird unterstellt, dass sich eine Person bezogen auf die anderen Bereiche des Bilds stärker bewegt. Durch das Erkennen der Relativbewegung kann man den Ort der Person im Bild feststellen. Mit Hilfe dieser Objektverfolgung kann man bewirken, dass das Beschneiden an der Kante oder den Kanten erfolgt, an denen die geringsten Veränderungen auftreten. Wahlweise oder zusammen mit der Objektverfolgung kann man auch eine Audioverfolgung dazu verwenden, die vorgenommene Bildbeschneidung zu führen. Da das Videophon **15** ein Mikrofonfeld aufweist, kann man mit Hilfe von normalen Triangulierungsverfahren, die auf den unterschiedlichen Zeitspannen beruhen, die der Schall zum Erreichen der verschiedenen Elemente des Mikrofonfelds benötigt, feststellen, wo sich die Person relativ zum Mikrofonfeld befindet. Da man den Ort des Mikrofonfelds relativ zur abgebildeten Szene kennt, kennt man auch den Ort der Person im Bild.

[0059] Die Funktionen des Videophons **15** werden über einen Berührungsbildschirm **74** auf dem Monitor gesteuert. Der Berührungsbildschirm **74** – es handelt sich um einen normalen Berührungsbildschirm aus Glas – liefert dem Berührungsbildschirmcontroller **76** Rohsignale. Die Rohsignale werden bekanntlich anhand der Ultra-

schallwellen erfasst, die auf dem Glas erzeugt werden, wenn der Benutzer das Glas an einem gegebenen Ort berührt. Der Berührungsbildschirmcontroller **76** nimmt die Rohsignale und wandelt sie in aussagekräftige Information hinsichtlich der X- und Y-Position auf der Anzeige um. Diese Information wird an den Hauptcontroller **50** übertragen.

[0060] Ist ein Fernseh- oder Videorekorderanschluss verfügbar, so wird das eingespeiste Fernseh- oder Spielfilmsignal an den Decodierer **34** gegeben. Dort wird das Signal wie jedes beliebige andere Videosignal behandelt, das das Videophon **15** empfängt. Das Fernsehprogramm oder der Spielfilm kann neben einer Szene vom Videoanschluss zu einem anderen Videophon **15** auf dem Bildschirm **54** erscheinen.

[0061] Der Audiostrom der Szene folgt im Wesentlichen einem parallelen oder vergleichbaren Pfad wie der Audio-Videostrom. Der Audiostrom wird jedoch von einem Audioempfänger **58**, beispielsweise einem Mikrofon, einer Soundkarte, einer Kopfgarnitur oder einem Hörer, an eine CS Crystal 4201 Audioschnittstelle **60** oder eine ähnliche Vorrichtung als Codec geliefert, der Analog/Digital- und Digital/Analog-Umsetzungen der Signale vornimmt und auch die Lautstärke und das Mischen steuert und das Audiosignal digitalisiert und es an einen DSP **62** des Typs TCI 320C6711 oder 6205 liefert. Der DSP **62** unterteilt den digitalisierten Audiostrom in Pakete und überträgt ihn an das FPGA **38**. Das FPGA **38** überträgt seinerseits den digitalisierten Audiostrom an die PCI-Schnittstelle **44**. Dort wird er für die Übertragung auf dem Netz **40** an die Netzschnittstellenkarte **56** weitergeleitet. Der Audiostrom, den das Empfangsvideophon **15** annimmt, wird an das FPGA **38** und an den DSP **62** übermittelt und an die Audioschnittstelle **60** gegeben, die das digitale Signal in ein analoges Signal umsetzt, das auf den Lautsprechern **64** wiedergegeben wird.

[0062] Die Netzschnittstellenkarte **56** versieht jedes Audiopaket und Videopaket, das über das Netz **40** übertragen wird, mit einer Zeitangabe. Die Geschwindigkeit, mit der die vom Videophon **15** empfangenen Audio- und Videosignale verarbeitet werden, ist so schnell, dass das menschliche Auge und Ohr beim Anhören keinerlei Fehlausrichtung der Audiosignale zum zeitlich zugeordneten Video der Szene erkennen kann. Die Verarbeitung der Audio- und Videoinformation der Szene ist auf 20–30 Millisekunden beschränkt, damit die Zuordnung der Audio- und Videosignale der Szene erhalten bleibt. Um sicherzustellen, dass die Audio- und Videosignale der Szene beim Empfang am Empfangsvideophon **15** synchron sind, wird die Zeitangabe eines jeden Pakets untersucht. Das Empfangsvideophon **15** richtet zugehörige Audiopakete und Videopakete aus und gibt sie zusammen im Wesentlichen zeitgleich wieder, damit

der Benutzer am Empfangsvideophon **15** keine Fehlausrichtung der Audio- und Videosignale der Szene erkennen kann.

[0063] Eine ENC-DSP-Platine enthält den eNV 420 MPEG-2-Codierer von IBM und Hilfsschaltungen, den DSP **62** für die Audiocodierung und -decodierung und die PCI-Schnittstelle **44**. Sie enthält die Hardware, die für die vollen Terminalfunktionen des Videophons **15** bei einem System **10** mit einer PC **68** Hochleistungsplattform und einem Bildschirm **54** nötig ist. Der Entwurf ist vollständig mit PCI 2.2 verträglich. Die Platine weist passende Schnittstellen für Kamera, Mikrofone) und Lautsprecher **64** auf. Der DSP **62** führt die Audiocodierung, Decodierung, Mischung, Stereoanordnung, Pegelsteuerung, Pausenfüllung, Paketierung und weitere Audiofunktionen, beispielsweise Stereo-AEC (AEC = Automatic Echo Cancellation), Strahlausrichtung, Rauschunterdrückung, Unterdrückung von Tastenklacks und Hallunterdrückung aus. Das FPGA **38** wurde mit den Celoxia-Werkzeugen (Handel-C) entwickelt und ist vollständig umkonfigurierbar. Das Layout unterstützt Teile im Bereich von 1–3 Millionen Gatter.

[0064] Die Platine enthält eine Chipschnittstelle für die Digitalkamera **47**, eine auf Hardware oder "Video-DSP" beruhende Mehrkanal-Schnittstelle für den Videodecodierer **34**, Videoüberlagerung mit Hilfe der DVI-Eingangs- und Ausgangsstecker bis hin zur vollständigen simplen Frame-Pufferfähigkeit mit Videoüberlagerung.

[0065] Verwendet man ein NTSC- oder PAL-Video-signal, so sollte der Codierer **36** einen hochwertigen Videostrom mit einer Auflösung von 640 × 480 und bevorzugt von 720 × 480 oder mehr erzeugen. Die Bitrate sollte gesteuert werden, damit die Höchstanzahl Bits je Frame begrenzt wird und eine Übertragungsverzögerung über das Netz **40** verhindert wird. Der Decodierer **34** muss mit dem Decodieren eines Teils beginnen, wenn er den ersten Makroblock an Daten empfängt. Eine gewisse Pufferung kann erforderlich sein, damit man geringere Störungen ausgleichen kann und damit ein verbessertes Bild erhält.

[0066] MPEG-2 ist weit verbreitet und wird im großen Stil eingesetzt. Es ist die Grundlage der DVD- und VCD-Codierung, (VCD = Video-CD) von digitalen Videorekordern und Zeitverschiebungsvorrichtungen, beispielsweise TiVo (Television In-/Output, Festplatten-Videorekorder), und auch von DSS (DSS = Digital Slow Scan) und weiteren digitalen TV-Distributionen. In der Regel wird man eine Videoübertragung mit 4 bis 50 Mbit/sec wählen. Aufgrund seiner weiten Verbreitung sind relativ preisgünstige hochintegrierte Lösungen für das Decodieren und neuerdings für das Codieren im Handel erhältlich.

[0067] MPEG-2 sollte man als Syntax für codiertes

Video und weniger als Standardkompressionsverfahren ansehen. Die Spezifikation legt die Syntax und die Codierverfahren fest. Beim Gebrauch der Verfahren hat man jedoch einen sehr großen Spielraum, solange man der definierten Syntax folgt. Aus diesem Grund sind Verallgemeinerungen zu MPEG-2 häufig irreführend oder ungenau. Man muss tiefer in die Einzelheiten der jeweiligen Codierverfahren und der beabsichtigten Anwendung eindringen, um die Leistungsfähigkeit von MPEG-2 für eine besondere Anwendung zu bewerten.

[0068] Für das Videophon-Projekt **15** sind die Themen Codierung und Decodierung mit geringer Verzögerung von Interesse und ebenso die auf das Netz **40** bezogenen Fragen. Im MPEG-2-Algorithmus gibt es drei Hauptprobleme, die man verstehen muss, wenn man hochwertiges Video mit geringer Verzögerung über ein Netz **40** erzielen will:

- die GOP-Struktur (GOP = Group Of Pictures, Bildgruppe) und ihre Auswirkung auf die Verzögerung;
- die Auswirkung der Bitrate, der Größenänderung der codierten Frames und des VBV-Puffers (VBV = Video Buffering Verifier) auf die Verzögerung und die Anforderungen an das Netz **40**;
- Die Auswirkung der GOP-Struktur auf die Qualität bei Paketverlusten.

GOP-Struktur und Verzögerung

[0069] MPEG-2 definiert drei Arten von codierten Frames, nämlich I, P und B. Die am häufigsten verwendete GOP-Struktur besteht aus 16 Frames: IPBBPBBPBBPBBPBB. Das Problem bei dieser Struktur besteht darin, dass jeder fortlaufende B-Frame, weil ein B-Frame aus dem vorhergehenden und dem nachfolgenden Frame durch Bewegungsschätzung entsteht, erfordert, dass die folgenden Frames erfasst werden, bevor das Codieren des B-Frames beginnen kann. Da jeder Frame 33 Millisekunden dauert, kommen im Vergleich zu einer GOP-Struktur ohne B-Frames bei dieser Struktur mindestens 66 Millisekunden zusätzliche Verzögerung hinzu. Damit gelangt man zu einer GOP-Struktur mit geringer Verzögerung, die nur I- und/oder P-Frames enthält und in der MPEG-2-Spezifikation als SP@ML-Codierung (Simple Profile) definiert ist.

Bitrate, Größe der codierten Frames und VBV

[0070] Hat man die B-Frames beiseite gelassen, um die Codiervverzögerung so gering wie möglich zu halten, so besteht die GOP aus I-Frames und P-Frames, die auf die I-Frames bezogen sind. Da ein I-Frame vollständig innenbildcodiert ist, benötigt man sehr viele Bits hierfür, und weniger Bits für die folgenden P-Frames.

[0071] Man beachte, dass ein I-Frame 8 Mal so

groß sein kann wie ein P-Frame und 5 Mal so groß wie die nominale Bitrate. Dies wirkt sich direkt auf die Anforderungen an das Netz **40** und auf die Verzögerung aus. Besteht eine Bandbreitenbegrenzung, so wird der I-Frame an der Engstelle des Netzes **40** gepuffert. Dies führt zu einer zusätzlichen Verzögerung von mehreren Frame-Zeitspannen für den Transport über den begrenzten Abschnitt. Dieser Puffer muss an den Empfänger angepasst werden, da die Entleerrate durch das Video vorgegeben wird und nicht durch die Bandbreite des Netzes **40**. Das Probebild, das für die obigen Daten verwendet wurde, war eine Büroszene mit wenig Bewegung. Bei einem stark bewegten Inhalt mit Szenenänderungen werden den Frames abhängig vom Inhalt mehr oder weniger Bit zugewiesen, und bei Szenenänderungen treten einige große P-Frames auf.

[0072] Zum Steuern dieses Verhaltens implementiert MPEG-2 den VBV-Puffer (VBV = Video Buffering Verifier), der einen gewissen Grad an Einfluss auf das Verhältnis zwischen der maximalen codierten Framegröße und der nominalen Bitrate erlaubt. Durch eine strenge Begrenzung des VBV, durch die die I-Frames auf weniger als das 2fache der Größe begrenzt werden, die die nominale Bitrate angibt, kann man die zusätzliche Pufferungsverzögerung auf eine zusätzliche Framedauer beschränken. Das Beschränken der VBV-Größe geht zu Lasten der Bildqualität. Große I-Frames nimmt man als gute Grundlage für die folgenden P-Frames. Bei kleineren Bitraten (kleiner als 4 Mbit) nimmt die Qualität stark ab, wenn man die Größe der I-Frames beschränkt. Bei 2 Mbit ist beispielsweise die mittlere Framegröße 8 Kbyte. Auch die doppelte Größe reicht nicht aus, ein 320×240 JPEG-Bild mit guter Qualität zu codieren, das ähnlich wie ein I-Frame DCT-komprimiert ist.

[0073] Geht man dazu über, nur I-Frames zu codieren, so erlaubt dies eine gleichmäßigere Größe der codierten Frames, die Qualität wird jedoch noch stärker beeinträchtigt. Das ausschließliche Codieren von I-Frames mit geringer Bitrate schöpft den größten Teil der Kompressionsfähigkeit des MPEG-2-Algorithmus nicht aus.

[0074] Die MPEG-2-Spezifikation definiert den CBR-Modus (CBR = Constant Bit Rate, konstante Bitrate) und VBR-Modus (VBR = Variable Bit Rate, veränderliche Bitrate) und erlaubt veränderliche GOP-Strukturen innerhalb eines Stroms. Der CBR-Modus ist dafür gedacht, eine gleichmäßige Anzahl Bits für jede GOP zu erzeugen, wobei bei Bedarf Fülldaten verwendet werden. VBR ist dafür gedacht, eine gleichmäßige Qualität zu erlauben, wobei eine Schwankung der Codierungsbandbreite zulässig ist. Damit kann der Strom für schwierig zu codierende Bereiche mehr Bits anfordern, solange dies durch geringere Bitraten bei einfacheren Abschnitten ausgeglichen wird. VBR kann mit einem Durchlauf oder mit

zwei Durchläufen implementiert werden. Variable GOP-Strukturen erlauben beispielsweise das Anbringen von I-Frames an Szenenübergangsgrenzen, damit sichtbare Artefakte durch die Kompression vermieden werden. Aufgrund der Forderung nach geringen Verzögerungen und der Notwendigkeit, etwas voraus zu sehen, um VBR oder variable GOPs zu implementieren, sind diese Vorgehensweisen für eine Anwendung im Videophon **15** nur wenig interessant.

[0075] Da die P- und B-Frames in einer üblichen GOP-Struktur vom I-Frame und den vorhergehenden P- und B-Frames abhängig sind, beeinflusst ein Datenverlust alle auf den Fehler folgenden Frames bis zum nächsten I-Frame. Dies beeinflusst auch die Wartezeit am Anfang, wenn beispielsweise in einem DSS-System **10** Kanäle umgeschaltet werden, wobei der Decodierer **34** auf einen I-Frame warten muss, bevor er mit der Wiedergabe eines Bilds beginnen kann. Aus diesem Grund muss man die GOP-Länge, die Struktur und die Bitrate auf die Anwendung und das Versorgungssystem **10** abstimmen. Im Fall einer Echtzeitzusammenarbeit über IP wird ein unzuverlässiges Transportprotokoll verwendet, beispielsweise RTP oder UDP (UDP = User Datagram Protocol), da ein verspätetes Paket wie ein verlorenes Paket behandelt werden muss, weil man die Verzögerungszeit nicht zur Verfügung hat, die für das Arbeiten mit einem zuverlässigen Protokoll mit Handshake und erneuter Übertragung erforderlich ist. Es wurden verschiedene Untersuchungen über die Auswirkung von Paketverlusten auf die Videoqualität vorgenommen. Die Ergebnisse zeigen, dass für übliche IPB-GOP-Strukturen ein einprozentiger Paketverlust zu 30 Prozent Frameverlust führt. Kürzere GOP-Strukturen und schließlich Ströme ausschließlich aus I-Frames (mit Qualitätsverlust) reduzieren diese Verluste etwas. Mit FEC-Techniken (FEC = Forward Error Correction) kann man bei Verlusten etwa Abhilfe schaffen. Ein Problem bei MPEG-2 ist jedoch sicherlich, dass es nicht sehr tolerant gegenüber Datenverlust ist.

[0076] Eine GOP-Struktur für die Framecodierung mit dem Namen Continuous P löst alle angesprochenen Probleme und liefert bei relativ geringen Bitraten eine exzellente Videoqualität für das Videophon **15**. Die Codierung Continuous P nutzt die Möglichkeit zur Innenbildcodierung von Makroblöcken eines Frames innerhalb eines P-Frames. Codiert man eine Pseudo-Zufallsmenge von Makroblöcken mit 16×16 Pixeln in jedem Frame und die anderen mit Bewegungscodierung, so werden Äquivalente zu I-Frame-Bits in jedem Frame verteilt. Implementiert man die Pseudozufalls-Makroblockauswahl so, dass sichergestellt ist, dass alle Blöcke häufig aktualisiert werden, so werden der Anlauf und die Szenenveränderungen in vernünftiger Weise behandelt.

[0077] IBM hat diesen Algorithmus für den Codierer

S420 implementiert und stellt die Vollbild-DCT-Aktualisierungsrate auf 8 Frames (3,75 Mal in der Sekunde). Die Ergebnisse für übliche Büro- und Konferenzinhalte sind ziemlich beeindruckend. Die Codiervverzögerung, die Größenänderung der codierten Frames und das Paketverlustverhalten sind für das Videophon **15** nahezu ideal. Eine Prüfung der codierten Probedbilder zeigt, dass für Szenenveränderungen und hochdynamischen Inhalt Artefakte durch den Codierer **36** offenkundig sind. Für den üblichen Inhalt aus sprechenden Köpfen bei einer Zusammenarbeit ist die Qualität aber sehr gut.

[0078] Audio in guter Qualität ist eine unerlässliche Vorbedingung für eine wirksame Kommunikation. Gute Qualität ist dabei definiert als Vollduplex, eine Bandbreite von 7 kHz (ein Telefon hat 3,2 kHz), ein Signal-Rausch-Abstand von mehr als 30 dB, und es dürfen keine wahrnehmbaren Echos, Begrenzungen oder Verzerrungen auftreten. Die Installation ist sehr einfach und kommt mit so wenig Kabeln wie möglich aus. Eine eingebaute Diagnose erkennt Schwierigkeiten und gibt an, wie sie zu beseitigen sind. Der Schall aus den Lautsprechern **64** ist frei von lauten Knall- und Dröhngeräuschen und zu hohen oder zu niederen Lautstärken.

[0079] Ein Audiosignal von fehlenden oder verspäteten Paketen kann ausgehend vom vorhergehenden Audiosignal "aufgefüllt" werden. Der Audiopuffer sollte ungefähr 50 Millisekunden umfassen und einen Kompromiss zwischen den Schwankungen des Netzes **40** und der zusätzlichen Audioverzögerung darstellen. Die derzeitige Paketgröße von 320 Abtastwerten oder 20 Millisekunden könnte verkleinert werden, um die Wartezeit beim Codieren und Decodieren zu verkürzen. 20 Millisekunden stellt jedoch eine Standarddatenlänge für RTP-Pakete dar.

[0080] Einige der im Weiteren beschriebenen Prozeduren sind in käuflichen Produkten verfügbar. Wegen der Kosten und der Integration werden sie jedoch auf einem DSP **62** implementiert. In einer anderen Ausführungsform kann ein zweiter DSP **62** die akustische Echounterdrückung anstelle nur eines DSP **62** vornehmen, der diese Funktion mit übernimmt.

[0081] Das Audiosystem **10** besitzt einen Sendeabschnitt und einen Empfangsabschnitt. Der Sendeabschnitt besteht aus den folgenden Bestandteilen.

Mikrofone

[0082] Eine der hauptsächlichen Beschwerden über den Lautsprecherklang betrifft den hohlen Klang, den man am anderen Ende hört. Dieser hohle Klang entsteht durch den Raumnachhall und wird am besten durch das Verhältnis der reflektierten (zurückgeworfenen) Schallleistung zur direkten Schallleistung beschrieben. Das derzeit beste Verfahren zum Verbes-

sern der Aufnahme besteht darin, die Mikrofone nahe am Sprecher anzuordnen und damit die direkte Schallleistung zu erhöhen. In einer Büroumgebung kann man Mikrofone am Monitor des PC **68**, am Terminal des Videophons **15** und an einer Präsentationsstafel anbringen.

Automatische Verstärkungsregelung

[0083] Die Verstärkung des Vorverstärkers eines jeden Mikrofons wird automatisch so eingestellt, dass man den ADC-Bereich vollständig ausnutzt. Die Verstärkung des Vorverstärkers muss an andere Audio-prozeduren übermittelt werden, beispielsweise AEC und Rauschminderung.

CODEC

[0084] In seiner einfachsten Form handelt es sich um eine ADC-Vorrichtung. Einige Firmen, etwa Texas Instruments und Analog Devices Inc haben CODECS mit analogen Verstärkern und analogen Multiplexern. Auf dem Chip befindet sich auch ein D/A-Umsetzer mit ähnlichen Steuerungen. Die im vorhergehenden Abschnitt beschriebene automatische Verstärkungsregelung ist im CODEC implementiert und wird vom DSP **62** gesteuert.

Rauschminderung

[0085] Zum Verbessern des Signal/Rausch-Verhältnisses kann man zwei Rauschminderungsverfahren einsetzen. Das erste Verfahren wird allgemein als Störaustastung bezeichnet. Dabei schaltet man den Kanal abhängig vom Pegel des vorhandenen Signals ein und aus. Das zweite Verfahren ist die adaptive Rauschunterdrückung (ANC, ANC = Adaptive Noise Cancellation), die unerwünschtes Rauschen aus dem Mikrofonsignal entfernt. In einer Büroumgebung könnte man ANC dazu verwenden, verstärkte Durchsagen, Gebläsegeräusche und in machen Fällen auch Tastengeräusche zu entfernen.

[0086] Rauschminderungs- und Austast-Algorithmen sind in kommerziellen Audioeditor-Paketen verfügbar, beispielsweise Cool Edit und Goldwave, die auch Sondereffekte anwenden können, Kratz- und Knackgeräusche aus Aufnahmen und auch Zischen aus Bandaufzeichnungen entfernen.

Unterdrückung akustischer Echos

[0087] Ein Echo wird wahrgenommen, wenn die Stimme des Sprechenden nach mehr als 50 Millisekunden zum Sprecher zurückkommt. Echos sind sehr störend und müssen deshalb entfernt werden. Die zwei Quellen für Echos sind Leitungsecho und akustisches Echo. Das Leitungsecho entsteht durch die Eigenschaften eines Zweileitungs-Telefonsystems **10**. Die PSTN entfernt dieses Echo mit Hilfe ei-

nes Leitungsechounterdrückers (LEC, LEC = Line Echo Canceller). Verwendet man ein System **10** mit Telefonhörer, so treten akustische Echos zwischen dem Telefonhörer und dem Mikrofon auf. Der Schall des Sprechers am entfernten Ort wird vom entfernten Mikrofon aufgenommen und zum Sprecher zurückgeführt. Das Unterdrücken von akustischen Echos (AEC) ist schwieriger als LEC, da die Raumakustik schwerer zu modellieren ist und sich schlagartig verändern kann, wenn sich Personen bewegen. Es gibt zahlreiche AEC-Produkte, die von Einzelgeräten wie ASPI EF1210 bis zu Objektmodulen von Signal Works reichen, die dafür optimiert sind, auf Plattformen von der Art des DSP **62** zu laufen.

Automatisches Mischen

[0088] Automatisches Mischen bedeutet die Auswahl, welche Mikrofonsignale zu mischen sind, und das Senden des einkanaligen Mischerausgangssignals an den Codierer **36**. Die Auswahlkriterien beruhen darauf, das Mikrofon zu verwenden, das sich in der Nähe der lautesten Quelle befindet, oder die Mikrofone zu benutzen, die Schall empfangen, der oberhalb eines Grenzpegels liegt. Automatische Mischer sind von verschiedenen Anbietern im Handel erhältlich und werden in Telekonferenzsystemen und Teleausbildungssystemen eingesetzt.

Codierung

[0089] Zum Reduzieren der Datenübertragungs-Bandbreite wird das Audiosignal auf eine geringere Bitrate komprimiert, indem man die üblichen Signaleigenschaften und die menschliche Sprachwahrnehmung ausnutzt. Derzeit bietet der Codec G.722 die beste Audioqualität (7 kHz Bandbreite bei 14 Bit) bei einer vernünftigen Bitrate von 64 kBit/sec.

RTP-Übertragung

[0090] Die codierten Audiodaten werden in Abschnitte von 20 Millisekunden unterteilt und als Real-Time Protocol-Pakete (RTP) gesendet. RTP wurde insbesondere für den Echtzeit-Datenaustausch entwickelt, der für VoIP (VoIP = Voice over IP) und Telefonanwendungen nötig ist.

[0091] Der Empfangsabschnitt besteht aus den folgenden Bestandteilen.

RTP-Empfang

[0092] RTP-Pakete, die Audioströme von einem oder mehreren entfernten Orten enthalten, werden in ihren jeweiligen Puffern abgelegt. Fehlende oder verspätete Pakete werden erkannt. Diese Information wird an den Gap Handler weitergeleitet. Pakete in der falschen Reihenfolge sind ein Sonderfall der verspäteten Pakete, und wie verspätete Pakete werden sie

wahrscheinlich verworfen. Wahlweise kann man einen Puffer vorsehen, der die Wiedergabe des Audiosignals um mindestens eine Paketlänge verzögert. Die Puffergröße muss beschränkt werden, damit die Gesamtverzögerung nicht mehr als 100 Millisekunden beträgt.

Decodierung

[0093] Der G.722-Audiostream wird zu PCM-Abtastwerten für den CODEC decodiert.

Behandlung von Lücken

[0094] In jedem Netz werden RTP-Pakete verloren oder beschädigt. Daher "füllt" der Gap Handler die fehlenden Daten abhängig vom Spektrum und der Statistik der vorhergehenden Pakete auf. Es sollten zumindest Nullen in den Datenstrom eingebettet werden, um Daten zu erzeugen. Man kann aber auch einen spektralen Interpolations- oder Extrapolationsalgorithmus zum Auffüllen der Daten verwenden.

Pufferung

[0095] Netzschwankungen erfordern eine Pufferung, damit eine kontinuierliche Audiowiedergabe möglich wird. Die Größe dieses Puffers (und damit die Verzögerungszeit) wird man wohl anhand eines Kompromisses zwischen der Kurzzeitschwankungs-Statistik und der Auswirkung der Verzögerung einstellen.

Steuerung der Abtastrate

[0096] Die nominale Abtastrate für das Terminal eines Videophons **15** beträgt 16 kHz. Es können jedoch geringfügige Differenzen auftreten, die bewältigt werden müssen. Sei angenommen, dass das Videophon **15** Nord mit exakt 16,001 kHz abtastet, das Videophon **15** Süd dagegen mit 15,999 kHz. Damit sammelt das Terminal Süd je Sekunde einen Abtastwert mehr, als es an den Lautsprecher ausgibt, und das Terminal Nord gerät in ein Defizit von gleicher Größe. Mit einer Langzeitstatistik im Empfangspuffer kann man den Unterschied in der Abtastrate erkennen, und der geeignete Interpolationsfaktor (für das Videophon **15** Nord) oder Dezimierungsfaktor (für das Videophon **15** Süd) kann berechnet werden.

Lautstärkeregelung

[0097] Das Einstellen der Lautstärke aus den Lautsprechern **64** erfolgt in der Regel durch die Zuhörer am entfernten Ort. Ein besserer Weg könnte darin bestehen, den Schall aus den Lautsprechern **64** automatisch abhängig davon einzustellen, wie laut er an den Mikrofonen im Raum ankommt. Andere Faktoren können einbezogen werden, beispielsweise das Hintergrundgeräusch und die persönlichen Vorlieben

des Zuhörers.

Stereoanordnung

[0098] Entfernte Sprecher an verschiedenen Stellen kann man im Hörfeld anordnen. D. h., eine Person vom Ort A wäre entsprechend von links zu hören, die Person vom Ort B von der Mitte und die Person vom Ort C von rechts. Durch diese Anordnung kann man leichter verfolgen, wer gerade spricht.

Lautsprecher

[0099] Die Qualität des Schalls hängt in gewissem Umfang von der Qualität der Lautsprecher **64** und von deren Gehäuse ab. In jedem Fall werden für das Terminal des Videophons **15** Lautsprecher **64** mit eingebautem Verstärker verwendet.

Unterscheidung

[0100] Derzeit gängige Konferenzsysteme wie etwa Polycom Soundstation bieten eine zufriedenstellende Vollduplex-Audioqualität, die jedoch bandbegrenzt ist. Die Bandbreite ist auf 3500 Hz begrenzt, und die sich ergebende Klangqualität beansprucht das Gehör insbesondere beim Unterscheiden von Reibelauten.

[0101] Das Videophon **15** erweitert die Bandbreite auf 7 kHz und mischt mehrere Mikrofone automatisch, damit der Raumhall so gering wie möglich wird. Sprechen drei oder mehr Personen, so wird jeder Teilnehmer am fernen Ort an einen unverwechselbaren Platz im Stereoklangfeld gesetzt. Zusammen mit der hochwertigen Audioaufnahme und der erhöhten Bandbreite empfindet man eine Konferenz über das Netz **40** rasch so, als wenn man persönlich zusammenträfe.

[0102] Das Audiosystem **10** verwendet mehrere Mikrofone für eine bessere Schallaufnahme und einen Breitbandcodierer (G.722), damit die Klangtreue besser ist als bei derzeit angebotenen gebührenpflichtigen Systemen. Zusätzlich wird für Konferenzen mit mehreren Teilnehmern eine Stereoplatzierung der entfernten Sprecher implementiert und ein System **10** zum Auslöschen akustischer Echos, das einen freihändigen Betrieb erlaubt. Die Lautstärkeeinstellung im Raum wird automatisch gesteuert, wobei der Endanwender mit einer einzigen Einstellung den Gesamtlautstärkepegel vorgeben kann.

[0103] Im Netz **40** der Videophone **15** verbindet ein Gateway **70** Nicht-SIP-Teile mit der SIP-Umgebung. Oft liegen elektrische Unterschiede und Differenzen beim Protokoll vor. Die meisten Gateways **70** verbinden andere Telefon- oder Videokonferenzvorrichtungen mit dem System **10** der Videophone **15**.

[0104] Die Gateways **70** unterscheiden sich durch die Schnittstellen. Auf einer Seite befindet sich ein Netz **40**; für ein Videophon **15** ist es ein Ethernet oder ATM. Die Außenseite kann eine analoge Telefonleitung oder ein RS-232-Port sein. Die Gateways **70** unterscheiden sich durch Typ, Anzahl und Eigenschaften der Ports. Auf der Seite des Netzes **40** werden Transportprotokolle wie RTP oder AAL2 verwendet, und Signalisierungsprotokolle wie SIP, Megaco oder MGCP.

[0105] Auf der Außenseite können abhängig von den vorhandenen Schnittstellen zahlreiche verschiedene Protokolle eingesetzt werden. Einige Beispiele wären ISDN (Q.931) oder eine POTS-Signalisierung (POTS = Plain Old Telephone Service). PSTN-Gateways **70** verbinden vor Ort PSTN-Leitungen in die Videophone **15** des Systems **10**. PBX-Gateways **70** (PBX = Private Branch Exchange, Nebenstellenanlage) erlauben es den Videophonen **15** des Systems **10**, ein Privattelefon zu emulieren, damit die Kompatibilität zu vor Ort vorhandenen PBX gegeben ist. POTS-Gateways **70** verbinden simple Analogtelefone mit den Videophonen **15** des Systems **10**. H.323-Gateways **70** verbinden ein H.323-System **10** mit den SIP-basierten Videophonen **15** des Systems **10**. Dabei handelt es sich um einen Gateway **70** ausschließlich für die Signalisierung. Die Umsetzung von H.261 nach MPEG erledigt der Medienserver **66**.

[0106] Drei mögliche Technologien für das Videophon **15** sind das Session Initiation Protocol (SIP), das Session Description Protocol (SDP) und das Realtime Transport Protocol (RTP).

[0107] SIP ist ein Signalisierungsprotokoll zum Aufbauen, Verwalten und Beenden von Sprach- und Videositzungen über Paketnetze.

[0108] SDP ist zum Beschreiben von Multimediasitzungen gedacht und dient der Sitzungsankündigung, der Einladung zur Sitzung und anderen Formen des Multimedia-Sitzungsaufbaus. SIP verwendet SDP zum Beschreiben von Mediensitzungen.

[0109] RTP liefert Transportfunktionen vom einen Ende des Netzes **40** zum anderen und ist geeignet für Anwendungen, die Echtzeitdaten übertragen, beispielsweise Audio-, Video- oder Simulationsdaten, und zwar über Multicast- oder Unicastdienste des Netzes **40**. SIP verwendet RTP für den Mediensitzungstransport.

[0110] Das Videophon **15** kann Konferenzen mit drei oder mehr Teilnehmern abwickeln, ohne irgendeine Konferenzbrücke oder MCU zu verwenden. Dies wird durch den Gebrauch von ATM-Point zu Multipoint-Strömen erreicht, die SIP aufbaut. Wird im Einzelnen der MPEG-2-Strom und der Strom mit der geringen Framerate für die Übertragung auf dem

Netz **40** in Pakete zerlegt, so nennt der Kopfeintrag eines jeden Pakets bekanntlich die Adressen aller Empfangsvideophone **15** der Konferenz. Werden die Pakete über das Netz **40** übertragen, so erzeugt SIP aus dieser Information die notwendige Verbindungsinformation für die verschiedenen Pakete, damit diese ihre gewünschten Bestimmungsvideophone **15** erreichen.

[0111] Als Beispiel für eine Konferenz, bei der keinerlei Konferenzbridges verwendet werden, seien zehn Videophone **15** an verschiedenen Orten die Teilnehmer an einer Konferenz. Jedes Videophon **15** erzeugt einen Audiostrom, einen MPEG-2-Strom und einen Strom mit geringer Framerate. Da keines der Videophone **15** irgendeinen dieser Ströme zurück an sich selbst sendet, kommuniziert in einer Konferenz der Videophone **15** mit zehn Teilnehmern jedes Videophon effektiv mit neun anderen Videophonen **15**. Da es vorkommen könnte, dass ein Videophon **15** mit sich selbst kommuniziert, kann zum Maximieren der Bandbreitenausnutzung das von irgendeinem Videophon **15** erzeugte Video und auf Wunsch das von einem Videophon **15** erzeugte Audio im Wesentlichen so gesehen oder gehört werden, wie es auf den anderen Videophonen **15** erscheint, jedoch über einen internen Kanal, der im Weiteren beschrieben wird und keinerlei Bandbreite des Netzes **40** verbraucht.

[0112] In der Konferenz empfängt jedes Videophon **15** neun Audiodatenströme sowie drei MPEG-2-Datenströme und sechs Datenströme mit geringen Frameraten. Auf Wunsch kann der Empfänger bis zu neun Ströme mit geringen Frameraten auswählen, so dass die Anzeige **54** nur die kleineren Bilder eines jeden Videophons **15** darstellt. Der Empfänger kann auch bis zu vier MPEG-2-Datenströme wählen. Dabei wird die Anzeige **54** mit vier Bildern von vier Videophonen **15** der Konferenz gefüllt. Es werden keine Bilder von Datenströmen mit geringen Frameraten gezeigt, da für sie kein Platz auf der Anzeige **54** vorhanden ist, wenn vier MPEG-2-Ströme wiedergegeben werden. Stellt man drei MPEG-2-Ströme dar, so kann man noch sechs Ströme mit geringen Frameraten wiedergeben. Jeder Strom ist wie oben beschrieben aufgebaut und wird wie beschrieben an den verschiedenen Videophonen **15** empfangen.

[0113] Wird gewünscht, mehr als vier große Bilder einer Konferenz darzustellen, so erreicht man dies dadurch, dass zusätzliche Videophone **15** so miteinander verbunden werden, dass die Bildschirme der einzelnen Videophone **15** nebeneinander aufgereiht sind, siehe [Fig. 7](#). Ein Videophon **15** kann als Master arbeiten, und jedes zusätzliche Videophon **15** wird beim Anschließen zum Slave des Master-Videophons **15**, das die Darstellung der großen und kleinen Bilder auf den Anzeigen **54** der einzelnen Videophone **15** steuert.

[0114] Als Protokoll zum Festlegen, welche Person als Großbild und welche Person als Kleinbild auf den Bildschirmen der Konferenzvideophone **15** dargestellt wird, besteht ein bevorzugtes Protokoll darin, dass die drei letzten Sprecher groß dargestellt werden und alle weiteren Teilnehmer klein abgebildet werden. D. h., die derzeit sprechende Partei und die beiden vorhergehenden Sprecher werden groß dargestellt. Da jedes Videophon **15** der Konferenz alle Audioströme der Konferenz empfängt, kann jedes Videophon **15** mit seinem Hauptcontroller **50** feststellen, wo zu einem gegebenen Zeitpunkt gesprochen wird und die Netzschnittstellenkarte **56** anweisen, den MPEG-2-Strom anzunehmen, der zu dem Videophon **15** gehört, von dem das Sprechen stammt, und den zugehörigen Strom mit der geringen Framerate nicht anzunehmen. In einem anderen Protokoll wird ein Videophon **15** als Führer- oder Moderatorvideophon **15** eingesetzt. Das Führevideophon **15** wählt aus, welche großen und kleinen Bilder jedes andere Videophon **15** sieht. In noch einem anderen Protokoll ist die Auswahl der Großbilder und Kleinbilder festgelegt und bleibt während der ganzen Konferenz erhalten. Das Protokoll kann darin bestehen, dass jedes Videophon **15** wählt, wie es die empfangenen Bilder darstellt. Es werden sowohl der MPEG-2-Strom als auch der Strom mit geringer Framerate auf dem Netz **40** zu den Empfangsvideophonen **15** der Konferenz transportiert. Damit sind beide Videoströme an jedem Empfangsvideophon **15** verfügbar und können abhängig vom gewählten Protokoll für die Anzeige **54** dargestellt werden.

[0115] Um die Bandbreite noch wirksamer zu nutzen und die Audioverarbeitung dadurch zu unterstützen, dass weniger Verarbeitungsanforderungen an jegliches Sendevideophon **15** oder Empfangsvideophon **15** ausgegeben werden, wird hinsichtlich der Audioströme, die jedes Videophon **15** überträgt, ein Audiostrom nur dann von einem Videophon **15** übertragen, wenn der Audiopegel am Sendevideophon **15** über einem vorbestimmten Dezibel-Grenzwert liegt. Unterstellt man, dass der Grenzwert so eingestellt wird, dass er beim Sprechen erreicht oder überschritten wird, so dass nur Audioströme übertragen werden, deren Schall ausreichend laut ist, so schließt dies nicht nur unbeteiligte Hintergrundgeräusche vom Senden und Empfangen aus, die im Wesentlichen nichts beitragen und nur Bandbreite verbrauchen, sondern es unterstützt auch die Auswahl des MPEG-2-Stroms, der zum Sprechen gehört, da nur Audioströme empfangen werden, auf denen gesprochen wird.

[0116] Wünscht ein gegebenes Videophon **15** wie erwähnt, sein eigenes Bild zu sehen, das an die anderen Videophone **15** ausgesendet wird, so wird der vom FPGA **38** gebildete Strom mit geringer Framerate an einen lokalen Speicher im Videophon **15** gesendet. Dabei erfolgt keinerlei Kompression, d. h. der

Strom mit geringer Framerate liegt genau so vor, wie er vom Videophon **15** in Pakete zerlegt und in das Netz **40** gesendet wird. Diesen lokalen Speicher bearbeitet der Hauptprozessor mit Software und bewirkt, dass er als Kleinbild auf dem Bildschirm **54** dargestellt wird.

[0117] Zudem steuert das Videophon **15**, welche der Audio- oder Videoströme, die es aus dem Netz **40** empfängt, zu hören bzw. zu sehen sind. In Situationen, in denen mehr Parteien an der Konferenz teilnehmen als der Benutzer des Videophons **15** zu sehen oder zu hören wünscht, kann der Benutzer des Videophons **15** wählen, nur eine Untermenge der Audio- oder Videoströme zu sehen oder zu hören, die in der Gesamtkonferenz enthalten sind. In einer Konferenz mit 100 Teilnehmern entscheidet der Benutzer, drei der Videoströme als Großbilder und 20 der Videoströme als Kleinbilder auf dem Bildschirm zu sehen, womit insgesamt 23 von 100 möglichen Bildern dargestellt werden. Der Benutzer des Videophons **15** wählt, dass er die drei lautesten Sprecher als Großbilder sehen will. Anschließend wählt er über den Berührbildschirm, dass 20 der Konferenzparteien, die auf einer Seite des Berührbildschirms aufgelistet sind, auch als Kleinbilder dargestellt werden. Es können auch andere Protokolle gewählt werden, beispielsweise dass die 20 Bilder, die als Kleinbilder dargestellt werden, die letzten 20 Sprecher in der Konferenz sind, und zwar ab dem Zeitpunkt, zu dem die Konferenz beginnt und jede Partei ihre einleitenden Worte spricht. Durch die Kontrolle der Anzahl der dargestellten Videoströme wird die Konferenz strukturiert und die Nutzung der Ressourcen des Videophons **15** wird besser zugewiesen.

[0118] Hinsichtlich der verschiedenen auf dem Bildschirm dargestellten Bilder kann jedem Bild ein Auswahlkriterium zugewiesen werden. Beispielsweise kann ein Moderator des Konferenzgesprächs ein Bild auswählen. Zwei der Bilder kann man von den derzeit letzten bzw. lautesten Sprechern der Konferenz abhängig machen. Das andere Bild kann man einer Person zuweisen, die der Benutzer aus allen anderen Konferenzteilnehmern auswählt. Auf diese Weise kann jeder Teilnehmer oder Benutzer der Konferenz potentiell eine unterschiedliche Auswahl der Bilder aus der Gesamtzahl der Konferenzteilnehmer sehen. Die größtmögliche dann benötigte Bandbreite muss einen Videostrom umfassen, der an das Netz gesendet wird, und vier Videoströme, die aus dem Netz empfangen werden, und zwar unabhängig von der Anzahl der Konferenzteilnehmer.

[0119] Hinsichtlich der Audioströme kann man am Videophon **15** die Beschränkung einführen, dass nur die Audioströme zum Hören ausgewählt werden, die zu den drei lautesten Sprechern gehören, während ihr zugehöriges Bild auf dem Bildschirm dargestellt wird. Der DSP **62** kann die empfangenen Audioströ-

me untersuchen und nur die Wiedergabe der drei Audioströme erlauben, die zu den lautesten Sprechern gehören. Gleichzeitig kann er die Netzschnittstelle **42** anweisen, nur die ersten Videoströme der Großbilder zu empfangen, die zu den drei Audioströmen gehören, die die lautesten Sprecher aufwiesen. Ganz allgemein betrachtet herrscht umso mehr Verwirrung und umso weniger Verständnis, je mehr Personen gleichzeitig sprechen. Daher übt der Benutzer eine gewisse Kontrolle über die Audioströme aus und organisiert sie in bestimmter Weise.

[0120] Als Teil der Steuerung der Audioströme sendet jedes Videophon **15** wie erwähnt nur dann einen Audiostrom aus, wenn das Geräusch in der Nähe des Videophons **15** über einem Grenzwert liegt. Der Grenzwert ist bevorzugt dynamisch und hängt vom Geräuschpegel der drei lautesten Audioströme ab, die zu einem gegebenen Zeitpunkt zu den drei lautesten Sprechern gehören. Dies folgt, weil für den Audiostrom, der als einer der Audioströme mit den drei lautesten Sprechern eingestuft wird, die Geräuschpegel anderer Audioströme überwacht und hinsichtlich ihres Geräuschpegels erkannt werden müssen. Der DSP **62** überprüft nach dem Empfang der Audioströme über das Netz **40** aus der Netzschnittstelle **42** den Audiostrom und erkennt die drei Ströme mit dem lautesten Geräusch. Der DSP vergleicht auch die Geräuschpegel der drei empfangenen Audioströme, die den drei lautesten Sprechern zugeordnet sind, mit dem Geräuschpegel der Szene in der Nähe des Videophons **15**. Ist der Geräuschpegel aus der Szene in der Nähe des Videophons **15** größer als irgendeiner der empfangenen Audioströme, so sendet das Videophon **15** seinen Audiostrom an das Netz **40**. Diese Art der unabhängigen Untersuchung durch den DSP **62** erfolgt in jedem Videophon der Konferenz und stellt damit eine über die ganze Konferenz verteilte Analyse dar. Jedes Videophon untersucht unabhängig von allen anderen Videophonen seine empfangenen Audioströme selbst, die das entsprechende Videophon **15** per Definition nur dann aussendet, wenn es festgestellt hat, dass das Geräusch aus seiner Szene so laut ist, dass es zu einem gegebenen Zeitpunkt als eines der drei lautesten Geräusche gelten kann. Jedes Videophon **15** nimmt daraufhin diese empfangene Audiostrominformation und verwendet sie als Bezug für den Vergleich mit dem eigenen Geräuschpegel. Jedes Videophon **15** legt somit seinen Grenzwert selbst fest.

[0121] Eine andere Art, diese verteilte Analyse auszuführen, besteht darin, dass jedes Videophon, nachdem es mit dem DSP **62** den Grenzwert festgestellt hat, den es für richtig hält, diesen Grenzwert an alle anderen Videophone der Konferenz senden kann. Damit können alle Videophone nachprüfen, welchen Grenzwert alle anderen Videophone haben wollen, und können beispielsweise diese Grenzwerte mitteln und daraus einen Grenzwert festlegen, den sie für

ihre Szene verwenden.

[0122] Geht man so vor, dass man die Videoströme der drei lautesten Sprecher wählt, so können Augenblicke auftreten, in denen Teilnehmer zur gleichen Zeit laut zu sprechen beginnen und ein Durcheinander erzeugen, in dem man nichts verstehen kann. Dadurch steigt jedoch das Geräusch im Grenzwertpegel, wodurch in sehr kurzer Zeit die Audioströme beseitigt werden, die weniger Geräusch als andere erzeugen, so dass wiederum nur die Audioströme der drei gewichtigsten Sprecher gewählt und gehört werden und die anderen nicht gewählt werden. Dadurch wird ein Teil der Geräusche entfernt, den die anderen Audioströme beitragen könnten. Daraus folgt, dass es Zeiten geben kann, in denen das Videophon **15** mehr als drei Audioströme empfängt, da mehr als 3 Videophone einen Geräuschpegel haben können, der in diesem Augenblick über dem Grenzwert liegt. Daher darf jedes dieser Videophone zu diesem Zeitpunkt einen Audiostrom erzeugen und an das Netz **40** senden. Dieser Vorgang kommt jedoch wie eben erklärt nach der Änderung des Grenzwerts zum Stillstand. Die verteilte Analyse der Audioströme ist nicht wie hier beschrieben auf das Videophon **15** beschränkt, sondern auf jede beliebige Art von Audiokonferenz anwendbar, und zwar unabhängig davon, ob auch Videoströme vorhanden sind oder nicht.

[0123] Übereinstimmend damit, dass man das Schwergewicht auf das Einsparen von verwendeter Bandbreite legt und zum Sparen der Bandbreite nur erforderliche Beiträge sendet, erfolgt das Beschneiden eines Bilds am Codierer **36** und nicht am Empfangsvideophon **15**. In den Fällen, in denen das Sendevideophon **15** weiß, wie sein Bild auf den Empfangsvideophonen **15** erscheint, beschneidet der Codierer **36** das große Bild der Szene, bevor es gesendet wird. Dadurch sind wesentlich kleinere Anteile des Bilds zu senden, und es wird weniger Bandbreite verbraucht. Erfolgt das Beschneiden am Empfangsvideophon **15**, so bearbeitet der Hauptprozessor das empfangene Bild mit Software, bevor es an den Anzeigecontroller **52** gegeben wird.

[0124] Man kann eine zweite Kamera an das Videophon **15** anschließen, um eine andere Ansicht der Szene zu liefern. Beispielsweise kann man in einem Raum die erste Kamera oder Hauptkamera so anordnen, dass sie auf das Gesicht des Betrachters oder des Sprechers gerichtet ist. Es können sich jedoch weitere Personen im Raum aufhalten, die die Person, die das Videophon **15** im Raum bedient, den anderen Zuschauern an den Empfangsvideophonen **15** zeigen möchte. Die zweite Kamera kann beispielsweise in einer oberen Ecke des Raums angeordnet werden, so dass die zweite Kamera einen wesentlich größeren Abschnitt des Raums überblicken kann als die Hauptkamera. Die Signale der zweiten Kamera kann man in den Decodierer **34** eingeben. Der Decodierer

34 weist mehrere Ports für den Empfang von Videoeinspeisungen auf. Liegt der Strom von der zweiten Kamera bereits digitalisiert vor, so kann man ihn wahlweise den Verarbeitungselementen des Videophons **15** zuführen, und zwar über ähnliche Kanäle wie bei der Hauptkamera. Bevorzugt legt jedes Videophon **15** seine Aussendungen selbst fest, so dass die Wahl, welche Kameraeingaben übertragen werden, vom Betrachter getroffen werden, der das Videophon **15** bedient. Wahlweise kann man an ein entferntes Empfangsvideophon **15** die Fähigkeit delegieren, zu steuern und zu wählen, welcher Strom von welcher Kamera an einem gegebenen Videophon **15** zu übertragen ist. Die Steuersignale vom steuernden Videophon **15** werden über das Netz **40** übertragen und vom jeweiligen Videophon **15** empfangen, das dann den gewählten Strom für die Übertragung bereitstellt. Anstelle einer zweiten Kamera kann man über das Videophon **15** jedes beliebige andere Videosignal liefern, beispielsweise das Videosignal von einer DVD, einem VCR oder einer Präsentationstafel-Kamera.

[0125] In einer bevorzugten Ausführungsform arbeitet das Videophon **15** in einem Einblickmodus. Im Einblickmodus nimmt die Kamera des Videophons **15** ein Standbild der vor ihr liegenden Szene auf und überträgt dieses Bild an die anderen Videophone **15**, die vorher für den Empfang des Bilds gekennzeichnet worden sind, beispielsweise auf einer Liste der Videophone **15** im Kurzwahlmenü. Wahlweise wird im Einblickmodus das aufgenommene Standbild im Videophon **15** gehalten und auf Anforderung an jedermann geliefert, der dieses Videophon **15** für einen Anruf aussucht. Der Idealfall ist – da sich dies mit dem bevorzugten Gebrauch des Videophons **15** deckt – dass der Benutzer eines jeden Videophons **15** sämtliche Aussendungen des Videophons **15** kontrolliert, und dass er einfach wählen kann, den Einblickmodus auszuschalten oder das Bild festzulegen, das gesendet wird. Erfolgt ein aktiver Anruf, so wird der Einblickmodus ausgeschaltet. Dadurch tritt kein Konflikt zwischen dem Einblickmodus und dem aktiven Anruf auf, bei dem die Kamera einen kontinuierlichen Bildstrom aufnimmt. Im Einblickmodus kann das Standbild der Szene in vorbestimmten Zeitintervallen aufgenommen werden, beispielsweise jeweils nach einer Minute, nach fünf Minuten, nach 30 Minuten usw. Im Einblickmodus kann zu einem vorbestimmten Zeitpunkt vor der Aufnahme des Standbilds, beispielsweise fünf oder zehn Sekunden vor der Aufnahme des Bilds, eine hörbare Sequenz ausgegeben werden, die jede Person, die sich vor der Kamera befindet, darauf aufmerksam macht, dass gleich ein Bild aufgenommen wird und dass sie präsentabel aussehen sollte. Die hörbare Sequenz kann ein Piepton, ein Pfiff oder ein anderes aufgezeichnetes Geräusch oder eine Ansage sein. Auf diese Weise wird durch die Verwendung des Einblickmodus den anderen Videophonen **15** ein Einblick in die Sze-

ne vor der Kamera des Videophons **15** ermöglicht, und den anderen Videophonen **15** wird die Anwesenheit von Personen im Blickfeld der Kamera angezeigt.

[0126] Als weiteres Beispiel für einen Anwesenheitssensor kann der Ort der automatischen Linse der Kamera bezogen auf das Blickfeld vor der Kamera als Anwesenheitssensor dienen. Befindet sich niemand vor der Kamera, so fokussiert die automatische Linse der Kamera auf ein Objekt oder eine Wand in ihrem Blickfeld. Befindet sich eine Person vor der Kamera, so fokussiert die automatische Linse auf diese Person. Dadurch befindet sich die Linse in einer Position, die sich von der Position unterscheidet, in der sich keine Person vor der Linse befindet. Die Kamera kann ein Signal, das den Fokus der Linse angibt, an das FPGA **38** senden, das dann veranlasst, dass die Fokuginformation an eine vorbestimmte Liste von Empfängervideophonen **15** gesendet wird, beispielsweise die Videophone in der Kurzwahlliste des Sendevideophons **15**, damit die Empfängervideophone **15** informiert werden, ob sich der Betrachter vor dem Videophon **15** befindet und angezeigt wird, dass jemand anwesend ist.

[0127] Das Videophon **15** ist auch tauglich für Videomail. Wird ein Videoanruf von einem Videophon **15** zu einem anderen Videophon **15** versucht, und antwortet das Empfängervideophon **15** nach einer vorbestimmten Zeitspanne nicht auf den Videoanruf, beispielsweise nach viermaligem Läuten, so antwortet ein Videoserver **66**, der dem Empfängervideophon **15** zugeordnet ist, auf den Videoanruf. Der Videoserver **66** beantwortet den Videoanruf vom Sendevideophon **15** und sendet dem Sendevideophon **15** eine aufgezeichnete Audionachricht oder eine Audionachricht mit einem vorher aufgezeichneten Videobild vom Empfängervideophon **15**, das nicht geantwortet hat. Der Videoserver **66** gibt die Nachricht wieder und liefert dem Anrufer eine Audio- oder Audio- und Videosequenz, die ihn auffordert, seine Nachricht nach einem vorbestimmten Kennzeichen, beispielsweise einem Piepton, zu hinterlassen. Wird das vorbestimmte Kennzeichen ausgegeben, so hinterlässt der Anrufer eine Nachricht, die ein Audiostatement und ein Videobild des Anrufers enthält. Die Video- und die Audionachricht wird im Speicher am Videoserver **66** abgelegt. Die Nachricht kann beliebig lang sein, oder sie kann auf eine vorbestimmte Zeitspanne für die Nachricht begrenzt sein, die man festlegen kann. Nach Ablauf der vorbestimmten Zeitspanne oder wenn der Anrufer fertig ist und das Gespräch beendet hat, speichert der Videoserver **66** die Videonachricht und sendet ein Signal an das Empfängervideophon **15**, das den ursprünglichen Anruf nicht beantwortet hat, dass eine Videonachricht auf den Betrachter des Empfängervideophons **15** wartet. Diese Meldung kann ein Text sein oder ein Videobild, der bzw. das auf dem Bildschirm **54** des Empfänger-

videophons **15** erscheint. Es kann auch einfach eine Signallampe sein, die eingeschaltet wird und den Betrachter des Empfängervideophons **15** darauf aufmerksam macht, dass eine Videomail für den Betrachter vorliegt.

[0128] Will der Betrachter die Videomail ansehen, so braucht er nur den Bereich auf dem Berührbildschirm **74** zu wählen, der die Videomail aktiviert. Dem Benutzer wird eine Anzahl Optionen zum Behandeln der Mail angeboten, zu der auch das Lesen der Videomail gehört. Dadurch wird ein Signal an den Videoserver **66** gesendet, damit er die Videomail für den Betrachter am Bildschirm **54** des Videophons **15** wiedergibt. Der Bildstrom, den der Videoserver **66** sendet, folgt ebenfalls dem Weg, der oben für Videoströme beschrieben wurde, zum und durch das Empfängervideophon **15**, damit er dargestellt wird. Will der Bediener des Videophons **15** eine Nachricht auf dem Videoserver **66** aufzeichnen, mit der Videoanrufe beantwortet werden, falls der Bediener die Videoanrufe nicht beantwortet, so berührt der Benutzer eine Fläche auf dem Berührbildschirm **74**, durch die der Videoserver **66** den Bediener auffordert, zu einem vorbestimmten Zeitpunkt entweder eine Audionachricht oder eine Videonachricht und eine Audionachricht aufzuzeichnen. Dies tut der Benutzer dann, um die Nachricht zu erzeugen.

[0129] Das Videophon **15** bietet einen Betrieb der Lautsprecher **64** bei einem vorbestimmten Pegel, ohne dass der Benutzer irgendeine Lautstärke einstellen muss. Die Lautsprecher **64** des Videophons **15** können mit Hilfe des Mikrofons kalibriert werden. Nimmt das Mikrofon zu laute Geräusche auf, so senken der Hauptcontroller **50** und der DSP **62** den Pegel des Audioausgangssignals für die Lautsprecher **64**, damit der Geräuschpegel sinkt. Durch das Festsetzen eines vorbestimmten und gewünschten Pegels steuert das Videophon **15** automatisch die Lautstärke im Raum ohne Eingriff des Benutzers.

[0130] Man kann das Videophon **15** so programmieren, dass es eine Aufforderung zum Sprechen an eine bestimmte Person erkennt und daraufhin das vorbestimmte Sprachmuster, das für die Erkennung verwendet wird, als Ton oder Signal am Empfangsvideophon **15** einsetzt, damit der Betrachter am Empfangsvideophon **15** informiert wird, dass ein Gespräch mit dem Empfangsvideophon **15** angefordert wird. Beispielsweise kann der Ausdruck "Hey Craig" im Videophon **15** dafür verwendet werden, zu erkennen, dass ein Gespräch zwischen Craig und dem Sendevideophon **15** aufgebaut werden soll. Sagt der Betrachter "Hey Craig", so veranlasst er das Sendevideophon, automatisch ein Gespräch zu Craig aufzubauen, das dann den Ausdruck "Hey Craig" zum Empfangsvideophon **15** von Craig sendet. Um anzuzeigen, dass ein Gespräch mit Craig angefordert wird, läutet das Empfangsvideophon **15** von Craig

nicht, sondern am Videophon **15** von Craig wird anstelle des Klingeltons, der normalerweise ausgegeben wird, um Craig aufmerksam zu machen, der Ausdruck "Hey Craig" wiederholt ausgegeben. Die Funktionen zum Ausführen dieses Vorgangs übernehmen der Hauptcontroller **50** und der DSP **62**. Das Statement "Hey Craig" wird vom Betrachter angesagt und wie oben erläutert an den Server **66** übertragen. Der Server **66** erkennt durch eine Untersuchung der Statements den Ausdruck als einen Befehl, ein Gespräch mit der im Befehl genannten Partei aufzubauen. Der Server **66** verwendet nun die Adressinformation im Videophon **15** von Craig dazu, das Gespräch mit dem Videophon **15** von Craig aufzubauen, und bewirkt, dass am Videophon **15** von Craig der erzeugte Ton bzw. das erzeugte Signal "Hey Craig" ist.

[0131] Bekanntlich kann der Codierer **36** Beginn und Ende eines jeden Frames erkennen. Empfängt der Codierer **36** die Daten, so codiert er die Daten für einen Frame und speichert die Daten bis der Frame vollständig ist. Aufgrund des Algorithmus, den der Codierer **36** verwendet, dient der gespeicherte Frame als Grundlage zum Ausbilden des folgenden Frames. Der gespeicherte Frame dient als Bezugsframe für den nächsten zu codierenden Frame. Dies ist im Wesentlichen deshalb so, weil die Veränderungen zwischen einem Frame und dem folgenden Frame bei dieser Codierung im Mittelpunkt stehen und nicht der gesamte Frame von Beginn an. Der codierte Frame wird nun wie oben erklärt ohne jegliche Pufferung, ausgenommen für die Paketierung, direkt zur Paketierung gesendet, damit sämtliche Verzögerungen so klein wie möglich werden. Wahlweise kann der Codierer **36**, um die Übertragungsgeschwindigkeit der Daten weiter zu steigern, beim Codieren der Daten für den Frame die codierten Daten für die Paketierung anordnen, ohne darauf zu warten, bis der gesamte Frame codiert ist. Die codierten Daten werden zum Bilden des Frames auch gespeichert, siehe die oben erklärten Gründe, damit ein Bezugsframe für den Codierer **36** vorhanden ist. Getrennt davon werden die Daten beim Codieren an die Paketierung gesendet und beim Vorbereiten der Paketierung zu einem Frame zusammengesetzt. Ist das Paket fertig für die Übertragung, so kann es vorkommen, dass nur ein Teil des Frames im Paket enthalten ist. Der verbleibende Teil des Frames wird in einem eigenen Paket übertragen, und der Frame wird nicht aufgebaut, bevor beide Pakete mit der Frameinformation am Empfangsvideophon **15** angekommen sind.

[0132] Die Videophone **15** sind an das Netz **40** angeschlossen, siehe [Fig. 1](#). Die Videophone **15** unterstützen 10/100-Ethernetverbindungen und wahlweise ATM-Verbindungen mit 155 Mbps sowohl auf Kupferleitungen als auch auf Multimode-Lichtleitern. Jedes Terminal eines Videophons **15** ist in der Regel mit dem PC **68** eines Benutzers verbunden. Die Rolle des Videophons **15** besteht darin, die Audio- und Vi-

deoaspekte eines (Konferenz-)Gesprächs zu liefern. Der PC **68** wird für alle anderen Funktionen verwendet. Das Herstellen eines Gesprächs über die Videophone **15** kann automatisch eine Microsoft Netmeeting-Sitzung zwischen den zugeordneten PCs **68** aufbauen, so dass die Benutzer über Windows-gestützte Programme zusammenarbeiten können, beispielsweise eine Power-Point-Präsentation oder eine Tabellenkalkulation, dass sie Graphiken auf einer elektronischen Präsentationstafel austauschen können, Dateien übertragen können oder ein textbasiertes Simultandiskussionsprogramm verwenden können usw. Der PC **68** kann unabhängig davon, wie das Terminal des Videophons **15** angeschlossen ist, mit Ethernet verbunden sein. Er kann natürlich auch mit einem ATM-LAN verbunden sein. Der PC **68** und das zugeordnete Sendevideophon **15** kommunizieren über das Netz **40** miteinander. Der PC **68** und das zugeordnete Sendevideophon **15** kommunizieren gegenseitig, damit der PC **68** weiß, mit wem das Sendevideophon **15** spricht. Der PC **68** kann nun mit dem PC **68** des Empfangsvideophons **15** kommunizieren, mit dem das Sendevideophon **15** spricht. Damit kann der PC **68** auch ein Gespräch für das Videophon **15** herstellen.

[0133] Der größte Teil der Funktionalität des Systems **10** befindet sich auf einem Server und besteht aus Software, die auf dem Proxy-Server des Videophons **15** läuft, der bevorzugt ein SIP-Proxy-Server ist. Einen Server **66** benötigt man zum Bereitstellen der Grundfunktionalität. Einen zweiten Server benötigt man für den redundanten Betrieb, d. h. für das Aufrechterhalten der Dienste für den Fall, dass ein Server **66** ausfällt. Die Software in den Servern und im Terminal des Videophons **15** wird in diesem Fall automatisch auf den Reserveserver **66** ausgelagert. Mit dieser Konfiguration können die Terminals der Videophone **15** mit jedem anderen Terminal eines Videophons **15** am Netz **40** und mit beliebigen anderen Telefonen, die bevorzugt im Netz registrierte SIP-Telefone sind, Gespräche führen oder Gespräche von dort empfangen.

[0134] Medienserver bieten den Benutzern einen Satz Dienste für einen Satz Medienströme an. Der Medienserver **66** wird von einem Merkmalsserver **66** (bevorzugt einem Merkmalsserver **66**) kontrolliert. Er dient dazu, Quellen und Sinken für Medienströme als Teil der verschiedenen vom Benutzer aufrufbaren Funktionen bereitzustellen. Auf dem Medienserver **66** werden die folgenden Dienste bereitgestellt:
Konferenz-Bridging,
Aufzeichnung und Wiedergabe,
Umcodierung,
Töne und Ankündigungen.

[0135] Der Medienserver **66** ist ein Kästchen, das auf dem LAN oder WAN sitzt. Im Allgemeinen weist er keine weiteren Verbindungen dazu auf. Es handelt

sich bevorzugt um ein SIP-Gerät. Die Merkmalsserver liegen im Signalisierungspfad von den Terminals der Videophone **15**. Der Medienpfad geht jedoch direkt vom Medienserver **66** zur Vorrichtung.

[0136] Bei Betrieb kann der Benutzer eine Funktion anfordern, beispielsweise Videomail. Der Merkmalsserver **66** liefert dann die Benutzerschnittstelle und die Signalisierungsfunktion, der Medienserver **66** stellt die Mechanismen für Multimedia-Eingabeaufforderungen (falls verwendet) bereit und die Aufzeichnung und Wiedergabe von Meldungen.

[0137] Damit ein Terminal eines Videophons **15** Gespräche zu jedem (Video-)Telefon ohne Protokoll oder Standard(video)telefon (beispielsweise SIP) herstellen bzw. von dort empfangen kann, wird ein Gateway **70** zugefügt, beispielsweise ein SIP-Gateway. Einen analogen Vierleitungs-Gateway **70** kann man entweder direkt mit der PSTN verbinden oder mit den analogen Leitungen der lokalen PBX. Es gelten die normalen Regeln für das Einrichten von abgehenden Leitungen. In der Regel wird eine Anschlussleitung für jeweils sechs Benutzer eingerichtet. D. h., es wird unterstellt, dass jeder beliebige Anwender sein Telefon zum Wählen eines externen Anschlusses innerhalb jeder Stunde für 10 Minuten verwendet. Dient das Videophon **15** als Erweiterung einer vorhandenen PBX insofern eingehende Anrufe betroffen sind, so benötigt man für jedes Videophon **15** eine Analogleitung.

[0138] Der Benutzer des Videophons **15** kann auch auf Quellen von TV-Programmen zugreifen, beispielsweise CNN. Der Videoserver **66** der Videophone **15** ermöglicht diesen Dienst. Der Server **66** unterstützt den Anschluss eines einzigen Videokanals, auf den dann alle Anwender der Videophone **15** am Netz **40** zugreifen können. Der Videokanal entspricht zwei normalen Konferenzsitzungen. Ein Tuner kann den Kanal einstellen, der verfügbar ist. Für jeden weiteren neuen Kanal, den der Kunde gleichzeitig verfügbar haben will, sollte ein neuer Videoserver **66** für die Videophone **15** zugefügt werden.

[0139] Der Server **66** für die Videophone **15** (bevorzugt SIP) enthält auch eine Datenbank für Benutzerdaten, die einen lokalen Cache für die Benutzer-Kontaktinformation enthält. Diese Datenbank kann mit der Hauptkontakt-Datenbank des Benutzers synchronisiert werden. Die Synchronisierung kann beispielsweise mit Benutzern von Outlook/Exchange und von Lotus Notes verwendet werden. Ein eigenes Programm, das auf jeder NT-gestützten Serverplattform **66** läuft, nimmt die Synchronisierung vor. Unabhängig von der Anzahl der bedienten Sites ist nur ein Server **66** erforderlich.

[0140] In der Regel sind die Terminals der Videophone **15** über mehrere Sites verteilt, siehe [Fig. 2](#),

die über ein Fernverkehrsnetz **40** verbunden sind. Ein Server **66** reicht dazu aus, mehr als 100 Videophone **15** auf einem einzigen Campus zu bedienen. Nimmt die Gesamtanzahl der Videophone **15** an einer Site zu, so müssen in manchen Fällen mehr Server installiert werden.

[0141] Auch wenn die Videophone **15** über mehrere Sites verteilt sind, können sie auf einen Zentralserver gestützt arbeiten; dies ist jedoch keine empfohlene Anordnung, und zwar wegen der benötigten WAN-Bandbreite und der Abhängigkeit vom WAN. Bevorzugt hat jede Site mindestens einen Server **66**, der bevorzugt ein SIP-Server **66** ist, falls SIP verwendet wird. Für vorsichtigeren Naturen besteht die einfachste und unkomplizierteste Konfiguration darin, dass jede Site duplizierte Server aufweist, die bevorzugt jeweils SIP-Server sind. Der Einsatz eines Zentralservers **66** als Alternative zu Servern in entfernten Sites ist jedoch auch möglich.

[0142] Videophone **15** können überall im Netz **40** PSTN- oder PBX-gestützte abgehende Gespräche von einem einzigen zentralen Gateway **70** herstellen. Soll das Videophon **15** auch als Erweiterung einer lokalen PBX arbeiten und eingehende Gespräche annehmen, so muss man an jedem Ort einen PSTN-Gateway **70** bereitstellen. Für jedes Videophon **15** an dieser Site muss ein Port am Gateway **70** vorhanden sein.

[0143] Ein zentraler CNN-Server **66** kann TV-Kanäle an jedes beliebige Videophon **15** am Netz **40** verteilen. Trotzdem kann es zu bevorzugen sein, eigene Server an den Sites aufzunehmen, anstatt das WAN mit dieser Bandbreite zu belasten.

[0144] Ein Videophon **15** kann entweder an ein 10/100 Ethernet-Netz **40** oder ein ATM-Netz **40** bei 155 Mbit/sec angeschlossen werden (es sind Glasfaser und Kupferkabel möglich). Ein an ATM angeschlossenes Videophon **15** benutzt eine IP-Steuerungsebene zum Erzeugen der ATM-Adressen der Endpunkte für ein Gespräch, und es verwendet anschließend eine ATM-Signalisierung zum Aufbau des Trägerkanals zwischen diesen Endpunkten. Der Trägerkanal wird mit einem Switched Virtual Circuit (SVC) hergestellt, wobei die vollständigen QoS-Anforderungen spezifiziert werden.

[0145] Jeder Videostrom liegt zwischen 2 Mbps und 6 Mbps duplex. Dies hängt von den Einstellungen und der ausgehandelten Bandbreite ab. Da die Anzeigevorrichtung mehr als einen einzigen Videostrom darstellen kann, nimmt die insgesamt erforderliche Anschlussbandbreite für jedes Videophon mit der Anzahl der Teilnehmer an einem Gespräch zu. Eine senderseitige Beschneidung stellt sicher, dass die maximal erforderliche Bandbreite ungefähr das 2,5fache der Bandbreite für einen einzigen Video-

strom beträgt. Befinden sich mehrere Videophone **15** an einer Site, so gilt das normale Verhältnis zwischen Benutzern und Anschlussleitungen auch für Sitzungen mit Videophonen **15**. Anders ausgedrückt nimmt man von einem Benutzer eines Videophons **15** an, dass er im Mittel in jedem Gespräch mit zwei anderen Personen spricht, d. h. zwei Ströme benötigt, und dass er das Videophon **15** im Mittel 10 Minuten pro Stunde benutzt. Bei einer mittleren Codierungsrate von 3 Mbps führt dies zu einem WAN-Bandbreitenbedarf von 6 Mbps, von dem man erwarten darf, dass er bis zu 6 Benutzer unterstützt.

[0146] Das Videophon **15**, siehe [Fig. 3](#), arbeitet in einem Ethernet-Netz **40**, das 'p'-enabled ist, falls die Terminaldichte der Videophone **15** gering ist. Das System **10** mit den Videophonen **15** baut über den ATM-Teil des Netzes **40** einen SVC auf, der die beiden Videophone **15** miteinander verbindet, und nutzt das Ethernet, das 'p'-enabled ist, dazu, sicherzustellen, dass über den Ethernet-Teil der Verbindung eine ausreichende Quality of Service (Quality of Service = Dienstgüte in Kommunikationsnetzen) geliefert wird.

[0147] Die wesentlichen Elemente des Systems **10** mit den Videophonen **15** sind in [Fig. 4](#) dargestellt. Gemeinsam bilden sie ein Multimedia-Zusammenarbeitswerkzeug, das die Fähigkeit geographisch zerstreuter Teams zum Austausch wesentlich verbessert. Derartige Teams werden in jedem großen Unternehmen zunehmend normal. Die Werkzeuge, die eine wirksame und effiziente Zusammenarbeit unterstützen, unterscheiden sich jedoch wenig von den Werkzeugen von vor zehn Jahren und sind in vieler Hinsicht unbefriedigend. Das Videophon **15** geht die zahlreichen Probleme bei existierenden Systemen in übergreifender Weise an und erzeugt eine sprunghafte Verbesserung der Zusammenarbeit über Entfernungen hinweg. Dies wird durch neu verfügbare Technologien ermöglicht. Das System hebt sich durch die Quality of Service und die passende Mischung von Funktionen ab, die durch die Entwicklung einer ausgezeichneten Benutzerschnittstelle einsetzbar werden. Es ist für Erweiterungen entworfen, weil die Architektur auf Standards beruht.

[0148] Die Audio- und Videoströme werden wie erläutert über das Netz vom Ursprungsvideophon **15** zu den Endvideophonen **15** übertragen, wobei beispielsweise bekannte SIP-Techniken verwendet werden. SIP-Meldungen kann man mit IP-Routingtechniken über heterogene Netze leiten. Für Medienströme in heterogenen Netzen ist es wünschenswert, dass ein direkterer Weg vorhanden ist. In Fällen, in den das Ursprungsvideophon **15** einer Konferenz an ein Ethernet angeschlossen ist und ein Endvideophon **15** der Konferenz an ein ATM-Netz, siehe [Fig. 15](#), erfolgt bevorzugt die folgende Adressierung der Pakete, die das Netz zwischen den Ursprungs- und Endvideophonen durchlaufen. Das Ursprungsvideophon **15**

sendet ein Paket auf das Ethernet, mit dem es über die IP-Adresse des Ursprungsvideophons kommuniziert. Die Pakete erreichen einen Ursprungsgateway **81**, der das Ethernet mit dem ATM-Netz verbindet. Am Ursprungsgateway **81** wird die IP-Adresse des Ursprungsvideophons **15** aus dem Paket entnommen, und der Ursprungsgateway **81** fügt dem Paket die ATM-Adresse des Ursprungsgateways **81** zu und sendet das Paket an das Endvideophon **15** weiter. Empfängt das Endvideophon **15** das Paket, so speichert es die ATM-Adresse des Ursprungsgateways **81** aus dem Paket und sendet an den Ursprungsgateway **81** ein Rückpaket, das anzeigt, dass es das Paket empfangen hat und dass die ATM-Adresse des Endvideophons **15** enthält. Der Ursprungsgateway **81** entnimmt beim Empfang des zurückgesendeten Pakets die ATM-Adresse des Endvideophons **15** und fügt die IP-Adresse des Ursprungsgateways **81** in das Rückpaket ein. Der Ursprungsgateway **81** sendet nun das Rückpaket wieder an das Ursprungsvideophon **15**.

[0149] Auf diese Weise kennt jeder kritische Knoten des Wegs die jeweiligen Adressen eines jeden kritischen Knotens auf dem Gesamtweg zwischen dem Ursprungsvideophon **15** und dem Endvideophon **15**. Jeder Knoten auf dem Weg kennt zumindest die Adresse des nächsten Knotens auf dem Weg. Auf Wunsch können zusätzliche Adressen in den jeweiligen Paketen gehalten werden, wenn sie sich auf dem Weg bewegen, damit jeder Knoten des Wegs hinsichtlich der Adressen der kritischen Knoten mehr kennt als den folgenden Knoten, zu dem das Paket läuft. Dies ist so, weil, wenn sich das Paket von Knoten zu Knoten bewegt, und insbesondere in diesem Beispiel vom Ursprungsvideophon **15** zum Ursprungsgateway **81**, von dort zum Endvideophon **15** und zurück zum Ursprungsgateway **81** und zum Ursprungsvideophon **15**, jeder Knoten die kritische Adresse des vorhergehenden Knotens speichert, von dem das jeweilige Paket empfangen wurde, und seine eigene Adresse bezogen auf den Typ des Netzes, zu dem der folgende Knoten gehört, einfügt. Folglich sind alle kritischen Adressen, an die jeder Knoten das Paket zum folgenden Knoten versenden muss, entlang des Weges verteilt.

[0150] Dieses Beispiel für die Übertragung eines Pakets vom Ursprungsvideophon **15** auf einem Ethernet zu einem Endvideophon **15** auf einem ATM-Netz ist auch in umgekehrter Richtung anwendbar, wenn das Ursprungsterminal oder -videophon **15** mit einem ATM-Netz verbunden ist und das Endvideophon **15** mit einem Ethernet.

[0151] In ähnlicher Weise kann der Weg ein Ursprungsvideophon **15** verbunden mit einem Ethernet und ein Endvideophon **15** verbunden mit einem Ethernet umfassen, wobei dazwischen von den Paketen ein ATM-Netz durchquert wird, siehe [Fig. 16](#). In

einem derartigen Fall befinden sich zwei Gateways an jedem Übergang, an dem sich eine Schnittstelle zwischen dem Ethernet und dem ATM-Netz befindet. Wie erklärt wird bei diesem Vorgang einfach ein zusätzlicher Knoten in den Weg eingefügt, wobei der Ursprungsgateway **81** seine eigene ATM-Adresse in das Paket einfügt und es an den Endgateway **83** sendet, der die ATM-Adresse des Ursprungsgateways speichert und dem Paket die IP-Adresse des Endgateways zufügt. Das Paket wird nun auf dem Ethernet zum Endvideophon **15** gesendet. Mit dem Rückpaket geschieht das Gleiche in umgekehrter Reihenfolge. Jeder Gateway speichert die jeweilige Adressinformation vom vorhergehenden Gateway oder Endvideophon **15** und fügt seine eigene Adresse in das Rückpaket ein, das schließlich an das Ursprungsvideophon **15** gesendet wird. Dabei speichern der Ursprungsgateway **81** und das Ursprungsvideophon **15** die ATM-Adresse des Endgateways **83** bzw. des Ursprungsgateways **81**, wodurch die jeweiligen Adressen in jeder Verbindung des Gesamtwegs wirksamer gespeichert und rasch an nachfolgende Pakete einer Verbindung gesendet werden.

[0152] Der Hauptcontroller **50** und die Netzschnittstelle **42** des Videophons **15** können beispielsweise die Adresse des Videophons **15** jedem Paket zufügen, das von ihnen an das Netz **40** gesendet wird, und zwar mit Hilfe der gleichen Vorgehensweisen, die Fachleuten bekannt sind, und bei denen SIP-Routinginformation (oder irgendeine andere verwendete Standard-Routinginformation) im Paket abgelegt wird. Die Netzschnittstelle **42** speichert auch die Adressinformation, die sie von einem Paket von einem Knoten auf dem Netz empfängt, in einem lokalen Speicher. Bei einem Gateway im Netz **40** kann man das gleiche Vorgehen verwenden. Bekanntlich besitzt der Gateway Kontrollvorrichtungen und Datenverarbeitungsvorrichtungen, die ein Paket an seinen endgültigen Bestimmungsort versetzen. Eine Netzschnittstelle **42** und ein Hauptcontroller **50** des Kontrollmechanismus des Gateways, die hinsichtlich der SIP-Routinginformation mit bekannten Vorgehensweisen arbeiten, speichern von einem Paket empfangene Adressinformation und setzen ihre eigene Adressinformation bezüglich eines Netzes **40**, in das das Paket zu senden ist, in das Paket. Beispielsweise kann die Adressinformation des Gateways oder des Videophons **15** in einem Feld abgelegt werden, das sich im Kopfeintrag des Pakets befindet. Obwohl in dem Beispiel von Videophonen **15** als Ziel und Ursprung die Rede ist, darf man nicht übersehen, dass man in diesem Gesamtschema jede beliebige Vorrichtung als Knoten verwenden kann, die Pakete erzeugt und empfängt.

[0153] Das Virtual Presence Video-Phone (Videophon) **15** ist eine Desktop-Einrichtung am Netz **40**, die ein privates Kommunikationsterminal darstellt. Es ersetzt das Telefon auf dem Schreibtisch des Benut-

zers, bietet alle Merkmale eines modernen PBX-Terminals zusammen mit der Einfachheit einer Benutzerschnittstelle und dem leichten Gebrauch, der durch den großen Berührungsbildschirm **74** des Videophons **15** ermöglicht wird.

[0154] Das Videophon **15** fügt sämtlicher persönlicher Kommunikation eine Videodimension zu und verwandelt das Geschehnis in eine virtuelle Anwesenheit. Früher war die Videoqualität auf Videokonferenzsystemen für eine transparente Technologie nicht hoch genug. Das Videophon **15** ist das erste persönliche Videophon, dessen Videoqualität so hoch ist, dass der geeignete Eindruck entsteht. Die effektive Echtzeit-Videokommunikation muss nicht nur eine Bildqualität liefern, die nahe an der Qualität des ausgestrahlten Fernsehens liegt, sondern es muss auch die Verzögerungszeit sehr gering gehalten werden. Soll ein natürlicher Gesprächsverlauf zustande kommen, so ist Lip Sync ebenfalls wichtig. Alle diese Probleme sind in den Entwurf des Videosubsystems des Videophons **15** eingegangen. Im Videophon **15** wird die neueste Technologie für den Codierer **36** und den Decodierer **34** verwendet und eigens für diese Anwendung konfiguriert. Anders ausgedrückt liegt das Videophon **15** so nahe wie möglich an der "Anwesenheit".

[0155] Das Videophon **15** verbessert auch die Leistungen eines herkömmlichen Hörer Telefons stark, und zwar durch den Einsatz eines klangtreuen Audiokanals nahezu in CD-Qualität, der eine kristallklare Stimme liefert. Stereo-Audiokanäle bieten eine räumliche Auflösung des Schalls für jeden Teilnehmer. Eine hochentwickelte Stereoechoauslöschung unterdrückt nicht nur die Geräusche von den Lautsprechern **64** der Einheiten, sondern ermöglicht es dem Sprecher auch, ein Gespräch mit normalen Konversationslautstärken zu führen, und dies auch in einem geräuscherfüllten Raum.

[0156] Das Videophon **15** unterstützt direkt den Aufbau von bis zu 4 Videokonferenzgesprächen (d. h. ein 5-Weg-Gespräch) mit entfernten Parteien oder ein Audiokonferenzgespräch mit bis zu 10 Teilnehmern. Jeder Benutzer kann sehen, ob alle anderen Mitglieder seiner Arbeitsgruppe erreichbar sind. Das Videophon **15** verwendet bevorzugt das Session Initiation Protocol (SIP) als Mittel zum Herstellen, Modifizieren und Beenden von Mehrstrom-Multimediasitzungen. Das Videophon **15** kann über einen Gateway **70** ein Audiogespräch mit jedem beliebigen anderen SIP-Telefon oder jedem beliebigen anderen Telefon herstellen.

[0157] Das Videophon **15** stellt hohe Anforderungen an das Netz **40**, mit dem es verbunden ist. Die Videogespräche des Videophons **15** erfordern ein Netz **40**, das kontinuierlich hohe Bandbreiten liefern und die Bandbreite, die Verzögerungszeit und die

Schwankungen garantieren kann. Marconi plc ist spezialisiert auf das Bereitstellen von Netzen, die Anwendungen mit hoher Quality of Service unterstützen. Eine Konferenzraumversion des Videophons **15** ist ebenfalls verfügbar.

[0158] Das Videophon **15** ist ein Kommunikationsterminal (Plattform), das vollständig in einen PC **68** des Benutzers integrierbar ist, der Computerplattform. Eine Anwendung des Videophons **15** für den PC **68** bietet eine Anzahl Integrationsdienste zwischen dem PC **68** und dem zugehörigen Terminal des Videophons **15**. Dazu gehört das automatische Herstellen von NetMeeting-Sitzungen zwischen den Parteien in einem Konferenzgespräch über das Videophon **15**, falls dies freigegeben ist, damit Anwendungen wie die Präsentationstafel oder Präsentationen usw. gemeinsam verwendet werden können. Dazu gehören auch weitere Fähigkeiten einschließlich dem "Drag-and-Drop-Wählen" einer Nummer auf dem PC **68** durch das Videophon **15**.

[0159] Ein Satz Server, von denen jeder bevorzugt ein SIP-Server ist, liefert die Gesprächssteuerung und die Implementierung der Merkmale für die Vorrichtungen des Netzes **40**. Es handelt sich um Softwareserver, die auf Standardcomputerplattformen laufen und Redundanz bieten. Auf diesen Servern läuft auch eine lokale Kopie der Kontaktinformations-Datenbank des Benutzers und die Datenbank mit den Vorlieben des Benutzers. Auf diesen Servern verfügbare Anwendungen bieten Zugriff auf gemeinsame Verzeichnisse oder andere Verzeichnisse, auf die mit LDAP (LDAP = Lightweight Directory Access Protocol) zugegriffen werden kann.

[0160] Ein Synchronisierserver **66** erhält die Synchronisation zwischen der Haupt-Kontaktdatenbank des Benutzers und der lokalen Kopie auf dem Server **66** (bevorzugt SIP). Die Synchronisation von Outlook Exchange oder Lotus Notes wird unterstützt. Ein Satz Mediengateways **70** werden hin zum analogen oder digitalen PSTN-Netz **40** verwendet. Ein Satz Mediengateways **70** stellt die Verbindung zu den gängigsten PABX-Anlagen (PABX = Private Automatic Branch Exchange, Wahl-Nebenstellenanlage) her, und zwar einschließlich der Sprachnachrichtensysteme, die zu diesen PABX gehören.

[0161] Der Medienserver **66** bietet dem Terminal des Videophons **15** eine Anzahl Dienste. Auf Wunsch wirkt er als Bridging-Konferenzserver **66** für eine Videokonferenz mit mehr als 4 Teilnehmern. Er kann auch eine Codeumsetzung zwischen den Standards des Videophons **15** und anderen üblichen Audio- oder Videoformaten liefern, z. B. H320/H323. Er kann Möglichkeiten zum Aufzeichnen und Wiedergeben bieten und erlaubt es, Sitzungen aufzuzeichnen oder wiederzugeben. Er kann als Quelle von Tönen und Ankündigungen dienen.

[0162] Benötigt wird eine Firewall gemäß dem verwendeten Standard, beispielsweise eine SIP-Firewall, damit die dynamisch erzeugten RTP-Ströme gesteuert durch die Standard-Proxysoftware (beispielsweise SIP-Proxysoftware) sicher weitergeleitet werden. Ein TV-Server **66** dient als Quelle für die TV-Verteilung und erlaubt es den Benutzern der Videophone **15**, irgendeinen unterstützten Kanal zu wählen, beispielsweise CNN.

[0163] Das Videophon **15** ist für Ethernet- und ATM-Desktops geeignet. Das Videophonterminal **15** unterstützt End-to-End-ATM-SVCs und verwendet sie zum Herstellen von Verbindungen mit dem geforderten Niveau der Quality of Service. Das Videophon **15** unterstützt auch IP-Anschlussmöglichkeiten über LANE-Dienste (LANE = Local Area Network Emulation). Um hierfür die verlangte QoS zu garantieren ist LANE 2 erforderlich. Das Videophon **15** bietet ATM-Durchleitungen zu einem an ATM angeschlossenen Desktop-PC **68** oder eine ATM-Ethernet-Durchleitung zum Anschluss des PC **68** über Ethernet.

[0164] Das Videophon **15** erfordert die Unterstützung von End-to-End-QoS. Für ein an das Ethernet angeschlossenes Videophon **15** muss der Benutzeranschluss 802.1 p, Diff-Serv und/oder IntServ oder einen besseren Standard unterstützen. Ist das Ziel über ein ATM-Netz **40** erreichbar, so wird ein Ethernet- oder ATM-Gateway **70** bereitgestellt. Der SIP-Proxyserver **66** und die SIP-Signalisierung erzeugen den ATM-Endpunkt so nahe wie möglich am Terminal des Zielvideophons **15**, d. h. seiner ATM-Adresse, falls es an ATM angeschlossen ist, oder am nächstgelegenen ATM-Ethernet-Gateway **70**. Die Signalisierung stellt einen SVC mit der geeigneten QoS über den ATM-Abschnitt des Netzes **40** her. Dieser SVC wird mit dem besonderen Ethernetstrom verbunden, der die geeignete Prioritätsanzeige am entfernten Ende erzeugt.

[0165] Die Produktlinie der Videophone **15** besteht aus mehreren Terminals (Endgeräten), einem Satz Server, die Merkmale bieten, die nicht in die Endgeräte eingebaut sind, und einem Satz Gateways **70**, die die Produkte mit vorhandenen Einrichtungen und äußeren PSTN-Diensten verbinden. Die vom System **10** gelieferte Grundfunktionalität besteht in:

- Telefondienste, bei denen Video in allen Gesprächen "über das Netz" sowie Video und Audio in sehr hoher Qualität verfügbar sind.
- Mehrparteien-Konferenzdienste mit Audio und Video, spontan oder vorgeplant, vollständige Selbstbedienung, vollständige Integration in die Telefondienste.
- Anwesenheitsdienste mit verschiedenen Werkzeugen, die die Erreichbarkeit zur Zusammenarbeit feststellen.
- Dienste für gemeinsame Oberflächen wie elek-

tronische Präsentationstafel, gemeinsame Applikationen, gemeinsame Dokumente, Rundum-Versendung von Präsentationen.

– Andere als nützlich erachtete Dienste wie Video-TV-Ausstrahlungen (Mikes Nachrichten an die Truppe). Interaktives Online-Training usw. Auf Wunsch sind auch Sitzungs-Aufzeichnungsdienste verfügbar.

[0166] Das Videophon **15** ist ein Telefon mit gewaltigen neuen Funktionen, nicht der Computer, der versucht, das Telefon nachzuahmen. Dies erlaubt es, den Computer gleichzeitig für die Aufgaben zu nutzen, für die er geeignet ist, und bietet eine flexible und zugleich anwendungsspezifische Kommunikationsvorrichtung. Die Benutzerschnittstelle und den Hardwareentwurf kann man auf diese Anwendung abstimmen und eine sofort betriebsbereite hochgradig zuverlässige Kommunikationseinrichtung vergleichbar zum derzeitigen Telefon bereitstellen. PCs werden diese Eigenschaften niemals haben. Dieser Ansatz erlaubt auch eine Kontrolle über die Betriebsumstände der Vorrichtung. Die Supportprobleme im Zusammenhang mit Hardware- und Softwarekonfigurationsfragen beim PC **68** treten nicht mehr auf.

[0167] Untersuchungen der menschlichen Seite haben wieder und wieder gezeigt, dass die Audioqualität der wichtigste Einzelfaktor für eine effektive und transparente Kommunikation ist. Obwohl man noch einen Hörer braucht, bieten Freisprechaudio in hervorragender Qualität einschließlich Echounterdrückung (AEC), automatischer Verstärkungsregelung (AGC, AGC = Automatic Gain Control), Breitband-Audiofähigkeit (G.722 mit 8 kHz Bandbreite oder mehr), Stereoausgabe und Integration in die Tonausgabe des PC **68** ein neues Niveau der wirksamen Zusammenarbeit über die Entfernung hinweg. Es ist auch ein Mikrofonfeld mit hoher Qualität vorhanden, das zum Begrenzen von blechernen Geräuschen entworfen und bearbeitet wurde.

[0168] Es wird eine einfache, saubere, intuitive und vollständig flexible Plattform für die visuelle Ausgabe und Tasten/Auswahl-Eingabe verwendet. Beim ersten Videophonmodell ist dies ein hochwertiger TFT-Farbbildschirm mit 17 Zoll Diagonale und dem Verhältnis 16 zu 9 und einer Auflösung von 1260 × 768 oder mehr. Darüber liegt ein langlebiges Berührungsfeld mit mittlerer Auflösung. Eine helle (mehr als 200 Nit) Aktivmatrixfläche mit vergrößertem Betrachtungswinkel (größer als +−60°) wird für die Darstellung von Full-Motion-Video (Full-Motion = mit maximaler Framenanzahl) verwendet, das in einer Büroumgebung eine angenehme Ansicht bietet. Man kann Bildschirme verwenden die größer, heller und schneller sind und einen höheren Kontrast und einen größeren Betrachtungswinkel aufweisen.

[0169] Im Videophon **15** wird eine TFT-Farb-LCD-Anzeige verwendet, die eine dem PC **68** ähnliche Architektur hat und eine VGA-Schnittstelle für den Bildschirm **54**, die auf einem Intel Celeron/440 MMX und einem Lynx-VGA-Controller beruht.

[0170] Eine hochwertige Digitalkamera mit 480 Zeilen und Zeilenfolgeabtastung dient dazu, zumindest ein 640 × 480-Video mit 30 Frames pro Sekunde zu liefern. Im Videophon **15** wird die MPEG-2-Codierung verwendet und die Technologie des Videocodierers **36** für Set-Top-Boxes ausgenutzt. Es können mehrere unterschiedliche Bitraten erzeugt werden. Dadurch kann man die Videoqualität an die verfügbaren Ressourcen für Gespräche zwischen zwei Teilnehmern anpassen sowie an den Teilnehmer mit der höchsten Qualität, falls ein Teilnehmer mit mehreren anderen spricht oder viele Parteien miteinander sprechen. Ein integriertes hochwertiges Kameramodul ist nahe am Bildschirm angeordnet. Dabei ist ein externer Videoeingang (Firewire) bereitgestellt, der den Einsatz zusätzlicher Kameras, Videorekorder oder anderer Videoquellen erlaubt.

[0171] Eine vorhandene 10/100-BaseT-Ethernet-Verbindung zum Desktop ist die einzige erforderliche Verbindung für die Kommunikation mit dem LAN, WAN, dem Desktop-PC **68** und diversen Servern, Proxies und Gateways **70**. Zeitkritische RTP-Ströme für Audio und Video werden mit 802.1 p mit Priorität versehen, wobei die Mechanismen innerhalb des Ethernetbereichs des LAN für QoS geliefert werden. DiffServ wird ebenfalls unterstützt, wobei RSVP (RSVP = Resource Reservation Protocol) eine Option ist. Damit man keine weitere Verdrahtung zum Desktop aufbauen muss, enthält das Videophon **15** einen kleinen 10/100-Ethernetverteiler, mit dem man den vorhandenen Desktop-Port sowohl für das Telefon als auch den PC **68** verwenden kann.

[0172] Das Videophon **15** unterstützt auch eine ATM-Schnittstelle. Die Schnittstelle beruht auf der Verwendung der HE155 Mbit/sec-Karte mit entweder einer Kupferkabel- oder einer Lichtleiterschnittstelle. Das Videophon **15** bietet einen ATM-Durchreichport für den Anschluss des Videophons **15**, das mit ATM verbunden ist, an einen Desktop, der mit ATM verbunden ist, oder für den Anschluss an einen PC **68**, der mit Ethernet verbunden ist.

[0173] Die Kompromisse zwischen Kosten und Leistungsfähigkeit für die Konferenzraum-Umgebung unterscheiden sich offensichtlich von den Verhältnissen beim Desktop. Videoprojektion, zahlreiche Kameras mit Schwenk/Neigung/Zoom aus der Ferne, zahlreiche Mikrofone, zahlreiche Videokanäle, Präsentationstafeln mit Rückprojektion und andere für die Konferenzraum-Umgebung geeignete Produkte sind in das Konferenz-Videophon **15** integriert. Die

Konferenzraum-Umgebung und der Desktop arbeiten nahtlos und transparent zusammen. Diese Umgebung verwendet anstelle des Desktops sehr häufig OEM-Anlagen (OEM = Original Equipment Manufacturer, Originalgerätehersteller), die Schnittstellen zur gleichen Infrastruktur und zu den gleichen Standards haben. Der Hardwareentwurf ist im Wesentlichen der Gleiche, wobei noch Audiounterstützung für mehrere Mikrofone und Videounterstützung für mehrere Kameras und Bildschirme hinzukommt. Wahlweise kann man eine Anwendung auf dem PC **68** verwenden, die entweder von der Maus oder dem Berührungsbildschirm **74** gesteuert wird, falls der PC **68** einen Berührungsbildschirm **74** hat, der die Verbindung zu einem kostengünstigen SIP-Telefon herstellt. Für diejenigen Desktops bzw. diejenigen Orte, die die beschriebenen Zusammenarbeitsmöglichkeiten nicht benötigen, kann man ein Standardtelefon verwenden, das mit dem System **10** arbeitet, ohne dass eine zusätzliche Verdrahtung oder eine PBX erforderlich ist.

[0174] Benutzt man den SIP-Standard (Session Initiation Protocol), so werden die Terminalgeräte von einem oder mehreren Servern unterstützt, die Registrierung, Ort, Benutzerprofil, Anwesenheit und verschiedene andere Proxydienste bieten. Diese Server sind kostengünstige Linux- oder BSD-Maschinen, die an das LAN angeschlossen sind.

[0175] Das Videophon **15** ist das Telefon. Damit muss eine Grundmenge an PBX-Funktionen bereitgestellt werden, zu denen Weiterverbinden, Umleitung, Konferenzen mit 3 (und auch 4, 5, ...) Teilnehmern, Caller ID+, Rufnummernaufzeichnung usw. gehören. Einige dieser Merkmale kann man auf einen SIP-Erweiterungsmechanismus mit dem Namen "CPL" aufsetzen, der eine Sprache zum Abwickeln von Gesprächen in sicherer und erweiterbarer Weise darstellt.

[0176] Das Videophon **15** bietet aktive Anwesenheit und sofortige Benachrichtigung. Als das vielleicht umwälzende Werkzeug zum Verbessern der täglichen Zusammenarbeit in verteilten Gruppen ermöglicht die Anwesenheitserkennung den Leuten, festzustellen, welche Personen anwesend sind und womit sie gerade beschäftigt sind. Sie bietet die Grundlage für Gespräche mit sehr geringem Drumherum. Telefon-Zettel und das herkömmliche Wählen von Nummern sind nicht mehr nötig. Dies ermutigt Gruppen, als Gruppe zu kommunizieren und nicht über die trennenden Telefongespräche zwischen zwei Personen, die derzeit üblich sind. Die Integration mit Instant Messaging (Echtzeit-Email) bietet einen verzögerungsfreien Weg, kurze Textnachrichten auszutauschen, wobei wohl die Tastatur des PC **68** für die Eingabe verwendet wird.

[0177] Das Videophon **15** bietet eine verteilte und

redundante Architektur. Dies ist das Telefonsystem **10**, das zuverlässig sein muss. Es sollte möglich sein, es zentral mit örtlichen Erweiterungen zu verwalten, wobei verteilte Server allen Benutzern "sofortige" Antwort bieten. Jeder der unterschiedlichen SIP-Proxyfunktionen wird beispielsweise, falls SIP verwendet wird, eingesetzt, so dass man sie beliebig in einem Satz physikalischer Server kombinieren kann, wobei sich redundante Versionen im Netz **40** befinden.

[0178] Für gemeinsame Oberflächen und gemeinsame Anwendungen wird Microsoft NetMeeting verwendet. Eine Computer/Telefon-Schnittstelle (CTI) für den PC **68** und den PDA kann verwendet werden, die beispielsweise Merkmale wie integrierte Kontaktlisten, automatisches Wählen von ausgesuchten Telefonnummern oder Namen, Listenaufzeichnung der geführten Gespräche, automatische Kontakteinträge usw. aufweist.

[0179] SIP stellt eine Herausforderung für Firewalls dar, da die RTP-Flüsse dynamisch zugeordnete UDP-Ports verwenden, und die Adressen/Port-Information in SIP-Meldungen befördert wird. Dies bedeutet, dass die Firewall die SIP-Meldungen verfolgen muss und "winzige Löcher" in der Firewall für die entsprechenden Adressen/Port-Kombinationen öffnen muss. Wird NAT (NAT = Network Address Translation) verwendet, so müssen die Meldungen verändert werden, damit sie die geeigneten verschobenen Adressen bzw. Ports aufweisen. Diese Aufgabe kann man auf zwei Weisen erfüllen. Zum einen kann man die Fähigkeit dazu in die Firewall einbauen. Die drei führenden Anbieter von Firewalls (Checkpoint, Network Associates und Axxent) stellen dies bereit. Wahlweise kann man eine besondere Firewall verwenden, die parallel zur Hauptfirewall nur SIP behandelt. Im Handel sind Versionen einer solchen Firewall erhältlich, beispielsweise die Firewall von MicroAppliances. Man beachte, dass SIP oder NetMeeting bevorzugte Ausführungsformen sind, die dafür vorhanden sind, ihre erforderlichen jeweiligen Funktionen auszuführen. Wird die nötige Funktionalität bereitgestellt, kann man auch Alternativen dazu verwenden.

[0180] Fig. 5 zeigt die mechanischen Hauptkomponenten des Terminals eines Videophons **15**. Der Sockel bietet eine einfache Möglichkeit, die Höhe der Hauptanzeigetafel **54** einzustellen und die Anzeige in dieser Höhe zu befestigen. Der Höheneinstellbereich beträgt mindestens sechs Inch (1 Inch = 2,54 cm), damit eine Anpassung an unterschiedlich große Benutzer möglich ist. Dabei ist angenommen, dass der Sockel auf einem Tisch steht und die Höhe der Tischplatte genormt ist. Die Verbindung zwischen dem Sockel und der Haupteinheit muss einen gewissen Grad an Neigung gegen die Vertikale erlauben, die den Vorlieben des Benutzers entgegenkommt, und sich in diesem Winkel leicht befestigen lassen. Die Größe der Neigung sollte zwischen 0 und +15° gegen die

Vertikale liegen. Wahlweise kann man die Haupteinheit direkt an der Wand befestigen und benötigt dann keine Sockelbaugruppe.

[0181] Das Gehäuse der Haupteinheit bietet Platz für alle anderen Elemente im Entwurf des Videophons **15** einschließlich aller in **Fig. 5** dargestellten Elemente und der Terminalelektronik. Der Hörer kann rechts oder links am Gehäuse befestigt werden. Rechtshänder nehmen den Hörer gern in die linke Hand (da sie den Berührbildschirm **74** mit der rechten Hand steuern wollen und rechts schreiben) und Linkshänder in die rechte Hand. Obwohl die linke Seite den Normalfall darstellt, muss es möglich sein, den Hörer rechts anzubringen. Am Gehäuse ist ein Lautsprecherstecker vorgesehen, damit man die Lautsprecher **64** entfernt vom Videophon **15** montieren kann. Es sind Eingänge vorhanden, die die Lautsprecher-Ausgangssignale vom zugehörigen PC **68** aufnehmen. Dadurch kann das Videophon **15** die Audiosignale des PC **68** und des Videophons **15** kontrollieren. Man kann einen drahtlosen Anschluss zu den Lautsprechern **64** verwenden (über Bluetooth oder Sony-Standards).

[0182] Mit der Einheit wird ein Hörer mitgeliefert. Er sollte mit einem RJ9-Standardspiralkabel und einem Verbindungsstecker angeschlossen werden. Der abgelegte Hörer sollte leicht zu ergreifen sein und trotzdem nicht stören. Eine Option für den Hörer bietet eine Standardtastatur für den Hörer. Für eine bessere Beweglichkeit des Terminalbenutzers kann man einen drahtlosen Hörer verwenden.

[0183] Für den Anschluss eines Stereokopfhörers und eines Mikrofons ist ein Stecker vorhanden. Der Gebrauch von Kopfgarnituren für normale Telefongespräche nimmt zu. Dem Benutzer soll es freistehen, einen Kopfhörer und ein am Bügel befestigtes Mikrofon zu wählen oder nur einen Kopfhörer, wobei das Mikrofonfeld als Eingabevorrichtung verwendet wird. Für eine bessere Beweglichkeit des Terminalbenutzers kann man eine drahtlose Kopfgarnitur verwenden.

[0184] Als Schnittstelle zu PDAs und anderen IR-Geräten ist ein IR-Port an einer leicht zugänglichen Position am Hauptgehäuse vorhanden. Derzeit sind IR-Schnittstellen bei Telefonen und PDAs am weitesten verbreitet. Daher ist eine IR-Schnittstelle aus den gleichen Gründen nötig, die eine Bluetooth-Schnittstelle erforderlich machen.

[0185] Im Gehäuse ist ein Mikrofonfeld untergebracht. Das Feld darf beim normalen Betrieb des Terminals keine Fremdgeräusche erzeugen. Es sollte insbesondere nicht möglich sein, Handlungen des Benutzers am Berührbildschirm zu erfassen. Das Mikrofonfeld erlaubt es einem Benutzer, mit normaler Konversationslautstärke innerhalb eines Bogens

(von beispielsweise 6 Feet (1 Foot = 30,48 cm)) um die Vorderseite der Einheiten herum in einem Winkel von 110° in der horizontalen Ebene und in Anwesenheit von Hintergrundgeräusch mit einer vorbestimmten dB-Zahl zu sprechen. Die Einheit muss eine schlichte Anzeige aufweisen, ob das Mikrofon aktiv oder nicht aktiv ist, d. h. ein Äquivalent zu "aufgelegtem Hörer" oder "abgenommenem Hörer". Ein Benutzer des Videophons **15** will sich rückversichern, dass er ohne sein Wissen nicht abgehört wird. Dies stellt ein Audioäquivalent des mechanischen Kameraverschlusses dar.

[0186] Die Haupteinheit des Videophons **15** kann wahlweise einen Smartcard-Leser aufweisen, damit ein sicherer Zugang zum Terminal für persönliche Belange vorhanden ist. Für den Zugang zum Videophon **15** sind mehrere verschiedene Zugangskontrollmöglichkeiten nötig, und zwar vom einfachen Anmelden mit Hilfe eines Passworts auf dem Bildschirm bis zu Sicherheits-Fobs (Fobs sind kleine elektronische Geräte, die ein Maximum an Sicherheit bieten). Ein Smartcard-Leser stellt eines dieser Zugangsverfahren dar.

[0187] Es stellt selbstverständlich einen Vorteil dar, wenn man die Neigung und Schwenkung vom Bildschirm aus kontrollieren kann. Bevorzugt erfolgen Neigung und Schwenkung lediglich elektronisch und benötigen keine mechanischen Mechanismen. Die Kamerabefestigung sollte so nahe wie möglich an der Oberseite des Hauptbildschirms montiert sein, damit der Blickkontakt verbessert wird.

[0188] Die Kamera sollte eine Digitalkamera **47** sein, die 480p-Ausgangssignale erzeugen kann. Die Ausgangssignale der Kamera gelangen an einen MPEG-2-Codierer **36**. Die Kamera sollte dynamisch konfigurierbar sein, damit man die Kameraausgangssignale für die Eingabe in den Codierer **36** optimieren kann, und zwar für die gewählte Ausgangsdatenrate des Codierers **36**. In der Hauptsache wird die Kamera Gesichter aufnehmen. Daher ist die exakte Erfassung der Hauttönung in einem breiten Bereich von Beleuchtungsumständen ein wesentliches Merkmal.

[0189] Die Kamera sollte in einem breiten Bereich von Beleuchtungszuständen bis hinunter zu einem Wert von 3 Lux betreibbar sein. Die Kamera sollte einen automatischen Weißabgleich besitzen. Die Weißabgleichänderungen müssen langsam geschehen, damit Transiente des aufgenommenen Bilds keine unerwünschten Bildstörungen verursachen. Erst Änderungen, die länger als 5 Sekunden dauern, sollten den Weißabgleich verändern. Die Kamera sollte zwischen 18 Inch und 10 Feet scharfgestellt sein, d. h. eine große Schärfentiefe aufweisen. Sie sollte bevorzugt bis 20 Feet scharf abbilden. Sowohl der Benutzer als auch die Information (falls vorhanden) auf seiner Präsentationstafel müssen beide scharf abge-

bildet werden. Ein Autofokus, bei dem die Kamera dauernd der besten Scharfstellung nachjagt, wenn sich der Benutzer bewegt, erzeugt ein irritierendes Bild auf der Empfängerseite und muss daher vermieden werden.

[0190] Die Kamera sollte eine beschränkte Zoomfähigkeit besitzen, und zwar von einer Einstellung, in der sich ein Benutzer direkt vor der Kamera befindet, bis zu einer Einstellung, in der sich einige Benutzer gleichzeitig auf einem Videophon **15** befinden. Wahlweise kann man verschiedene Linsen bereitstellen. Man kann dafür das Linsensichtfeld festlegen, beispielsweise von einem Blickfeld von 30° bis zu einem Blickfeld von 75°.

[0191] Die Kamera sollte in der Lage sein, ein Bild einzulesen, das größer ist als für die Übertragung erforderlich, beispielsweise ein 1280 × 960 Bild. Dies erlaubt elektronisch einen gewissen Zoom und eine horizontale und vertikale Schwenkung. Damit werden elektromechanische Steuerungen für die Kamera überflüssig. Die Kamera sollte von geringer Baugröße sein, damit sich eine Montage über dem Bildschirm nicht schon allein durch die Größe der Kamera verbietet.

[0192] Ein langlebiger Berührbildschirm mit mittlerer Auflösung stellt das primäre Kommunikationsverfahren mit dem Videophon **15** dar und bildet die Vorderseite des Haupt-Bildschirms **54**. Die Anzeige wird häufig mit den Fingern berührt und muss daher beständig gegen wiederholte Reinigungsvorgänge sein, um Schmierer und andere Fingerabdrücke zu beseitigen, die sonst die Qualität des Bildschirms **54** beeinträchtigen würden. Es sollte leicht möglich sein, den Berührbildschirm zu kalibrieren, d. h. sicherzustellen, dass die Ausrichtung zwischen der auf dem Berührbildschirm berührten Fläche und der darunterliegenden Anzeige **54** die "Fehlberührungs-Anforderung" erfüllt.

[0193] Die Oberfläche des Berührbildschirms **74** muss geringstmögliche Reflexionen aufweisen, damit die Anzeige **54** auch dann klar bleibt, wenn sie einem Fenster gegenüberliegt. Gefordert wird, dass "Fehlberührungen" seltene Ereignisse sind. Die Anforderungen an die Auflösung des Berührbildschirms hängt damit stark von der kleinsten Darstellungsfläche **54** ab, die durch Berührung unterschieden werden soll. Die Auflösung und der Parallaxenfehler sollten zusammen so wirken, dass die Wahrscheinlichkeit für eine "Fehlberührung" durch diese Faktoren beim geübten Benutzer weniger als 5 Prozent beträgt (eine Fehlberührung bei 20 Wahlvorgängen). Wünschenswert ist, dass dieses Fehlberührungsverhältnis weniger als 2 Prozent beträgt, d. h. eine Fehlberührung bei 50 Wahlvorgängen.

[0194] Wenn dies angezeigt ist, muss der Anwender

eine hörbare oder sichtbare Rückmeldung bezüglich der erfolgreichen Berührung erhalten. Diese Töne können sich abhängig von der aktuellen Darstellung auf der Anzeige **54** des Berührbildschirms **74** ändern. Wird beispielsweise eine Tastatur verwendet, so sind tastaturähnliche Geräusche geeignet, bei Gebrauch einer Wählscheibe sind wohl andere Klänge naheliegend usw.

[0195] Eine hörbare Rückmeldung muss nicht in jedem Fall erforderlich sein, obgleich in der Regel eine hörbare oder sichtbare Anzeige einer erfolgreichen Berührung den Benutzer unterstützt. Der Benutzer sollte die Töne ein- und ausschalten können und die Töne, die Tondauer und den zur Berührung gehörenden Lautstärkepegel auf einem Einstellbildschirm festzulegen. Voreinstellungswerte sollten vorhanden sein. Den Berührbildschirm **74** kann man mit einem Stift und ebenso mit dem Finger verwenden.

[0196] Die Anzeige **54** sollte mindestens eine Farb-anzeige **54** mit Flachbildschirm und 17 Zoll-Diagonale (oder mehr) sein, wobei ein Seitenverhältnis von 16 × 9 bevorzugt wird, ein Seitenverhältnis von 16 × 10 aber noch annehmbar ist.

[0197] Die Bildschirmauflösung sollte mindestens 1280 × 768 betragen. Der Betrachtungswinkel sollte mindestens sowohl in der horizontalen als auch der vertikalen Ebene 6° von der Achse betragen. Das Bildschirm-Kontrastverhältnis sollte besser als 300:1 (typisch) sein. Die Farbauflösung sollte mindestens 6 Bit pro Farbe betragen, d. h. in der Lage sein, 262K Farben mit 6 Bit je Farbe darzustellen; dies ist für Prototypenheiten annehmbar. 8 Bit pro Farbe werden für die Produktionseinheiten bevorzugt, oder andere Werte, die äquivalent sind. Der Bildschirm **54** sollte so hell sein, dass man ihn in einem gut beleuchteten oder mit Tageslicht beleuchteten Raum ohne Anstrengung betrachten kann. Die Helligkeit sollte mindestens 300 cd/m² betragen. Die Anzeige **54** und die Decodierelektronik sollte in der Lage sein, hochauflösende 720P-Bilder von geeigneten Quellen für derartige Bilder im Netz **40** darzustellen.

[0198] Die Hintergrundbeleuchtung sollte bei mindestens 50 Prozent Helligkeit eine Mindestlebensdauer von 25000 Stunden haben. Wird die Hintergrundbeleuchtung abgeschaltet, weil das Terminal des Videophons **15** nicht aktiv ist, so sollte sie sich automatisch einschalten, wenn ein Gespräch ankommt und wenn der Benutzer den Berührbildschirm irgendwo berührt. Der Benutzer sollte die Zeitspanne einstellen können, nach der der Berührbildschirm bei fehlender Aktivität abschaltet. Es sollte auch die Einstellung "keine Abschaltung" möglich sein.

[0199] Die im Anschlussbereich des Videophons **15** erforderlichen Verbindungen sind in [Fig. 6](#) dargestellt. In den folgenden Abschnitten wird jeder benö-

tige Stecker kurz beschrieben.

[0200] Zwei 10/100-Ethernetstecker RJ **45** dienen dem Anschluss an das Netz **40** und an den zugehörigen PC **68**.

[0201] Ein optionaler Stecker im ATM-Personality-Modul sollte bereitgestellt werden, damit das Videophon **15** leicht Schnittstellen für 155 Mbit/sec sowohl auf Lichtleitern als auch Kupferkabeln unterstützen kann.

[0202] Ein USB-Port sollte bereitgestellt sein, damit man verschiedene optionale Peripheriegeräte leicht anschließen kann, beispielsweise eine Tastatur, eine Maus, eine kostengünstige Kamera usw.

[0203] Eine 1394-Schnittstelle (Firewire) sollte bereitgestellt sein, damit der Anschluss von externen Kameras (Firewire) oder anderen Videoquellen möglich ist. Die Schnittstelle sollte eine vollständige Inbandkontrolle der Kamera über die Firewire-Schnittstelle ermöglichen. Bei Bedarf sollten externe Umsetzer verwendet werden, um beispielsweise von S-Video auf Firewire-Eingangssignale umzusetzen. Es sollte möglich sein, diese Quelle anstelle der Hauptkameraquelle im Videophon **15** als Ausgabe für die Konferenz zu verwenden. Es sollte auch möglich sein, für diese Videoquelle den normalen oder den "CNN-Modus" festzulegen, d. h. den beschneidbaren oder den nicht beschneidbaren Modus. Eine XVGA-Videoausgabe sollte möglich sein, damit das Videophon **15** externe Projektoren mit einem Bild ansteuern kann, das das auf dem Hauptbildschirm **54** dargestellte Bild widerspiegelt.

[0204] Für die Audio-Ausgangssignale des PC sollte ein Audioeingang vorhanden sein. Um die Audiointegration des PC **68** und des Videophons **15** sicherzustellen, wird nur ein Satz Lautsprecher **64** eingesetzt. Die Töne des PC **68** durchlaufen den Audiokanal des Videophons **15**. Ein Stecker oder ein Steckerpaar sollen dazu dienen, einen Kopfhörer und ein daran befestigtes Bügelmikrofon anzuschließen. Es muss auch möglich sein, nur den Kopfhörer zu betreiben, wobei die eingebaute Mikrofonanordnung verwendet wird. Ist der Kopfhörerstecker nur schwer zugänglich, so sollte es möglich sein, den Kopfhörer eingesteckt zu lassen und über die Benutzersteuerung zu wählen, ob der Kopfhörer mit Audiosignalen versorgt wird oder nicht. Für einen externen linken und rechten Lautsprecher **64** sind Anschlüsse vorhanden. Es ist möglich, ein, zwei oder drei Videophoneinheiten **15** so zu verwenden, als wären sie eine einzige Funktionseinheit, siehe [Fig. 7](#).

[0205] In Anordnungen mit mehr als einem Videophon **15** arbeitet nur eine Einheit als Hauptkontrollbildschirm. Auf der anderen Einheit (bzw. den anderen Einheiten) wird Video dargestellt sowie die Steu-

ermöglichkeiten, die direkt zum dargestellten Video gehören. Für jede beliebige derartige Anordnung wird nur ein Satz Lautsprecher **64** benötigt.

[0206] Hinsichtlich der Mikrofoneingaben und der Audioströme sollten etliche Optionen bereitgestellt werden, beispielsweise von der Verwendung einer einzigen gemeinsamen Mikrofoneingabe bis zur Übertragung des Audios von jeder Mikrofonanordnung zu den Videoquellen an diesem Videophon **15**.

[0207] Für die Videoeingaben sollten etliche Optionen bereitgestellt werden. Die Voreinstellung sollte sein, die Ansicht vom "Schaltpult-Videophon **15**" zu übertragen. Steht mehr Bandbreite zur Verfügung, so kann jeder Benutzer das Video von dem Bildschirm erhalten, auf dem der Benutzer dargestellt ist. Dies führt zu einem natürlicheren Eindruck. Die gesamte Koordination der zahlreichen Videophon-Terminals **15** kann über die LAN-Verbindung erfolgen, d. h., es ist keinerlei zusätzliche Verkabelung zwischen den Einheiten erforderlich.

[0208] Das Videophon **15** liefert seinem Benutzer eine Anzahl Hauptfunktionen.

- Es ist das Bürotelefon.
- Es ist das Telefon des Benutzers.
- Es ist ein Videophon.
- Es ist ein Konferenztelefon.
- Es ist ein Videokonferenztelefon.
- Es bietet einen leichten Zugriff auf Kontakteinzelheiten und deren Verwaltung.
- Es bietet Zugriff auf und Verwaltung von Sprach- bzw. Videomail.

[0209] Die Funktionalität der Einheiten teilt sich in zwei Kategorien auf, nämlich Benutzerfunktionen und Systemfunktionen.

[0210] Benutzerfunktionen sind sämtliche Funktionen, auf die der Benutzer Zugriff hat.

[0211] Die Funktionen des Systems **10** sind diejenigen Funktionen, die Informationstechniker zum Einstellen des Monitors und zum Warten des Videophon-Terminals **15** benötigen und die für den normalen Benutzer nicht sichtbar sind. In der Tat besteht eine wichtige Aufgabe des Gesamtentwurfs darin, sicherzustellen, dass dem Benutzer eine sehr einfache Schnittstelle präsentiert wird, mit der er das Videophon **15** nahezu ohne Übung bedienen kann.

[0212] Im Folgenden wird der grundlegende Satz an Merkmalen definiert, der den kleinstmöglichen Satz an Merkmalen darstellt, die verfügbar sein müssen.

[0213] Das Videophon **15** arbeitet als herkömmliches Telefon, solange kein Benutzer am Terminal angemeldet ist. Dessen Funktionalität darf in keiner Weise davon abhängen, ob ein zugehöriger PC **68**

vorhanden ist.

[0214] Im Folgenden wird die Funktionalität des Videophons **15** als herkömmliches Telefon in einem Büro beschrieben.

[0215] Das Terminal kann eine herkömmliche Nebenstellenummer an der PABX haben, die die Site bedient.

[0216] Das Terminal kann einen eingehenden Anruf von jedem beliebigen Telefon annehmen, d. h. unterschiedslos von der PABX, dem Netz **40** des Videophons **15** oder irgendeinem externen Telefon.

[0217] Das Videophon **15** kann Anrufe von anderen kompatiblen SIP-Telefonen annehmen.

[0218] Ein eingehender Anruf erzeugt einen Klingelton gemäß der Konfiguration (siehe die Anforderungen an den Einstell-Bildschirm unten). Insbesondere unterscheidet sich der Klingelton für Anrufe am Videophon **15**, die Video enthalten, wahlweise vom Klingeln für reine Audiogespräche, und zwar unabhängig davon, ob sie von Videophonterminals **15** stammen oder nicht.

[0219] Ein eingehender Anruf erzeugt eine Anzeige für einen eingehenden Anruf im Statusfeld auf dem Bildschirm **54**. Die Anzeige **54** muss die gesamte im eingehenden Anruf enthaltene Information über die Kennung des Anrufers angeben oder anzeigen, dass eine derartige Information nicht vorhanden ist.

[0220] Es ist möglich, den eingehenden Anruf anzunehmen durch:

- a) das Drücken des Gesprächsannahmeknopfs auf der Statusanzeige **54** des eingehenden Anrufs;
- b) durch das Abheben des Hörers; dadurch werden stets alle angebotenen Optionen angenommen, d. h. Video und Audio.

[0221] Innerhalb eines Gesprächs kann der Anwender leicht vom Hörerbetrieb auf den freihändigen Betrieb (Mithören über Lautsprecher) überwechseln. Das Abheben des Hörers innerhalb eines Gesprächs sollte automatisch aus dem Lautsprecher-Mithörmodus in den Hörermodus umschalten. Das Auflegen des Hörers ohne eine erneute Wahl des Lautsprecher-Mithörmodus unterbricht das Gespräch.

[0222] Der Modus, d. h. über den Hörer oder freihändig, sollte auf dem Bildschirm dargestellt werden.

[0223] Der Gesprächs-Statusbalken kann die Gesprächsdauer darstellen.

[0224] Es ist möglich, die Lautstärke des eingehenden Anrufs über leicht zugängliche Steuermöglich-

keiten auf dem Hauptbildschirm **54** einzustellen. Die Lautstärken von Kopfhörer und Lautsprecher sollten unabhängig einstellbar sein.

[0225] Im Lautsprecher-Mithörmodus kann man den Hörer in die Hörerhalterung zurücklegen, ohne das Gespräch zu beenden.

[0226] Ein Gespräch wird dadurch beendet, dass:

- a) der Benutzer den Gesprächsendeknopf auf der Gesprächsstatusanzeige am Bildschirm **54** drückt;
- b) der Benutzer im Hörermodus den Hörer auflegt, wenn freihändiges Telefonieren nicht gewählt ist;
- c) der entfernte Teilnehmer das Gespräch beendet, vorausgesetzt, dies wird dem Videophon **15** verlässlich angezeigt.

[0227] HALTEN – Es sollte möglich sein, ein Gespräch in den Halte-Status zu versetzen und das Gespräch wieder aus dem Halte-Status herauszunehmen. Der Halte-Status sollte auf der Statusanzeige **54** dargestellt werden, und zwar mit einem Knopf, der es erlaubt, das gehaltene Gespräch wieder aufzunehmen.

[0228] GESPRÄCH WARTET – Zusätzliche eingehende Gespräche müssen in der Statusfläche des Bildschirms **54** eine Anzeige für eingehende Anrufe erzeugen. Es darf kein Anrufton erzeugt werden, solange dies nicht im Einstellungs Menü freigegeben ist.

[0229] Es ist möglich, einen neuen eingehenden Anruf im momentanen Betriebsmodus anzunehmen, d. h. über Hörer oder freihändig, und zwar mit dem Gesprächsannahmeknopf auf der Statusanzeige **54**.

[0230] Das Annehmen eines weiteren eingehenden Anrufs versetzt die derzeit laufenden Gespräche automatisch in den HALTE-Status.

[0231] Das Drücken des Knopfs "Halten beenden" bei irgendeinem Gespräch muss automatisch alle anderen Gespräche in den HALTE-Status versetzen.

[0232] Die Anzahl gleichzeitig eingehender Gespräche, die bearbeitet werden können, wird durch die Verfügbarkeit von Raum für die Statusanzeige **54** bestimmt. Es dürfen nicht weniger als zwei Anrufe sein.

[0233] Ist die Anzahl der momentan anstehenden Gespräche größer als die Anzahl, die abgewickelt werden können, so gilt für alle weiteren eingehenden Gespräche:

- a) sie erhalten das Besetztzeichen;
- b) sie werden sofort zur Sprachmail weitergeleitet;
- c) sie werden sofort zur eingestellten Weiterleitungsnummer durchgeschaltet;
- d) sie erhalten eine aufgezeichnete Nachricht.

[0234] Dies wird in den Benutzereinstellungen für das "Weiterleiten bei besetzt" festgelegt.

[0235] Werden eingehende Gespräche innerhalb der annehmbaren Grenze nicht innerhalb eines (konfigurierbaren) Intervalls beantwortet, so werden die Gespräche:

- a) zur Sprachmail weitergeleitet;
- b) zur vorkonfigurierten Weiterleitnummer durchgeschaltet;
- c) mit einer aufgezeichneten Nachricht beantwortet.

[0236] Dies wird in den Benutzereinstellungen für das "Weiterleiten, falls keine Antwort erfolgt" festgelegt.

[0237] RUFWEITERVERBINDUNG – Der Benutzer kann leicht jedes beliebige Gespräch an jede beliebige andere Nummer weiterverbinden. Die Weiterverbindungsfunktion versetzt das Gespräch in den Haltestatus und erlaubt das Wählen einer neuen Nummer. Beim Ertönen des Klingeltons hat der Benutzer die Option, die Weiterverbindung zu vervollständigen. Wahlweise kann der Benutzer mit der neuen Nummer sprechen und dann entweder die Weiterverbindung anstoßen oder zuerst alle (drei) Parteien in einem Konferenzgespräch verbinden. Erfolgt letzteres, so wird dem Benutzer eine Funktion zum Verlassen des Konferenzgesprächs bereitgestellt. Antwortet das angerufene Terminal nicht oder nur mit einer Sprachmail, so hat der Benutzer die Möglichkeit, zum ursprünglichen Gespräch zurückzukehren.

[0238] RUFUMLEITUNG – Das Telefon muss so einstellbar sein, dass eingehende Gespräche automatisch an eine vorkonfigurierte Nummer umgeleitet werden. Die Rufumleitung kann erfolgen:

- a) bedingungslos;
- b) wenn die Nummer besetzt ist;
- c) wenn keine Antwort erfolgt.

[0239] KONFERENZGESPRÄCHE – Es ist möglich, die Gespräche nur zu einer Audiokonferenz zusammenzuschalten, und zwar unabhängig vom Ursprung des Sprachanrufs. Es ist möglich, mindestens drei Gespräche zu einer Konferenz zu verbinden, d. h. eine Vierwege-Unterhaltung herzustellen. Es wird nur gefordert, zu jedem beliebigen Zeitpunkt nur eine Konferenz zu unterstützen; es muss jedoch möglich sein, ein weiteres eingehendes Gespräch wie oben bei "Gespräch wartet" beschrieben anzunehmen. Es ist hinnehmbar, wenn der Prototyp nur in der Lage ist, ein eingehendes Gespräch zu einer bestimmten Konferenz anzunehmen, d. h. man benötigt eine externe Bridge für Gespräche, die nicht von Videophonen stammen.

[0240] Optionen, die zur Statusanzeige **54** des eingehenden Anrufs gehören, erlauben es dem Benut-

zer, ein Gespräch einer Konferenzverbindung zuzufügen oder es daraus zu entfernen.

[0241] Es ist möglich, Gespräche einer Konferenz unabhängig davon, ob es sich um eintreffende oder abgehende Gespräche handelt, zuzufügen.

[0242] Legen entfernte Konferenzteilnehmer auf, so muss der Gesprächszweig automatisch gelöscht werden.

[0243] Die Gespräche können freihändig oder auch über den Hörer abgewickelt werden. Durch das Abheben des Hörers sollte die Wahlkastatur angezeigt werden, wenn gerade kein Gespräch stattfindet, und das Audio sollte mit dem Hörer verbunden werden. Man benötigt eine Tonwahlkastatur auf dem Bildschirm (d. h. die Zahlen von 1 bis 0 und "*" sowie "#"). Zusätzlich sollte ein Pausenknopf zum Einfügen einer Pause in eine gewählte Folge vorhanden sein (um durch PABX-Anlagen durchgeleitet zu werden, falls man den oder die Gateways **70** nicht so programmieren kann, dass diese Forderung entfällt). Man kann in Betracht ziehen, eine Taste "+" zuzufügen und Vorkehrungen zu treffen, die das Zeichen "+" automatisch in die internationale Ländervorwahl für den entsprechenden Ort umsetzen.

[0244] Man benötigt auch eine Taste zum Korrigieren von Eingabefehlern (z. B. die Taste [BACK]) und eine Lösch Taste zum Löschen der Eingabe. Ein kurzer Druck auf die Taste [BACK] sollte die letzte eingegebene Zahl entfernen. Auf eine längere Betätigung hin sollten weitere Zahlen entfernt werden. Ein Dauerdruk sollte das Zahlenregister löschen.

[0245] Die Zahlendarstellung **54** sollte automatisch auf das ortsübliche Zahlenformat formatiert werden. [Dies kann eine Benutzereinstellung zur Auswahl des Landes erfordern, in dem das Gerät betrieben wird, da jedes Land seine eigene Form hat. Wird ein internationaler Code eingegeben, so sollte dieser Code als Grundlage für das Formatieren des restlichen Teils der Zahl dienen.]

[0246] Besteht eine Verbindung zu Diensten, die die Tonwahlkastatur zum Auswählen von Merkmalen verwenden, so müssen die korrekten Töne in der Richtung dieses Dienstes erzeugt werden, wenn die Bildschirmkastatur oder die Tastatur am Hörer verwendet wird. Die Wahlkastatur muss diese Funktion unabhängig davon liefern können, wie das Gespräch aufgebaut wird.

[0247] WAHLWIEDERHOLUNG – Es ist möglich, die zuletzt gewählte Nummer durch eine einzige Berührung einer geeignet gekennzeichneten Funktion erneut zu wählen.

[0248] AUTOMATISCHE WAHLWIEDERHOLUNG

– Es ist möglich, einen automatischen Wahlwiederholungsmechanismus anzustoßen, beispielsweise durch das Halten des Knopfs [WAHLWIEDERHOLUNG]. Die automatische Wahlwiederholung wiederholt automatisch den Anruf in mehreren Versuchen, falls die vorhergehenden Versuche ein Besetzt-Zeichen zurückliefern.

[0249] WARTEN AUF VERBINDUNG – Bei einem Anruf an eine Vorrichtung, die dieses Merkmal unterstützt, ist eine Funktion "Warten auf Verbindung" verfügbar. Warten auf Verbindung ruft den Benutzer zurück, sobald der angerufene Teilnehmer erreichbar ist. Es soll eine Meldung erzeugt werden, die mitteilt "dieser Dienst ist nicht erreichbar", falls die angerufene Nummer "Warten auf Verbindung" nicht unterstützen kann.

[0250] Ist kein Benutzer am Videophon **15** angemeldet, so kann ein geeigneter Bildschirm zum Anmelden dargestellt werden.

[0251] Eine Protokollierung der eingehenden, abgehenden, häufigen und verpassten Gespräche sollte in einer geeigneten Ansicht des integrierten Wahlbildschirms dargestellt werden. Auf den Wahlbildschirmen sollte mit ein oder zwei Berührungen ein Zugriff auf die "erneute Wahl der letzten Nummer" stets möglich sein. Diese Protokolle werden im Weiteren genauer bestimmt.

[0252] Ein Benutzer muss sich am Terminal anmelden, damit er Zugriff auf den vollständigen Satz der verfügbaren Merkmale im Terminal des Videophons **15** hat. Es wird ein Anmelde-Bildschirm bereitgestellt, in den der Benutzer seinen Namen und sein Passwort eintragen kann. Es kann sich dabei um den Namen und das Passwort handeln, mit denen der Benutzer normalerweise auf das Netz **40** zugreift. Das Videophon **15** verwendet daher die Benutzer-Berechtigungsdienste der Site. Es müssen sämtliche Bildschirme bereitgestellt werden, die das IT-Personal zum Konfigurieren des Videophons **15** benötigt, damit es diese Benutzer-Berechtigungsdienste verwenden kann. Andere Verfahren zum Identifizieren des Benutzers sind verfügbar, beispielsweise der Gebrauch einer Smart Card oder ID-Fob. Es ist nicht erforderlich, dass der Benutzer stets an einem PC **68** angemeldet ist, bevor er sich an einem Videophonterminal **15** anmeldet.

[0253] Es können sich mehrere Benutzer an einem einzigen Videophon **15** anmelden, und es können unterschiedliche Klingeltöne für jeden Benutzer bereitgestellt werden. Die Anzeige des eingehenden Gesprächs sollte die angerufenen Parteien nennen und den Namen der anrufenden Partei angeben. Sind mehrere Benutzer an einem einzigen Videophon **15** angemeldet, so sind alle Rufweiterleitungsfunktionen nur dem Benutzer zugänglich, an den der Anruf ge-

richtet ist.

[0254] Ist der Benutzer bereits an seinem PC **68** angemeldet, so soll der Anmeldevorgang am Videophon **15** einen Zusammenhang zwischen dem PC **68** herstellen, an dem der Benutzer angemeldet war, und dem Terminal des Videophons **15**, vorausgesetzt dass dies vom PC **68** bestätigt wird. Ein Benutzer kann sich gleichzeitig an den Terminals mehrerer Videophone **15** anmelden. Das aktive Videophon **15** ist dasjenige Videophon, an dem sämtliche Gespräche für diesen Benutzer zuerst beantwortet werden.

[0255] Der Startseiten-Bildschirm enthält einen Statusbereich, der in allen Bildschirmen sichtbar ist (mit Ausnahme des Vollbildmodus). Der Status enthält den Namen des angemeldeten Benutzers oder die Angabe "kein Benutzer angemeldet". Der Status enthält zudem den "Anwesenheitsstatus" des Benutzers, Icons für die Video- und Audioübertragung, eine "Meldungsanzeige" für Sprachnachrichten und Datum und Uhrzeit.

[0256] Eine "Meldungsanzeige" ist hell dargestellt und blinkt, wenn sich eine noch nicht angehörte Sprachnachricht im Sprachmailsystem **10** des Benutzers befindet. Ein Drücken der Anzeige führt dazu, dass der Bildschirm zum Bearbeiten von Sprachnachrichten erscheint.

[0257] Das Berühren des Bereichs für Datum und Uhrzeit ermöglicht den Zugriff auf die Kalenderfunktionen.

[0258] Die Startseite besitzt einen Steuerbalkenbereich, der in allen Bildschirmen (mit Ausnahme des Vollbildmodus) sichtbar ist.

[0259] Der Steuerbalken erlaubt einen direkten Zugriff auf die am häufigsten verwendeten Steuermerkmale und Zugriff auf alle anderen Funktionen. An den Knöpfen sollten Icons verwendet werden. Es kann jedoch auch Text verwendet werden, um den Zweck der Funktion zu unterstreichen.

[0260] Die Steuertafel weist auch globale Einstellmöglichkeiten für das Mikrofon, die Kamera und die Lautsprecher **64** auf. Die Einsteller sollten ihren Betriebszustand deutlich darstellen, beispielsweise "Ein" oder "Aus". Icons sollten soweit wie möglich verwendet werden.

[0261] Ein Selbstbild ist verfügbar, das sowohl das Bild anzeigt, das die Kamera aufnimmt, als auch denjenigen Abschnitt, der auf der anderen Seite des aktiven Gesprächs zu sehen ist. Das Selbstbild kann ein- und ausgeschaltet werden, und man kann festlegen, ob es stets eingeschaltet ist oder nur dann, wenn ein aktives Gespräch aufgebaut ist.

[0262] Man kann zu jedem beliebigen Zeitpunkt das Kamerabild im Hauptvideobereich des Bildschirms darstellen, d. h., in einem Gespräch, wenn kein Gespräch geführt wird, usw. Das Bild sollte das Bild für ein einzelnes Videogespräch sein und alle anderen vorhandenen Videodarstellungen überlagern. Es sollte möglich sein, eine Vollbildversion dieses Videos anzufordern. Man kann dies als digitalen Spiegel betrachten, der es dem Benutzer erlaubt, sicherzustellen, dass er mit dem Bild, das die Kamera zeigt (oder zeigen wird) zufrieden ist.

[0263] Für Diagnosezwecke ist es erwünscht, dass der Benutzer das Bild auch nach dem Codieren und Decodieren sehen und sich von der Bildqualität überzeugen kann, die am anderen Ende zu sehen ist. Wird dieser Modus unterstützt, so sind das unmittelbare Kamerabild und das codierte und decodierte Bild nebeneinander zu sehen. Der Benutzer kann sein Selbstbildnis aufnehmen und es als Bild verwenden, das zu seiner Kontaktinformation gehört.

[0264] Der Hauptteil des Startbildschirms ist den integrierten Wählfunktionen zugewiesen. Es gibt vier wesentliche Unterfunktionen, nämlich eine Kurzwahl-darstellung **54**, eine Darstellung **54** für den Zugriff auf Verzeichnisse, eine Wahl-tastatur und einen Zugriff auf protokollierte Gespräche. Die Wahl-tastatur und der Zugriff auf protokollierte Gespräche nehmen nur eine geringe Bildschirmfläche ein und sind leicht zu bedienen. Dadurch wird die verfügbare Fläche für die Kurzwahl- und Kontaktseiten so groß wie möglich. Zuerst wird der Kurzwahlbereich ausführlich beschrieben. Alle gemeinsamen Anforderungen über alle wesentlichen Unterfunktionen hinweg werden nur für den Kurzwahlbereich ausführlich beschrieben und gelten für die drei anderen Funktionen genauso. Der Wahlbereich hat die Funktion, einen Benutzer auszuwählen, mit dem ein Gespräch geführt werden soll.

[0265] Der Kurzwahlbereich ist innerhalb des Rahmens, der mit den anderen Anforderungen für den Wahlbildschirm verträglich ist, so groß wie möglich. Angemessen sind mehr als 20 Kurzwahlplätze. Jeder Platz sollte so groß sein, dass man die Einzelheiten der an diesem Platz abgespeicherten Person aus der normalen Arbeitsentfernung vom Bildschirm, etwa 3 Feet, noch sehr leicht lesen kann.

[0266] Die in einem Kurzwahlplatz gespeicherte Information umfasst den Namen der Person, den "Anwesenheitsstatus" falls bekannt, die Nummer, die bei der Auswahl dieses Kurzwahlplatzes gewählt wird, und ein Icon, das anzeigt, ob der Benutzer Videogespräche unterstützt. In der ausführlichen Information ist auch abgespeichert, um welche Art von Video es sich handelt, z. B. Videophon **15**, kompatibel zu MPEG-2, H.261 usw.

[0267] Die Fläche bietet einen leeren Bereich, den man zum Anstoßen eines Gesprächs berühren muss. Falls verfügbar ist ein Miniaturbild der Person enthalten. Ein Verfahren zum Handhaben langer Namen (d. h. Namen, die nicht in den Raum passen, der dem Kurzwahlknopf zugewiesen ist) wird bereitgestellt.

[0268] Herkömmliche Telefonnummern im internationalen Standardformat, d. h. "+ Ländercode Vorwahl Nummer" werden automatisch auf die externe Zugangsnummer plus die internationale Zugangsnummer umgesetzt, die für den Anruf bei dieser Nummer erforderlich ist.

[0269] Die vollständigen Kontaktdetails, die zu einer Person auf der Kurzwahlseite gehören, sind verfügbar. Die Kontakteinzelheiten liefern sämtliche Nummern, unter denen der Benutzer zu erreichen ist, sowie ein Hilfsmittel zum Auswählen einer dieser Nummern als Voreinstellungsnummer, die auf der Kurzwahlseite verwendet wird. Es ist möglich, über die Verknüpfung mit der Kontaktseite eine andere Nummer für diesen Benutzer auszusuchen und zu wählen.

[0270] Die Benutzerinformation enthält eine Aufzeichnung der zuletzt geführten Gespräche für diese Person, beispielsweise die letzten 10 Gespräche unabhängig davon ob es sich um ankommende, verpasste oder abgehende Gespräch handelt. Nur die Information über das "letzte Gespräch" bereitzustellen wäre ein annehmbares Minimum an Funktionalität.

[0271] Man kann die Kontakteinzelheiten überarbeiten, die zu dem Kurzwahleintrag gehören, oder einen neuen Kontakteintrag für die Kurzwahlseite erzeugen. Man kann einen Eintrag aus den Kontaktbildschirmen, den Verzeichnissen oder den Anrufaufzeichnungs-Bildschirmen in die Kurzwahlseite kopieren. Man kann einen Eintrag von der Kurzwahlseite in die Kontakt- oder Verzeichnisbildschirme kopieren. Man kann einen Kurzwahleintrag löschen oder diesen Eintrag auf eine andere Kontaktseite verschieben (d. h. kopieren und anschließend das Original löschen).

[0272] Man kann die Anordnung der Benutzer auf der Kurzwahlseite einstellen. Es sollte auch auf irgendeine Weise möglich sein (Farbcodierung), zwischen verschiedenen Klassen von Kurzwahlbenutzern zu unterscheiden, d. h. zwischen Unternehmen, Familie, Kollegen, Zulieferern, Kunden. Die Kurzwahlseite kann ebenso Namen aus zahlreichen anderen Kategorien in der Kontaktinformation enthalten. Eine gewisse Form der automatischen Organisation ist möglich, beispielsweise Nachname, Vorname, Firma oder sortiert nach Klasse gefolgt von Nachname, Vorname, Firma usw.

[0273] Man kann eine Benutzergruppe als einzigen Kurzwahleintrag definieren. Es ist annehmbar, wenn die Gruppengröße auf den größtmöglichen Umfang eines Konferenzgesprächs beschränkt ist. Man kann die Verzeichnisansicht aus der Kurzwahlseite wählen. Die Verzeichnisansicht nimmt die gleiche Bildschirmfläche ein wie Kurzwahlseite. Man kann aus dem Bereich der Online-Verzeichnisse auswählen, auf die das Videophon **15** Zugriff hat. Die Voreinstellung ist das Verzeichnis von Outlook und/oder Lotus Notes, das die wesentlichen Kontakteinheiten des Benutzers enthält. Der Name des gewählten Verzeichnisses sollte dargestellt werden.

[0274] Die Kategorien, die der Benutzer in seinen Outlook- oder Notes-Kontaktlisten erstellt hat, können gewählt werden. Passt die Anzahl der Kategorien nicht in den Bereich der Anzeige **54**, so werden Knöpfe zum Aufwärts- oder Abwärtsrollen der Liste bereitgestellt. Die Liste sollte alphabetisch geordnet sein.

[0275] Die Kurzwahl-Kategorie ist die Kategorie, die zum Füllen der Kurzwahlseite verwendet wird. Es wird angezeigt, wenn die Kurzwahlseite gefüllt ist und es nicht mehr möglich ist, dieser Kontaktkategorie weitere Namen zuzufügen, es sei denn, sie ersetzen einen vorhandenen Eintrag. Die Fähigkeit ist vorhanden, die Kurzwahleinträge geordnet nach dem zuletzt geführten Gespräch aufzureihen, d. h. der am wenigsten verwendete Kurzwahleintrag befindet sich ganz unten. Dies würde es erlauben, zu sehen, welcher Eintrag sich für das Löschen am besten eignet, damit man eine häufiger verwendete Nummer eintragen kann.

[0276] Es ist möglich, einen Eintrag aus der gewählten Kategorie mit sehr wenig Benutzereingaben leicht zu finden und auszuwählen. Der Auswahlmechanismus für Einträge muss für relativ kurze Listen und für sehr lange Listen arbeiten (mehrere 10000 Namen). Der Mechanismus muss die Möglichkeit umfassen, eine Zeichenkette einzugeben, nach der zu suchen ist. Man kann die Sortierreihenfolge für die dargestellten Daten ändern, beispielsweise nach dem Nachnamen, dem Vornamen oder der Organisation. Es gibt ein Verfahren, Eingabefehler zu korrigieren und die gesamte Suche rasch neu zu starten.

[0277] Es ist erwünscht, dass jede Reihenfolge der Suchtasten Bedeutung hat und vom Benutzer verändert werden kann. Anders formuliert erlaubt beispielsweise das Drücken und Halten der ganz linken Suchtaste dem Benutzer, nach dem Nachnamen, dem Vornamen oder der Firma zu suchen (Die Suche nach einer erweiterten Attributliste ist möglich. Dies ist beispielsweise nützlich, um eine Person in einer bestimmten Abteilung zu finden oder an einem bestimmten Ort – "wer ist in Korea"). Die zweite Taste modifiziert die Suche mit der ersten Taste usw. Die

Tasten sind damit auf Firma, Nachnamen, Vornamen eingestellt, beispielsweise Marconi, und führen eine alphabetische Benutzersuche innerhalb der Nachnamen bei Marconi durch. Nach der Auswahl einer jeden Sortierkategorie gibt es natürlich eine implizierte Unterreihenfolge der Einträge mit den gleichen Werten in diesem Kategoriefeld. Ist der Nachname gewählt, so ist die implizierte Unterreihenfolge Vorname und anschließend Firma. Für Firma ist die implizierte Sortierreihenfolge Nachname, Vorname, und für Vorname beispielsweise Nachname, Firma.

[0278] Der Gesprächsprotokollschirm stellt die jüngsten Einträge von drei Gesprächskategorien dar, nämlich abgehende, eintreffende und verpasste Gespräche. Dabei wird deutlich angezeigt, welche Kategorie gewählt ist. Zusätzlich sollte eine Kategorie "Häufigkeit" vorhanden sein, die die Nummern nach der Häufigkeit der Verwendung auflistet, und zwar über die letzten (weniger als 200) Gespräche jedes beliebigen Typs. Vom Gesprächsprotokollschirm aus sollte man auf die Wahlstatistik zugreifen können. Die Untersuchung, ob es sinnvoll ist, einen wesentlich größeren Grad an Handhabbarkeit der Gesprächsprotokolldaten bereitzustellen, wird aufgeschoben.

[0279] Wird "Nachricht" berührt, so wird zumindest eine Verbindung zum Sprachnachrichtensystem **10** des Benutzers aufgebaut, die Sprachmails für diesen Benutzer werden eingelesen, und die Wahlstatistik wird dargestellt, damit man die Sprachmail mit Hilfe der üblichen Betätigung der Telefontasten bedienen kann. Der größere Teil des "Sprachnachrichten"-Bildschirms sollte Knöpfe für den Zugriff auf jedes Merkmal des Mailsystems **10** zeigen, beispielsweise für: folgende Nachricht, vorhergehende Nachricht, Wiedergabe der Nachricht, Weiterleiten der Nachricht, Beantworten der Nachricht, Anruf an den Absender usw., wozu innerhalb jeder Funktion die entsprechenden Tastendrucke gehören, beispielsweise: Beginn der Aufzeichnung, Ende der Aufzeichnung, Anhören der Aufzeichnung, Löschen der Aufzeichnung usw. Alle diese Funktionen müssen auf Knöpfen liegen, die in die entsprechenden DMF-Töne konvertiert werden.

[0280] Es ist erwünscht, dass die "Weiterleit"-Nummer oder jeder beliebige Sprachmailbefehl, der eine Liste der Benutzernummern erfordert, die eingegeben werden müssen, von der Kurzwahl- oder Verzeichnisdarstellung aus gewählt werden kann, und dass die Auswahl automatisch nur den geeigneten Teil der Benutzernummern eingibt. Dies kann insbesondere beim Weiterleiten einer Sprachnachricht an eine Gruppe nützlich sein. Der Benutzer kann am Videophon **15** die Uhrzeit und das Datum einstellen. Es ist erwünscht, dass die Uhrzeit und das Datum über geeignete Dienste des Netzes **40** automatisch eingestellt werden können.

[0281] Es ist erwünscht, dass Kalenderfunktionen verfügbar sind, die in die Benutzeranwendungen Outlook/Palm/Notes, Scheduler/Kalender integriert sind. Die Mindestanforderung wäre, einfach die festgelegten Termine zu jedem beliebigen Datum, für einen Tag, für eine Woche oder für einen Monat (wie bei Outlook oder Palm-Bildschirmen) zu sehen, wobei Änderungen und neue Einträge nur über die Outlook- oder Palm-Datenbanken möglich sind.

[0282] Es ist zu vermuten, dass einige wenige Benutzer keinen eigenen Kalender führen und in der Tat keinen PC **68** auf ihrem Schreibtisch haben, jedoch die Information einsehen müssen. Das Berühren der Fläche Benutzerstatus im Statusteil des Bildschirms erlaubt es einem Benutzer, seinen Status zu setzen. Der Benutzer findet eine Reihe Statusoptionen vor, aus denen er auswählen kann. Hierzu gehören:

- i) erreichbar;
- ii) besetzt, und zwar mit einem Gespräch, in dem kein weiteres Gespräch angenommen wird;
- iii) bitte nicht stören, d. h., der Benutzer führt kein Gespräch, wünscht aber nicht unterbrochen zu werden;
- iv) in fünf Minuten zurück;
- v) nicht im Büro;
- vi) im Urlaub.

[0283] Im Fall eines Einzelgesprächs unterstützt das Videophonterminal **15** Ströme zwischen einem eingehenden Strom und der größtmöglichen Anzahl von Strömen in einer Konferenz. Für Videokonferenzen unterstützt das Terminal mindestens vier Verbindungen zu anderen Teilnehmern als Teil eines Einzelgesprächs. Es ist möglich, mindestens zwei unabhängige Gespräche anzunehmen, die nur Audio enthalten, und zwar auch dann, wenn ein Videokonferenzgespräch mit höchstzulässigem Umfang läuft. Dadurch kann in einem Audiogespräch zurückgefragt werden, es kann gehalten und weitergeleitet werden. Das Videophon **15** kann mindestens drei gleichzeitige "Gesprächsfälle" unterstützen, d. h. bis zu drei unabhängige Gespräche. Es kann nur ein Gespräch aktiv sein, d. h. die Gesprächssteuerungen können zu einem Zeitpunkt nur auf ein Gespräch angewendet werden. Es können mehr als ein Gespräch angenommen werden, d. h. das Audio und Video des Benutzers werden an jedes angenommene Gespräch übertragen, und zwar unabhängig davon, ob es aktiv ist oder nicht. Laufende Gespräche können auch in den HALTE-Status versetzt werden, wenn das Audio und Video des Benutzers nicht an den Benutzer im HALTE-Status übertragen wird und das Audio und Video dieses Benutzers ebenfalls unterdrückt wird.

[0284] Der Status eingehender Gespräche wird im Steuer-Anzeigebereich **54** dargestellt. Die Gespräche selbst und die Steuerungen im Gespräch sind im Hauptbereich der Anzeige **54** dargestellt.

[0285] Als Gesprächsstatus sind möglich:

- i) eingehendes Gespräch;
- ii) angenommen und aktiv – das Audio des Benutzers (und das Video bei einem Videogespräch) wird gegebenenfalls über die verschiedenen Stummschaltungen mit diesem Gespräch verbunden, und die Gesprächssteuerungen wirken auf dieses Gespräch;
- iii) angenommen und nicht aktiv – wie oben, die Gesprächssteuerungen wirken jedoch nicht auf dieses Gespräch;
- iv) angenommen und im Haltestatus – das Audio des Benutzers (und das Video bei einem Videogespräch) wird nicht mit diesem Gespräch verbunden;
- v) angenommen und weitergeleitet.

[0286] Zu jedem Gespräch wird der Gesprächsstatus angezeigt. Es kann nur eines der angenommenen Gespräche aktiv sein. Ein angenommenes Gespräch wird durch das Berühren der Gesprächsanzeigefläche **54** aktiviert, die zu diesem Gespräch gehört, oder des Gesprächsstatus in der Steuertafel. Jedes vorherige aktive Gespräch wird dadurch inaktiviert. Eine zweite Berührung schaltet den aktiven Status ab. Eine Anzeige für ein eingehendes Gespräch zeigt an, ob das Gespräch eine Videoverbindung anbietet. Wird nichts angezeigt, so liegt lediglich ein Audiogespräch vor. Die Anzeige für ein eingehendes Gespräch zeigt den oder die Namen der Partei, die zu diesem eingehenden Gespräch gehört. Dies zeigt sofort, ob der Benutzer von einer Einzelperson angerufen oder eingeladen wird, an einer Konferenz teilzunehmen.

[0287] Der Benutzer kann mit dem eingehenden Gespräch in folgender Weise verfahren:

- i) das Gespräch als ausschließliches Sprachgespräch annehmen;
- ii) das Gespräch als Videogespräch annehmen (die Sprache gehört automatisch dazu);
- iii) das Gespräch an die Sprachmail weiterleiten.

[0288] Es ist eine Einstellung verfügbar, in der das Terminal des Videophons **15** eingehende Gespräche selbsttätig beantwortet, und zwar bis zur größtmöglichen Anzahl unterstützter Gespräche. Die selbsttätige Antwort stellt eine Audio- und eine Videoverbindung her, falls dies angeboten wird. Wird ein Gespräch abgewickelt, so sollte der Benutzerstatus automatisch auf "führt ein Gespräch" gesetzt werden. Der Benutzerstatus kehrt zu seinem vorherigen Status zurück (normalerweise "erreichbar"), sobald keine Gespräche mehr aktiv sind.

[0289] Der Benutzer kann konfigurieren, ob die Benutzergesprächsdaten ebenfalls verteilt werden. Hat der Benutzer bereits ein oder mehrere Gespräche angenommen und sind alle Gespräche im HALTE-Status oder nicht aktiv, so erzeugt dieses Ge-

sprach bei der Annahme einen neuen Gesprächsfall. Alle angenommenen aber nicht aktiven Gespräche sehen und hören weiterhin den Benutzer bei der Abwicklung des neuen Gesprächs. Ist eines der angenommenen Gespräche angenommen und aktiv, so wird das neue Gespräch diesem Gespräch zugeschaltet, und alle Parteien des Gesprächs werden in Konferenz mit dem neuen Anrufer geschaltet, wenn das Gespräch angenommen wird.

[0290] Hebt der Benutzer nach mehr als 10 Sekunden nicht ab, so wird der Anruf automatisch so weitergeleitet, wie dies in den Einstellungen für das "Weiterleiten bei fehlender Antwort" festgelegt ist. Wie oben ist der Weiterleitvorgang durch den Benutzer festgelegt, an den der Anruf gerichtet ist. Ist der Benutzerstatus durch "Bitte nicht stören" oder "Belegt" gekennzeichnet oder ist der "Belegt-Status" eingestellt, weil die größtmögliche Zahl an Gesprächen bearbeitet wird, so wird der Anruf "sofort" weitergeleitet, wie dies in den Einstellungen für "Weiterleiten bei belegtem Anschluss" und "Weiterleiten, falls keine Störung gewünscht wird" festgelegt ist und in den Einstellungen für "zeige weitergeleitete Gespräche" modifiziert ist, falls diese implementiert sind.

[0291] Abhängig von den Einstellungen für "zeige weitergeleitete Gespräche" kann der Benutzer wählen, dass er die Anzeige für ankommende Gespräche sieht (für mehr als 5 Sekunden) bevor diese weitergeleitet werden. (Dies bedeutet, dass der Benutzer nichts tun muss, es sei denn, er wünscht, das Gespräch anzunehmen. Im obigen Fall muss der Benutzer dagegen eine aktive Handlung vornehmen.) Dies funktioniert nicht, falls der Besetzt-Status vorliegt, weil das Videophon **15** bereits die größtmögliche Anzahl an Gesprächen bearbeitet.

[0292] Die Möglichkeit, eine (sehr kurze) Textnachricht zu erzeugen, die zusammen mit dem Anruf gesendet wird, stellt einen nützlichen Weg dar, mehr Information über die Wichtigkeit des Anrufs und seine voraussichtliche Dauer zu übertragen. Die Anforderungen im Zusammenhang mit dem Erzeugen einer Nachricht und dem Zufügen zu einem abgehenden Gespräch werden im Folgenden beschrieben. Die eingehende Anruf-Textnachricht sollte, falls vorhanden, zusammen mit dem eingehenden Anruf angezeigt werden. Die Anzeige **54** ist in der Lage, Textnachrichten für mehrere ankommende Gespräche gleichzeitig darzustellen. Die Textnachricht wird auch im Gesprächsprotokoll der eingehenden oder verpassten Gespräche abgespeichert.

[0293] Das Aushandeln der Gesprächsparameter ist auf diejenigen Parameter beschränkt, die man braucht, um das Gespräch innerhalb der Betriebsparameter des Netzes **40** und der derzeitigen Nutzung des Netzes **40** herzustellen. Es sind Einstellungen bereitgestellt, die es dem Benutzer erlauben, seine

Vorlieben für Gespräche zu anderen Videophonterminalen **15** festzulegen, beispielsweise stets Video anzubieten, niemals Video anzubieten, bei jedem Gespräch fragen, ob Video angeboten werden soll oder nicht.

[0294] Das Warten auf eine verfügbare Verbindung wird für Gespräche zu anderen Benutzern von Videophonen **15** unterstützt. Dadurch wird ein Anruf an den Benutzer angestoßen, sobald dessen Status auf "erreichbar" wechselt. Ist der anzurufende Benutzer eine Gruppe, so werden die Gespräche erst dann angestoßen, wenn alle Gruppenmitglieder "erreichbar" sind.

[0295] Ein Konferenzgespräch liegt vor, wenn eine Stelle in der Kurzwahl- oder Verzeichnisliste eine Gruppe von Personen darstellt, wobei jede Person an dem Gespräch teilnehmen will. Die vorgeschlagene Prozedur zum Implementieren dieses Merkmals besteht darin, einen Anruf nach dem anderen vorzunehmen, und nach einer aktiven Bestätigung der Anforderung, den Anruf in die Konferenz aufzunehmen. Dies liefert eine Ausweichmöglichkeit, falls das Gespräch über eine Sprachnachricht erfolgt. Ist die Bearbeitung des ersten Gesprächsteilnehmers abgeschlossen, d. h. das Gespräch ist aufgebaut oder wurde abgewiesen, so wird die folgende Nummer bearbeitet.

[0296] Es ist möglich, einen abgehenden Anruf zu erzeugen, der nur die Eigenschaft "halbduplex" besitzt, d. h., der Audio und/oder Video von der angerufenen Partei erfordert, auf den bei diesem Gesprächstyp aber weder mit Audio noch Video geantwortet wird. Dies nennt man Pull-Modus. In gleicher Weise kann man einen Push-Modus erzeugen, in dem der abgehende Anruf Audio und/oder Video sendet, bei dem aber weder ein Audio- noch ein Videorücklauf erforderlich ist. Diesen Modus kann man dazu verwenden, Inhalte gezielt an unbeaufsichtigte Terminals zu senden oder an Terminals, deren Benutzer in der Konferenz nur eine passive Rolle einnehmen.

[0297] Die Gesamtlautstärke der Lautsprecher **64**, des Hörers und des Kopfhörers kann man unabhängig einstellen. Der Lautsprecher lässt sich ein- und ausschalten. Das Abschalten der Lautsprecher schaltet auch die Mikrofone ab. Statusanzeigen stellen den Status des Lautsprechers und des Mikrofons dar.

[0298] Das Mikrofon lässt sich ausschalten und wieder einschalten. Statusanzeigen stellen den Status der Mikrofonausblendung dar.

[0299] Die Kamera lässt sich ausschalten und wieder einschalten. Statusanzeigen stellen den Status der Kameraausblendung dar.

[0300] Die Gesprächssteuerungen wirken nur auf das aktive Gespräch. Ein angenommenes Gespräch wird, falls es nicht aktiv ist, entweder durch das Berühren der Statusanzeige für das laufende Gespräch auf der Steuertafel oder sonstwo im Gesprächsanzeigebereich **54** aktiviert, jedoch mit Ausnahme der besonderen Steuerfunktionsbereiche für das laufende Gespräch. Dabei wird jedes weitere derzeit aktive Gespräch in den inaktiven Status versetzt. Das aktive Gespräch kann durch eine weitere Berührung des gleichen Bereichs inaktiviert werden. Eine Steuermöglichkeit zum Auflegen des aktiven Gesprächs ist vorhanden. In einem Konferenzgespräch werden alle Elemente des Gesprächsvorgangs gelöscht.

[0301] Damit die Konferenzsteuerung wirkt, muss ein Gespräch angenommen und aktiv sein. Das Berühren der Konferenzsteuerung fügt den derzeit aktiven Gesprächsvorgang mit dem folgenden aktivierten Gespräch zusammen. Die Konferenzsteuerung zeigt solange an, dass das Gespräch aktiv ist, bis entweder wieder darauf gedrückt wird, wodurch es inaktiv wird, oder ein anderer Gesprächsvorgang aktiviert wird. Nachdem alle Gespräche im nun aktiven Gespräch zum Konferenzgesprächsvorgang verbunden sind, wird das Gespräch zu einem einzigen Konferenzgespräch und die Anzeige für eine aktive Konferenzsteuerung erlischt. Nochmals zusammengefasst wählt die Konferenz das Gespräch, mit dem andere Gespräche verbunden werden, und sie wählt anschließend das Gespräch, das mit diesem Gespräch zu verbinden ist.

[0302] Die Vorgehensweise, mit der eine Partei aus einem Konferenzgespräch ausscheidet, besteht darin, den Hörer aufzulegen. Aus verschiedenen Gründen kann der Benutzer wünschen, jeden Teil eines Gesprächsvorgangs unabhängig steuern zu können. Dies kann man durch eine Möglichkeit zum Beenden der Konferenz erzielen. Beispielsweise erscheint durch das Berühren des Gesprächsvorgangs für mehr als drei Sekunden ein Untermenü, mit dem man die einzelnen Teilnehmer des Gesprächsvorgangs erkennen und sie für das Entfernen aus der Konferenz auswählen kann. Dieses Gespräch wird daraufhin aus der Konferenz herausgenommen und als eigenständiger Gesprächsvorgang geführt, auf den man alle normalen Steuerungshandlungen anwenden kann, insbesondere kann man das Gespräch beenden.

[0303] Die Übertragungsfunktion überträgt das aktive Gespräch. Berührt man die Übertragungssteuerung, so wird der integrierte Wahlbildschirm dargestellt und das aktive Gespräch in den Haltestatus versetzt. Es wird jedoch angezeigt, dass es in einen laufenden Gesprächsvorgang eingebunden ist. Die Übertragungssteuerung stellt es aktiv dar, bis ein zweites Mal darauf gedrückt wird – dies bricht die Übertragung ab – oder bis der Benutzer die Nummer

aussucht, an die er das Gespräch zu übertragen wünscht, und die Wahl Tasten drückt.

[0304] Nach dem Anstoßen des abgehenden Anrufs zeigt die Übertragungssteuerung eine Änderung des Status, wodurch ein Berühren der Steuerung eine "blinde" Übertragung bewirkt und der Gesprächsvorgang vom Bildschirm entfernt wird. Wahlweise kann der Benutzer warten, bis die angerufene Nummer antwortet. An dieser Stelle wird ein neuer Gesprächsvorgang erzeugt, der es dem Benutzer erlaubt, mit der angerufenen Partei zu sprechen. Die Übertragungssteuerung ändert ihren Status erneut und zeigt an, dass ein nochmaliges Drücken den Übertragungsvorgang abschließt und beide Gespräche beendet. Andernfalls wird gefordert, zurück zum Sprechen mit dem Gesprächsteilnehmer zu gehen, der übertragen wird, und den Übertragungsvorgang erneut zu beginnen, oder das Gespräch zu beenden. Die Übertragung ist der Hauptmechanismus, mit dem ein "Verwalter" ein Gespräch herstellt und es anschließend an den "Chef" übergibt. In diesem Fall ist es unerlässlich, dass es für den Verwalter nicht möglich ist, bei dem übertragenen Gespräch weiter "mitzuhören". Dies gilt insbesondere in einer Umgebung mit Sicherheitsanforderungen.

[0305] Man kann das aktive Gespräch in den HALTE-Status versetzen, indem man die HALTE-Steuerung berührt. Im HALTE-Status werden die abgehenden Video- und Audioströme unterbrochen und der entfernten Seite wird mitgeteilt, dass sie sich im HALTE-Status befindet. Die eingehenden Video- und Audioströme werden nicht mehr dargestellt. Der HALTE-Status wird in der Gesprächsstatusanzeige **54** im Steuerbalken dargestellt. Die Haltesteuerung zeigt an, dass ein Haltevorgang aktiv ist, wenn sich irgendein Gespräch im Haltestatus befindet. Befindet sich das aktive Gespräch im HALTE-Status, so beendet ein erneutes Drücken auf HALTEN den HALTE-Status und führt das Gespräch in den dargestellten Status zurück.

[0306] Auf der Hauptstuertafel befindet sich eine Steuerungsmöglichkeit, die den Startbildschirm aufruft und den Zugriff auf alle anderen Nicht-Gesprächs-Funktionen erlaubt. Es wird angezeigt, dass der Hauptbildschirm gewählt ist. Ein erneutes Drücken auf den Startbildschirm stellt die Anzeige der aktuellen Gespräche wieder her und bringt den Startbildschirm zum verschwinden. Für jede angenommene und dargestellte Partei in einem Gespräch werden eigene Steuerungsmöglichkeiten bereitgestellt und für jedes Gespräch angezeigt. Das Einstellen der Lautstärke des Audios von jedem einzelnen Benutzer ist erforderlich. Es ist möglich, das Audio oder Video eines jeden auf dem Bildschirm dargestellten Benutzers für sich zu unterdrücken. Es ist ein Statusanzeiger vorhanden, der anzeigt, ob die Audio- oder Video-Unterdrückung eingeschaltet ist.

[0307] Kann zu jedem beliebigen Zeitpunkt mehr als ein Gesprächsvorgang dargestellt werden, beispielsweise eine Konferenz mit zwei weiteren Personen und ein neues Gespräch mit einem anderen Benutzer, so ist es möglich, das Audio und/oder Video für einen kompletten Gesprächsvorgang zu unterdrücken. Beispielsweise kann man das Audio für die Konferenz mit den zwei Personen unterdrücken und inzwischen das zweite Gespräch führen.

[0308] Das Anfordern von Video auf einer reinen Audioverbindung, die Video unterstützen könnte, ist vorgesehen. Das Annehmen oder Zurückweisen einer Videoanforderung ist vorgesehen. Eine Videoverbindung wird hergestellt, falls der Verbindung zugestimmt wird. Ein Punkt auf der Einstellseite ermöglicht es dem Benutzer, Videoanforderungen stets anzunehmen oder stets abzulehnen.

[0309] Es ist möglich, die Trägerkanalparameter für jede Verbindung darzustellen, d. h. die eingehenden und abgehenden Codierraten für Video (falls vorhanden) und Audio. In einem Gespräch wirken die Steuerhandlungen nur auf das aktive Gespräch. Ein angenommenes Gespräch wird aktiviert, falls es inaktiv ist.

[0310] Es ist möglich, eine "Trägerkanal-Qualitätsüberwachung" für jeden Benutzer freizugeben. Diese Anzeige, die entfernt der Signalstärkenanzeige auf einem Mobiltelefon gleicht, zeigt beispielsweise 100 Prozent eines grünen Balkens, falls keine Fehler oder verlorenen Pakete auf den Audio- und Videokanälen auftreten, einen gelben Balken, falls die Verlustrate oder die Wartezeit eine vorbestimmte Rate übersteigt, und einen roten Balken, falls diese Werte eine höhere Grenze überschreiten. Die Mittelungszeit sollte kurz sein, beispielsweise 50 Millisekunden, da Fehler in diesem zeitlichen Rahmen das Video des Benutzers beeinträchtigen. Erhält beispielsweise der Empfänger Videoartefakte, sieht jedoch gleichzeitig, dass der Überwachungsbalken gelb oder rot wird, so weiß er, dass dies von einer Überlastung des Netzes **40** herrührt.

[0311] Das Anfordern einer Veränderung der Videocodierparameter, d. h. eine Erhöhung oder Verringerung der Codierrate, innerhalb eines Gesprächs ist vorgesehen. Das Annehmen oder Zurückweisen dieser Anforderung und ein Verfahren zum Verändern der abgehenden Videorate ist vorgesehen. Das Videophon **15** erzeugt eine einzige abgehende Videorate für alle Teilnehmer. Das Videophon kann unterschiedliche ankommende Raten in allen ankommenden Strömen annehmen.

[0312] Eine Anforderung für einen Zusatzweg und die Möglichkeit, diese Anforderung anzunehmen oder zurückzuweisen, ist bereitgestellt. Bei Annahme schaltet der Zusatzweg den Audiostrom von beiden Teilnehmern zu allen anderen Personen ab, wodurch

sie eine private Unterhaltung führen können. Sie können aber weiter alle Unterhaltungen hören und alle Teilnehmer sehen und von diesen gesehen werden. Die Fähigkeit, mit den Video- und Zusatzweg-Anforderungen Kurznachrichten in beide Richtungen senden, ist gegeben.

[0313] Unabhängig davon, ob das Gespräch ein eingehendes Gespräch oder ein abgehendes Gespräch ist, sollte der Bildschirmübergang zur Videodarstellung weich erfolgen. Das Audio kann dem Video vorausgehen. Das Video sollte nicht dargestellt werden, solange dieser Übergang noch nicht möglich ist. (D. h., es sollen keine springenden Bilder, halb formatierte Frames usw. im Übergang zum Video erscheinen.) Der Übergang zum Videobildschirm der Benutzeranzeige **54** sollte erst beginnen, wenn das Gespräch "steht", und nicht schon beim Aufbau des Gesprächs. Die Darstellung des Videos vom Benutzer sollte die Fläche des Bildschirms **54**, die der Benutzeranzeige **54** zugewiesen ist, bestmöglich ausnutzen. Eine Steuerung in der Darstellung **54** ermöglicht es, diese Darstellung **54** für einen Benutzer und einen Gesprächsvorgang in die Vollbildmodus-Darstellung **54** umzuwandeln. Eine Berührung irgendwo innerhalb der "Vollbilddarstellung" **54** führt zur Standarddarstellung zurück. Zusätzlich zu den bereits genannten Gesprächssteuermöglichkeiten sollte der Benutzername dargestellt werden. Die Darstellung **54** und der Gesprächsvorgang auf der Steuertafel müssen anzeigen, ob das Gespräch aktiv ist oder nicht, d. h. ob die allgemeinen Steuermöglichkeiten in einem Gespräch arbeiten oder nicht. Bei einem dargestellten Gesprächsvorgang erfolgt das Aktivieren oder Inaktivieren durch das Berühren des Gesprächsvorgangs oder irgendeiner anderen Stelle auf der Hauptanzeige **54** außerhalb der Steuerbereiche, die nur für das Gespräch gelten.

[0314] Der Übergang von einem Gesprächsvorgang mit zwei Parteien sollte sanft erfolgen und angestoßen werden, sobald das zweite Gespräch "steht". Die Darstellung **54** sollte die Darstellungsfläche **54**, die der Benutzeranzeige **54** zugewiesen ist, bestmöglich ausnutzen. Bei Bedarf kann man die Videos an jeder Kante beschneiden anstatt sie zu skalieren, damit sie in die verfügbare Fläche passen. Eine Vollbilddarstellung **54** für zwei oder mehr angezeigte Gespräche ist nicht erforderlich. Zusätzlich zu den genannten gesprächsinternen Steuermöglichkeiten sollte der Benutzername für jede Partei dargestellt werden. Es muss angezeigt werden, dass beide Parteien Teil eines einzigen Gesprächsvorgangs sind. Die Anzeige **54** und der Gesprächsvorgang auf der Steuertafel müssen anzeigen, ob das Gespräch aktiv oder inaktiv ist. Das ankommende Video kann fortschreitend beschnitten werden, damit es in die verfügbare Anzeigefläche Bildschirm **54** passt, wenn weitere Parteien in das Videogespräch aufgenommen werden.

[0315] Im Fall von zwei Gesprächsvorgängen mit jeweils einer einzigen Gesprächspartei liegen zwei getrennte Gespräche zu Einzelbenutzern vor, die beide dargestellt werden müssen. Die Bildschirmanzeige **54** und die Gesprächssteueranzeige machen deutlich, dass es sich um zwei getrennte und unabhängige Gespräche handelt, und sie zeigt auch das aktive Gespräch an, falls ein Gespräch aktiv ist. Wird eines der beiden Gespräche in den HALTE-Status versetzt, so wird das Gespräch nicht mehr dargestellt, und die Anzeige **54** kehrt zu einer einzigen Gesprächsdarstellung eines einzigen Gesprächsvorgangs zurück.

[0316] Zusätzlich zu den beschriebenen Darstellungsmöglichkeiten sollte der Benutzerbereich jede beliebige Kombination darstellen können, die unten folgt.

[0317] Vier Gesprächsvorgänge für jeweils eine einzige Gesprächspartei.

[0318] Drei Gesprächsvorgänge, wobei ein Gespräch ein Gespräch mit zwei Parteien sein kann, und die anderen Gesprächsvorgänge Gespräche mit Einzelparteien sind.

[0319] Zwei Gesprächsvorgänge, wobei ein Gespräch bis zu drei Teilnehmer haben kann oder zwei Gesprächsvorgänge Gespräche mit zwei Parteien sein können.

[0320] Für eine Darstellung **54** nach "CNN-Art" gleichen die Anforderungen den beschriebenen Einzelgesprächsvorgang mit einem Gespräch, und zwar einschließlich der Möglichkeit der Vollbilddarstellung **54**. Es ist auch möglich, das Gespräch nach "CNN-Art" in einer Bildschirmhälfte darzustellen und die andere Hälfte für eine oder zwei Benutzerdarstellungsflächen zu verwenden, letzteres für zwei unabhängige Gesprächsvorgänge oder einen einzigen Gesprächsvorgang mit zwei Parteien.

[0321] Die Fähigkeit, verschiedene Verschlüsselungsebenen für die Sprach- und Datenströme bereitzustellen, ist gegeben. Der Zugang zu Diagnose-, Prüf-, Mess- und Verwaltungseinrichtungen soll SMF (SMF = Simple Management Framework) verwenden. Anders ausgedrückt ist der Zugang zu allen Einrichtungen auf drei Wegen möglich, nämlich über SNMP (SNMP = Simple Network Management Protocol), über das Web und über eine intelligente Schnittstelle. Das Videophon **15** muss aus der Ferne verwaltet werden können, es darf für den täglichen Betrieb oder für Softwareaufrüstungen, die Fehler beseitigen, kein IT-Expertenwissen vor Ort erforderlich sein. Fehlerdiagnosen sind ebenfalls aus der Ferne möglich, und man kann feststellen, ob das Problem an der Hardware der Einheit liegt, an der Konfiguration der Einheit, der Software der Einheit, dem Netz **40** oder den Diensten des Netzes **40**. Für die Verwaltung

kann man einen IP-Anschluss voraussetzen, man muss sich jedoch auf eine Verbindung zum Videophon **15** mit relativ geringer Bandbreite einstellen.

[0322] Im Normalbetrieb soll das Videophon **15** beim Einschalten eine verkürzte Version des Tests für das Hardwaresystem **10** ausführen. Hat der Test ein negatives Ergebnis, so soll das Videophon **15** auf dem Bildschirm eine Hochlauffehler-Meldung anzeigen. Das Terminal kann in einen erweiterten Hardwarediagnosemodus versetzt werden. Dies kann durch den Anschluss einer Tastatur an den USP-Port erfolgen oder indem man beim Einschalten der Einheit auf die rechte obere Ecke des Berührbildschirms **74** drückt. Dieser Modus erlaubt den Zugriff auf das unterlagerte Betriebssystem **10** und weitere leistungsfähige Diagnosewerkzeuge, mit denen man feststellen kann, ob ein Hardwarefehler vorliegt oder nicht.

[0323] Man kann eine Reihe einfacher Tests aufnehmen, die der Benutzer ablaufen lassen kann, falls das Videophon **15** den Hochlauffest besteht, dem Benutzer jedoch nicht die korrekte Funktionalität bietet. Das Terminal bietet eine technische Schnittstelle im Zusammenhang mit einer lokalen Tastatur (und Maus), die die Diagnose bei Problemen mit der Einheit oder dem System **10** unterstützt. Dies liefert einen Zugriff zu den verschiedenen Diagnosen für Audio, Video usw.

[0324] Es ist möglich, aus der Ferne gesteuert neue Versionen der Terminalsoftware für das Videophon **15** sicher herunterzuladen. Sicher bedeutet, dass man zur vorhergehenden Version zurückkehren kann, falls in der heruntergeladenen Version Fehler auftreten, und zwar ohne lokale Eingriffe (d. h., ohne dass man eine CD installieren muss). Es ist möglich, über die Verwaltungsschnittstellen die Software-Versionsnummer der Software auf einem bestimmten Videophonterminal **15** zu lesen sowie die Hardware-Seriennummer der Einheit, die Baugruppen-Revisionsnummer und Seriennummer sowie die Baugruppen-Revisionsnummer der wesentlichen Unterbaugruppen zu lesen. Im Fall eines Absturzes des Systems **10** sollte das Videophon **15** Information speichern oder gespeichert haben, die das Feststellen der Absturzursache unterstützt. Diese Information muss online von einer entfernten Site für die Analyse abrufbar sein, nachdem das Videophon **15** erneut hochgelaufen ist.

[0325] Das Videophon **15** bewahrt ein fortlaufendes Protokoll aller Handlungen, Ereignisse und Statusänderungen nach dem Einschalten, und zwar innerhalb der Grenzen des Speicherplatzes, der diesem Vorgang zuweisbar ist. Es sollte möglich sein, zumindest einen Monat lang die Aktivitäten zu speichern, die dies wert sind. Diese Daten können in mehrere unterschiedliche Kategorien fallen, beispielsweise eine Si-

cherheitskategorie, die die Benutzerdaten enthält, beispielsweise dass die Nummern, die der Benutzer angerufen hat, nur von ihm freigegeben werden können. Allgemeinere Daten, beispielsweise die Anzahl der Gespräche, der Gesprächsstatus (d. h. die Anzahl der Gesprächsvorgänge und die Endpunkte je Vorgang), die Einstellungen des Decodierers **34** und des Codierers **36**, Fehlerberichte des Trägerkanals usw. stellen eine weniger sensitive Information dar. Es kann nützlich sein, wenn man jeden Tastendruck aufzeichnen kann, um die Diagnose eines Level-Problems des Systems **10** zu unterstützen und die Verkettung der Ereignisse nachzuvollziehen.

[0326] Das Videophon **15** kann den Austausch auf der Steuerflächenebene sowohl auf der IP-Ebene als auch auf der SIP-Ebene an ein entferntes Diagnose-terminal kopieren (dies ist gleichwertig zu einem Line-Monitor, der aus der Ferne an das Videophon-terminal **15** angeschlossen ist). Die Terminalverwaltung überwacht zahlreiche Parameter, beispielsweise die Qualität des Netzes **40**. Es muss möglich sein, Grenzwerte zu setzen und einen Alarm zu erzeugen, wenn diese Grenzwerte überschritten werden. Sowohl die ATM-Schnittstelle als auch die Ethernet-Schnittstelle weisen Standardmesswerte auf (beispielsweise ähnlich zu RMON (RMON = Remote Network Monitoring)), die für das Videophon **15** verfügbar sein sollten. Das Videophon **15** sollte in der Lage sein, diese Alarmsignale an ein oder mehrere Netzverwaltungssysteme zu senden.

[0327] Zum Zweck der Erläuterung ist die Erfindung in den obigen Ausführungsformen sehr ausführlich beschrieben worden. Die Einzelheiten dienen jedoch nur der Erläuterung. Fachleute können Änderungen vornehmen, ohne den Bereich der Erfindung zu verlassen, der in den folgenden Ansprüchen beschrieben ist.

Patentansprüche

1. Telekommunikationssystem (**10**), umfassend: ein Netz (**40**); und einen ersten Knoten (**80**), einen zweiten Knoten (**82**) und mindestens einen dritten Knoten (**84**), die miteinander über das Netz (**40**) kommunizieren, **dadurch gekennzeichnet**, dass der erste Knoten (**80**) dafür eingerichtet ist, einen ersten Videostrom einer Szene am ersten Knoten (**80**), einen zweiten Videostrom der Szene am ersten Knoten (**80**) und einen Audiostrom der Szene am ersten Knoten (**80**) an den zweiten Knoten (**82**) und den dritten Knoten (**84**) zu senden, und dass der zweite Knoten (**82**) und der dritte Knoten (**84**) dafür eingerichtet sind, den Audiostrom und entweder den ersten Videostrom oder den zweiten Videostrom wiederzugeben.

2. System (**10**) nach Anspruch 1, dadurch gekennzeichnet, dass der erste Videostrom eine Bildra-

te hat, die größer oder gleich 25 Bilder je Sekunde ist, und das der zweite Videostrom eine Bildrate hat, die kleiner als 25 Bilder je Sekunde ist.

3. System (**10**) nach Anspruch 1 und 2, dadurch gekennzeichnet, dass der zweite Knoten (**82**) und der dritte Knoten (**84**) einen Anzeigebildschirm (**54**) aufweisen, und dass der zweite und der dritte Knoten so gestaltet sind, dass der erste Videostrom als großes Bild auf einer Fläche dargestellt wird, die 20 oder mehr Prozent des Bildschirms (**54**) einnimmt, und dass der zweite Videostrom als kleines Bild auf einer Fläche dargestellt wird, die weniger als 20 Prozent des Bildschirms (**54**) einnimmt.

4. System (**10**) nach Anspruch 1 bis 3, gekennzeichnet durch einen vierten Knoten, der mit dem ersten Knoten (**80**), dem zweiten Knoten (**82**) und dem dritten Knoten (**84**) über das Netz (**40**) kommuniziert und dafür ausgelegt ist, einen Fernseh-Videostrom eines Fernsehkanals an den ersten Knoten (**80**), den zweiten Knoten (**82**) und den dritten Knoten (**84**) zu senden; und dadurch, dass der erste Knoten (**80**), der zweite Knoten (**82**) und der dritte Knoten (**84**) den Fernseh-Videostrom auf dem Bildschirm (**54**) zusammen mit dem ersten Videostrom darstellen können.

5. System (**10**) nach Anspruch 1 bis 4, dadurch gekennzeichnet, dass der zweite Knoten (**82**) so beschaffen ist, dass er einen ersten Videostrom einer Szene am zweiten Knoten (**82**), einen zweiten Videostrom der Szene am zweiten Knoten (**82**) und einen Audiostrom der Szene an den dritten Knoten (**84**) sendet; und dass der dritte Knoten (**84**) einen Anzeigebildschirm (**52**) umfasst, der dazu dient, das Bild zu kontrollieren, das auf dem Bildschirm (**54**) erscheint, und jeden ersten Videostrom von jedem Knoten neben den anderen Strömen auf dem Bildschirm (**54**) des dritten Knotens (**84**) wiederzugeben.

6. System (**10**) nach Anspruch 1 bis 5, dadurch gekennzeichnet, dass das Netz (**40**) ein Ethernet oder ein ATM-Netz (**40**) ist.

7. System (**10**) nach Anspruch 1 bis 6, das Vorrichtungen umfasst, die den ersten und den zweiten Videostrom und den Audiostrom vom ersten Knoten (**80**) und vom zweiten Knoten (**82**) über das Netz (**40**) für ein Videokonferenzgespräch senden, wobei für das Videokonferenzgespräch keine Konferenzbrücke oder MCU verwendet wird.

8. System (**10**) nach Anspruch 1 bis 7, dadurch gekennzeichnet, dass jeder Strom für den Videokonferenzgesprächs-Knoten ATM-Point zu Multipoint umfasst.

9. System (**10**) nach Anspruch 1 bis 8, dadurch gekennzeichnet, dass der dritte Knoten (**84**) so beschaffen ist, dass er vorab festlegt, ob der erste oder

der zweite Videostrom vom ersten Knoten (80) oder vom zweiten Knoten (82) dargestellt wird.

10. System (10) nach Anspruch 1 bis 9, dadurch gekennzeichnet, dass der dritte Knoten (84) so eingerichtet ist, dass er die Darstellung des ersten Videostroms vom ersten Knoten (80) oder vom zweiten Knoten (82) auswählt, falls ein Benutzer in der jeweiligen Szene am ersten Knoten (80) oder zweiten Knoten (82) gerade spricht, oder dass der dritte Knoten (84) vorbestimmt hat, den ersten Videostrom des ersten Knotens (80) oder des zweiten Knotens (82) darzustellen.

11. System (10) nach Anspruch 1 bis 10, dadurch gekennzeichnet, dass der erste Videostrom des ersten Knotens (80) und des zweiten Knotens (82) im Format MPEG-2 vorliegt, wenn er über das Netz (40) gesendet wird.

12. System (10) nach Anspruch 1 bis 11, dadurch gekennzeichnet, dass der erste Knoten (80) und der zweite Knoten (82) so eingerichtet sind, dass sie Continuous P dazu verwenden, den ersten Videostrom vom ersten Knoten (80) und vom zweiten Knoten (82) in das Format MPEG-2 umzusetzen.

13. System (10) nach Anspruch 1 bis 12, dadurch gekennzeichnet, dass der erste Knoten (80) und der zweite Knoten (82) so eingerichtet sind, dass sie die ersten Videostrome der Szene des ersten Knotens (80) bzw. des zweiten Knotens (82) beschneiden.

14. System (10) nach Anspruch 1 bis 13, dadurch gekennzeichnet, dass der erste Knoten (80) und der zweite Knoten (82) so eingerichtet sind, dass sie den ersten Videostrom ihrer jeweiligen Szenen dadurch beschneiden, dass sie einen Abschnitt des ersten Videostroms entfernen, der zu einem Ort der entsprechenden Szene gehört, an dem sich kein Benutzer befindet.

15. System (10) nach Anspruch 1 bis 14, dadurch gekennzeichnet, dass der erste Knoten (80) und der zweite Knoten (82) so eingerichtet sind, dass sie nur einen Audiostrom der Szene des ersten Knotens (80) bzw. des zweiten Knotens (82) senden, falls ein Rauschen über einem vorbestimmten Grenzwert auf der jeweiligen Szene liegt.

16. System (10) nach Anspruch 1 bis 15, dadurch gekennzeichnet, dass der erste Knoten (80) einen automatischen Anwesenheitssensor (76) umfasst, der feststellt, ob ein Benutzer in der Szene am ersten Knoten (80) vorhanden ist, und der eine Anwesenheitskennung (79) erzeugt, die angibt, ob sich der Benutzer am ersten Knoten (80) aufhält, und dass der erste Knoten (80) eine Vorrichtung enthält, die die Anwesenheitskennung (79) an den zweiten Knoten (82) und den dritten Knoten (84) sendet.

17. System (10) nach Anspruch 1 bis 16, dadurch gekennzeichnet, dass der erste Knoten (80) dafür eingerichtet ist, ein Warnsignal zu erzeugen, das jeden Benutzer in der Szene am ersten Knoten (80) daran erinnert, dass die Anwesenheitskennung (79) nach Ablauf einer vorbestimmten Zeitspanne erstellt wird.

18. System (10) nach Anspruch 1 bis 17, dadurch gekennzeichnet, dass der erste Knoten (80) eine Bildgebervorrichtung (30) umfasst, die ein Videobild der ersten Szene aufnimmt und den ersten Videostrom erzeugt.

19. System (10) nach Anspruch 1 bis 18, gekennzeichnet durch einen Codierer (36), der mit der Bildgebervorrichtung (30) kommuniziert und den ersten Videostrom ohne Bildpufferung in ein gewünschtes Format komprimiert und codiert.

20. System (10) nach Anspruch 1 bis 19, dadurch gekennzeichnet, dass der erste Knoten (80) ein Field Programmable Gate Array (38) enthält, das mit dem Codierer (36) kommuniziert und so ausgelegt ist, dass es den ersten Videostrom in Pakete unterteilt und auch den ersten Videostrom von der Bildgebervorrichtung (30) empfängt und den zweiten Videostrom des ersten Knotens (80) erzeugt und den zweiten Videostrom in Pakete unterteilt.

21. System (10) nach Anspruch 1 bis 20, dadurch gekennzeichnet, dass der erste Knoten (80) eine Netzschnittstelle (42) enthält, die mit dem Field Programmable Gate Array (38) und dem Netz (40) kommuniziert, und dass die Schnittstelle so eingerichtet ist, dass sie den ersten Videostrom des ersten Knotens (80) zum Netz (40) überträgt und den ersten Videostrom vom zweiten Knoten (82) empfängt und ihn an das Field Programmable Gate Array (38) sendet.

22. System (10) nach Anspruch 1 bis 21, dadurch gekennzeichnet, dass der erste Knoten (80) Mikrofonvorrichtungen (58, 60, 62) enthält, die Schall von der ersten Szene aufnehmen und den Audiostrom des ersten Knotens (80) erzeugen.

23. System (10) nach Anspruch 1 bis 22, dadurch gekennzeichnet, dass der erste Knoten (80) eine Lautsprechervorrichtung (64) enthält, die mit der Netzschnittstelle (42) kommuniziert und den Audiostrom vom zweiten Knoten (82) wiedergibt.

24. System (10) nach Anspruch 1 bis 23, dadurch gekennzeichnet, dass der erste Knoten (80) einen DSP (62) enthält, der dafür ausgelegt ist, dass er den Audiostrom des ersten Knotens (80) in Pakete zerlegt und den Audiostrom an das Field Programmable Gate Array (38) liefert, das dafür eingerichtet ist, den Audiostrom des ersten Knotens (80) an die Netzschnittstelle (42) zu übertragen, die dafür ausgelegt

ist, den Audiostrom des ersten Knotens (80) an das Netz (40) zu übertragen und den Audiostrom des zweiten Knotens (82) vom Field Programmable Gate Array (38) zu empfangen.

25. System (10) nach Anspruch 1 bis 24, dadurch gekennzeichnet, dass der erste Knoten (80) eine Audioschnittstelle (60) enthält, die dafür ausgelegt ist, den Audiostrom des ersten Knotens (80) von den Mikrofonvorrichtungen (58, 60, 62) zu empfangen, ihn zu digitalisieren und an den DSP (62) zu liefern, und den Audiostrom vom zweiten Knoten (82), den sie vom DSP (62) empfängt, in eine analoge Form umzusetzen, die die Lautsprechervorrichtung (64) wiedergibt.

26. System (10) nach Anspruch 1 bis 25, dadurch gekennzeichnet, dass die Netzschnittstelle (42) so beschaffen ist, dass sie Pakete des Audiostroms und Videostroms des ersten Knotens (80) mit einer Zeitangabe versieht, bevor sie an das Netz (40) gesendet werden, und dass sie die Pakete des Videostroms und des Audiostroms des zweiten Knotens (82), die der erste Knoten (80) empfängt, anhand der Zeitangaben ausrichtet, so dass beim Wiedergeben des Videostroms und des Audiostroms des zweiten Knotens (82) am ersten Knoten (80) der zugehörige Schall mit den Bild der Szene des zweiten Knotens (82) wiedergegeben wird.

27. System (10) nach Anspruch 1 bis 26, dadurch gekennzeichnet, dass der erste Knoten (80) einen Empfangsspeicher enthält, in dem der erste Videostrom vom zweiten Knoten (82) empfangen und gespeichert wird, und einen Hauptcontroller (50), der mit der Netzschnittstelle (42), dem Codierer (36), dem Field Programmable Gate Array (38) und dem DSP (62) verbunden ist und diese Geräte steuert, wobei der Hauptcontroller (50) so gestaltet ist, dass er der Netzschnittstelle (42) befiehlt, den ersten Videostrom des zweiten Knotens (82) zu wählen und ihn an den Empfangsspeicher zu senden, und der Hauptcontroller (50) so gestaltet ist, dass er den ersten Videostrom des zweiten Knotens (82) decodiert und expandiert, der im Empfangsspeicher abgelegt ist, und ihn an den Anzeigecontroller (52) sendet.

28. System (10) nach Anspruch 1 bis 27, dadurch gekennzeichnet, dass der erste Knoten (80) einen LCD-Controller (72) enthält, der an den Anzeigecontroller (52) angeschlossen ist, und der Anzeigebildschirm (54) einen Flachbildschirm enthält, der mit dem LCD-Controller (72) verbunden ist, wobei der LCD-Controller (72) so gestaltet ist, dass er den ersten Videostrom des zweiten Knotens (82) vom Anzeigecontroller (52) empfängt und den ersten Videostrom des zweiten Knotens (82) für die Darstellung auf dem Flachbildschirm aufbereitet.

29. System (10) nach Anspruch 1 bis 28, gekenn-

zeichnet durch einen Berührbildschirm (74), auf dem Tasten mit zugeordneten Funktionen erscheinen, und durch einen Berührbildschirm-Controller, der mit dem Berührbildschirm (74) und dem Hauptcontroller (50) verbunden ist und Informationen bezüglich der Tasten an den Hauptcontroller (50) überträgt, die ein Benutzer auf dem Berührbildschirm (74) berührt hat.

30. System (10) nach Anspruch 1 bis 29, dadurch gekennzeichnet, dass der erste Knoten (80) einen Decodierer (34) enthält, der den ersten Videostrom von der Bildgebervorrichtung der Szene des ersten Knotens (80) in digitale Form umsetzt und ihn an den Codierer (36) und das Field Programmable Gate Array (38) liefert, dass der Decodierer (34) mit dem Field Programmable Gate Array (38) und dem Codierer (36) verbunden ist, und dass der Decodierer (34) auch den Fernseh-Videostrom und andere analoge Videoströme empfängt und sie in digitale Form umsetzt.

31. System (10) nach Anspruch 1 bis 30, dadurch gekennzeichnet, dass die Bildgebervorrichtung eine analoge Videokamera (32) enthält, die mit dem Decodierer (34) kommuniziert, eine digitale Videokamera (47), die mit dem Codierer (36) und dem Field Programmable Gate Array (38) kommuniziert, oder eine Fire-Wire-Kamera (48), die mit dem Field Programmable Gate Array (38) kommuniziert, wobei das Field Programmable Gate Array (38) jeglichen Videostrom, den es von der Fire-Wire-Kamera (48) empfängt, an den Codierer (36) liefert.

32. System (10) nach Anspruch 1 bis 31, dadurch gekennzeichnet, dass der DSP (62) so ausgelegt ist, dass er eine Stereoecho-Auslöschung für den Audiostrom der Szene des ersten Knotens (80) liefert, und der Codierer (36) so ausgelegt ist, dass er für den ersten Videostrom des ersten Knotens (80) eine Auflösung von mindestens 720 × 640 Pixel bereitstellt, und die Mikrofonvorrichtungen und die Lautsprechervorrichtung im Vollduplexbetrieb arbeiten.

33. System (10) nach Anspruch 1 bis 32, dadurch gekennzeichnet, dass der DSP (62) so ausgelegt ist, dass er die Mikrofonvorrichtungen (58, 60, 62) hinsichtlich des Rauschpegels überwacht, den die Lautsprechervorrichtung (64) erzeugt, und die Lautsprechervorrichtung so einstellt, dass ein gewünschter Rauschpegel beibehalten wird.

34. System (10) nach Anspruch 1 bis 33, dadurch gekennzeichnet, dass der Hauptcontroller (50) so ausgelegt ist, dass er einen Befehl des Benutzers erkennt und automatisch einen weiteren Benutzer zu einer Videokonferenz anruft, und dass er den Befehl dazu verwendet, den anderen Benutzern anzukündigen, dass der weitere Benutzer für die Videokonferenz angefordert wird.

35. System (10) nach Anspruch 1 bis 34, dadurch gekennzeichnet, dass der erste Knoten (80) eine Videomail besitzt, falls der erste Knoten (80) einen Videoanruf nicht annehmen kann.

36. System (10) nach Anspruch 1 bis 35, gekennzeichnet durch einen Server (66), der so gestaltet ist, dass er mit dem Netz (40) und dem ersten Knoten (80) kommuniziert, dass er den Videoanruf für den ersten Knoten (80) empfängt, falls der erste Knoten (80) den Videoanruf nicht empfängt, dass er eine Videonachricht speichert, die zu dem Videoanruf gehört, den der erste Knoten (80) nicht empfängt, und dass er eine Videoanruf-Wartenachricht an den ersten Knoten (80) sendet, die angibt, dass eine Videonachricht darauf wartet, dem ersten Knoten (80) gezeigt zu werden.

37. System (10) nach Anspruch 1 bis 36, dadurch gekennzeichnet, dass der erste Knoten (80) zahlreiche Anzeigebildschirme (54) aufweist, die mit dem Anzeigecontroller (52) verbunden sind, und dass die Bilder der ersten Videoströme von unterschiedlichen Knoten auf den Bildschirmen (54) nebeneinander so dargestellt werden, als wären die zahlreichen Bildschirme (54) ein fortlaufender Bildschirm (54).

38. System (10) nach Anspruch 1 bis 37, dadurch gekennzeichnet, dass der erste Knoten (80) eine zweite Bildgebervorrichtung enthält, die den ersten Videostrom des ersten Knotens (80) erzeugt, der die Szene am ersten Knoten (80) unter einem anderen Blickwinkel sieht als die Bildgebervorrichtung.

39. System (10) nach Anspruch 1 bis 38, dadurch gekennzeichnet, dass der Hauptcontroller (50) so gestaltet ist, dass er den zweiten Videostrom des ersten Knotens (80) vom Field Programmable Gate Array (38) an den Anzeigecontroller (52) sendet, damit er auf dem Bildschirm (54) dargestellt wird und der Benutzer des ersten Knotens (80) die Szene am ersten Knoten (80) auf dem Anzeigebildschirm (54) sehen kann.

40. Videophon (15), umfassend:
eine Bildgebervorrichtung (30), die ein Videobild einer Szene aufnimmt und gekennzeichnet ist durch Vorrichtungen (36, 38, 42), die einen ersten Videostrom der Szene und einen zweiten Videostrom der Szene aus dem Videobild erzeugen und den ersten Videostrom und den zweiten Videostrom gleichzeitig an ein Netz (40) senden, wobei die Erzeugungsvorrichtungen (36, 38, 42) mit der Bildgebervorrichtung kommunizieren.

41. Verfahren für einen Videoanruf, umfassend den Schritt:
Aufnehmen eines Videobilds einer Szene, gekennzeichnet durch die Schritte:
Erzeugen eines ersten Videostroms der Szene und

eines zweiten Videostroms der Szene aus dem Videobild; und
gleichzeitiges Senden des ersten Videostroms und des zweiten Videostroms an das Netz (40).

42. System (10) nach Anspruch 1 bis 39, dadurch gekennzeichnet, dass der erste Videostrom eine Bitrate hat, die größer oder gleich 1,5 Mbit je Sekunde ist, und dass der zweite Videostrom eine Bitrate von weniger als 1,5 Mbit je Sekunde hat.

Es folgen 10 Blatt Zeichnungen

Anhängende Zeichnungen

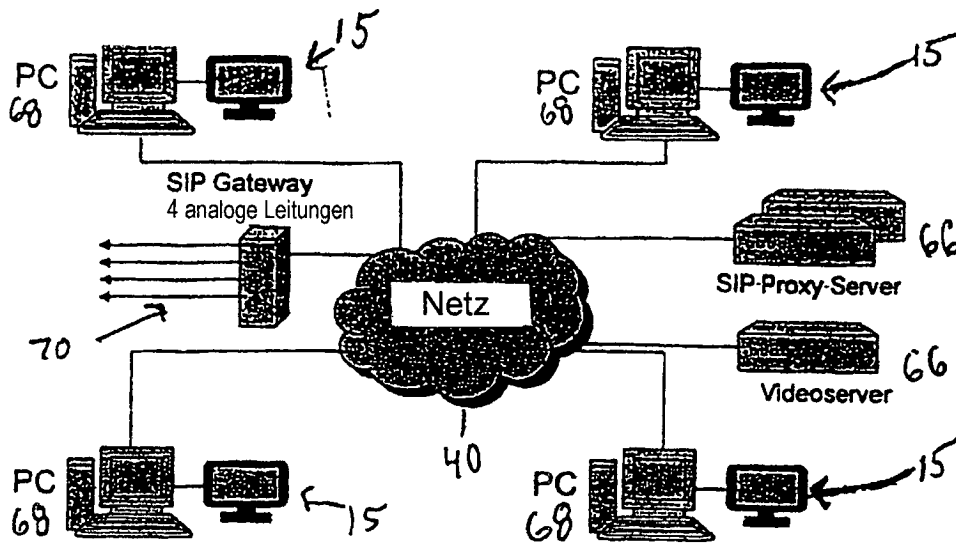


Fig 1

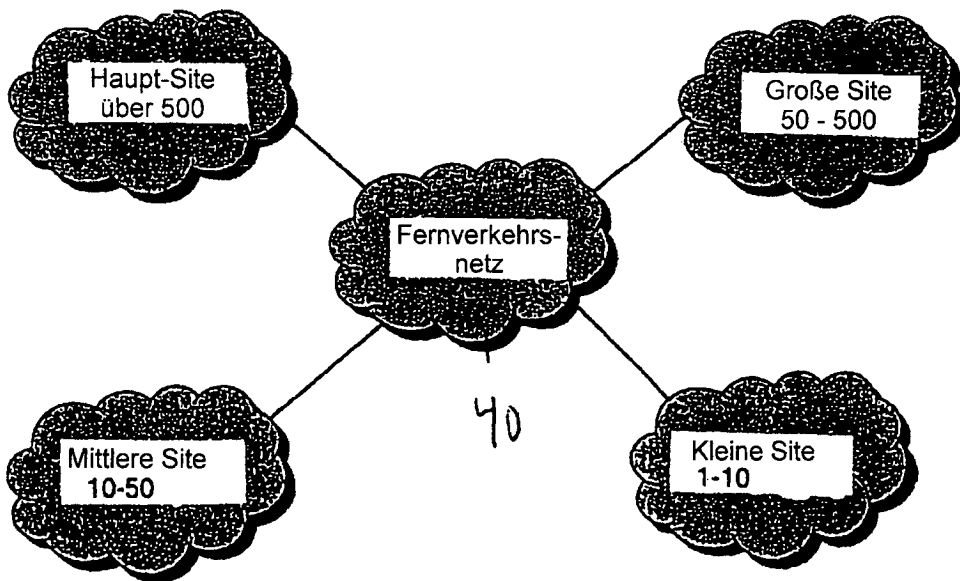


Fig 2

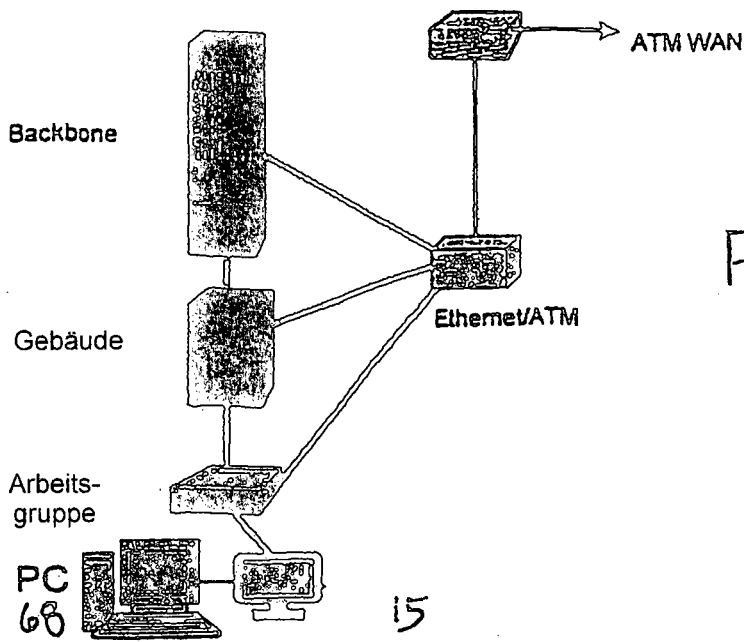


FIG 3

FIG 4

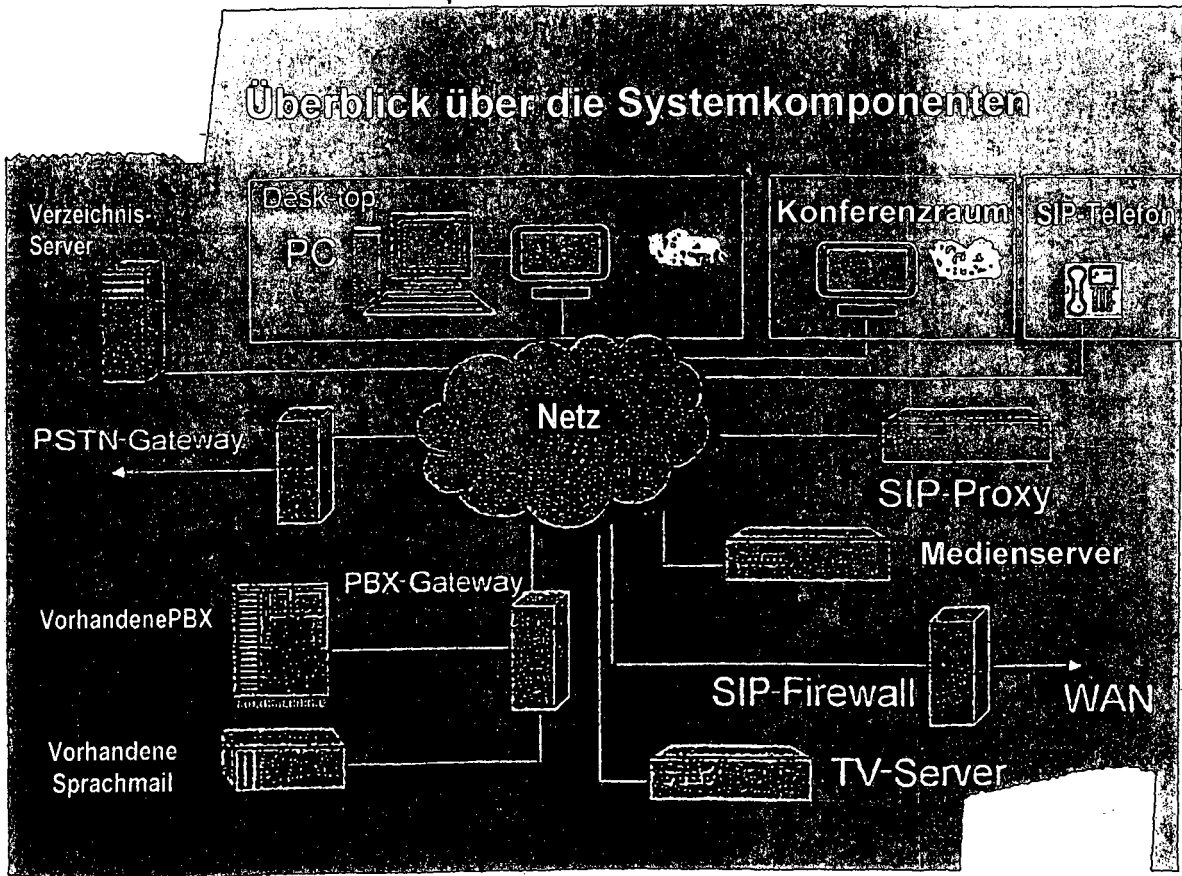


FIG 5a

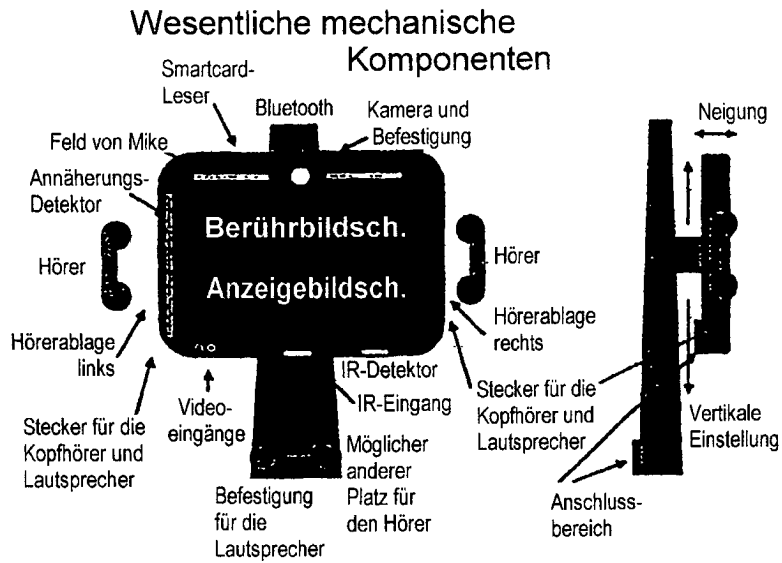


FIG 5b

Anschlussplatte

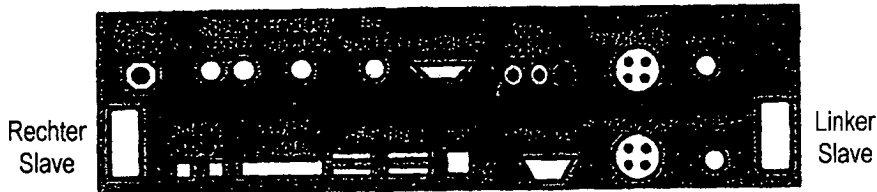


FIG 6

Anordnung mit mehreren Bildschirmen

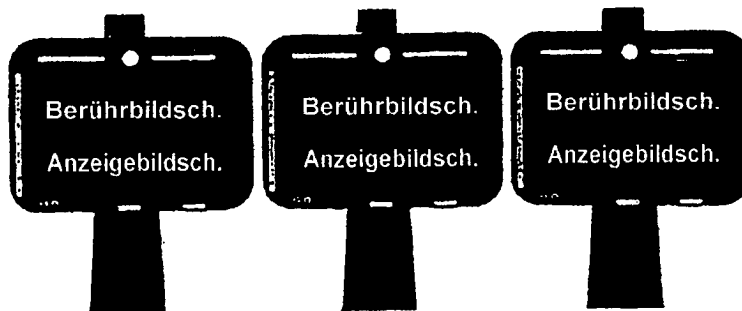
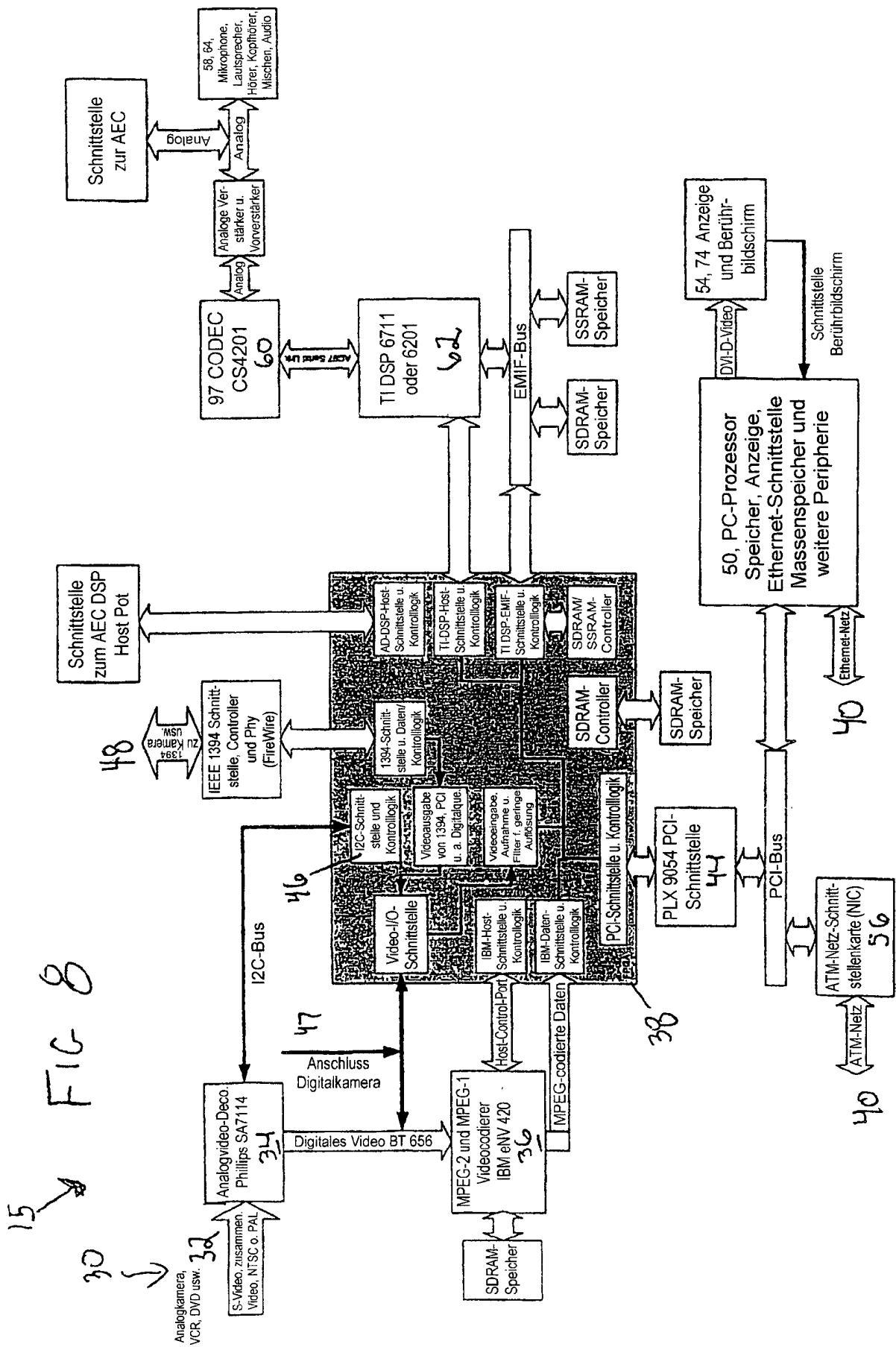


FIG 7



15 → FIG 8

Systemarchitektur

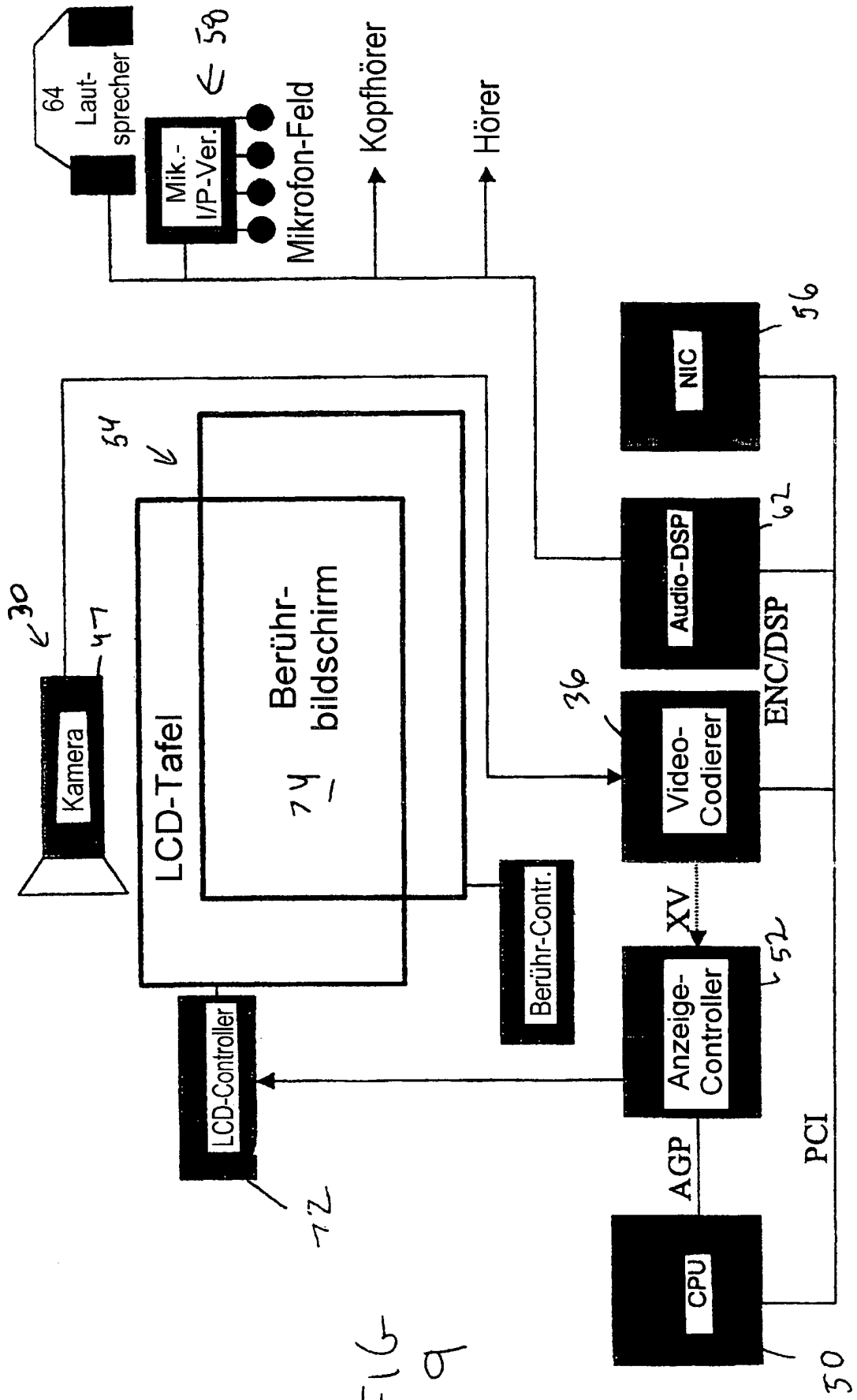


FIG 9

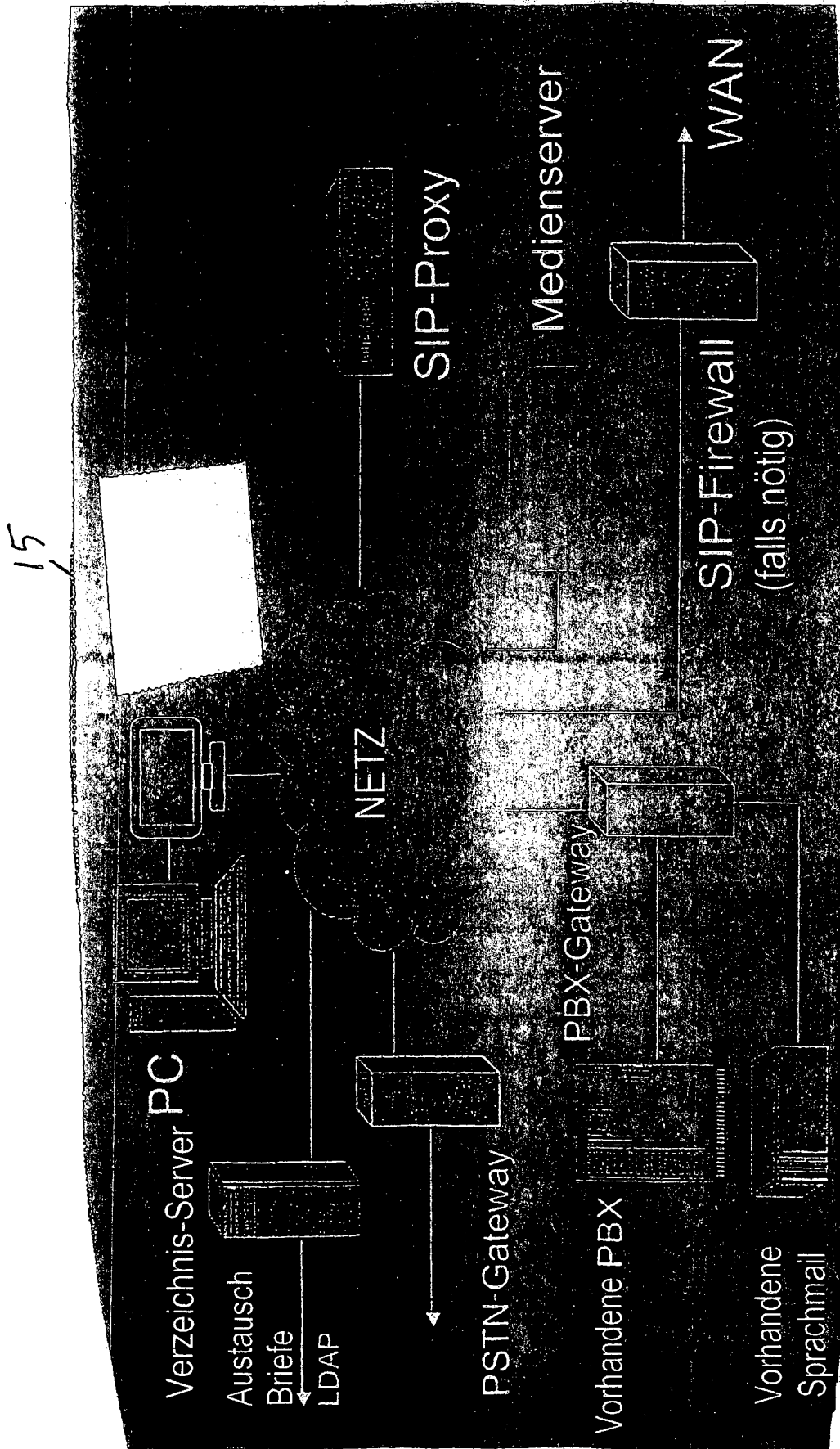


FIG. 10

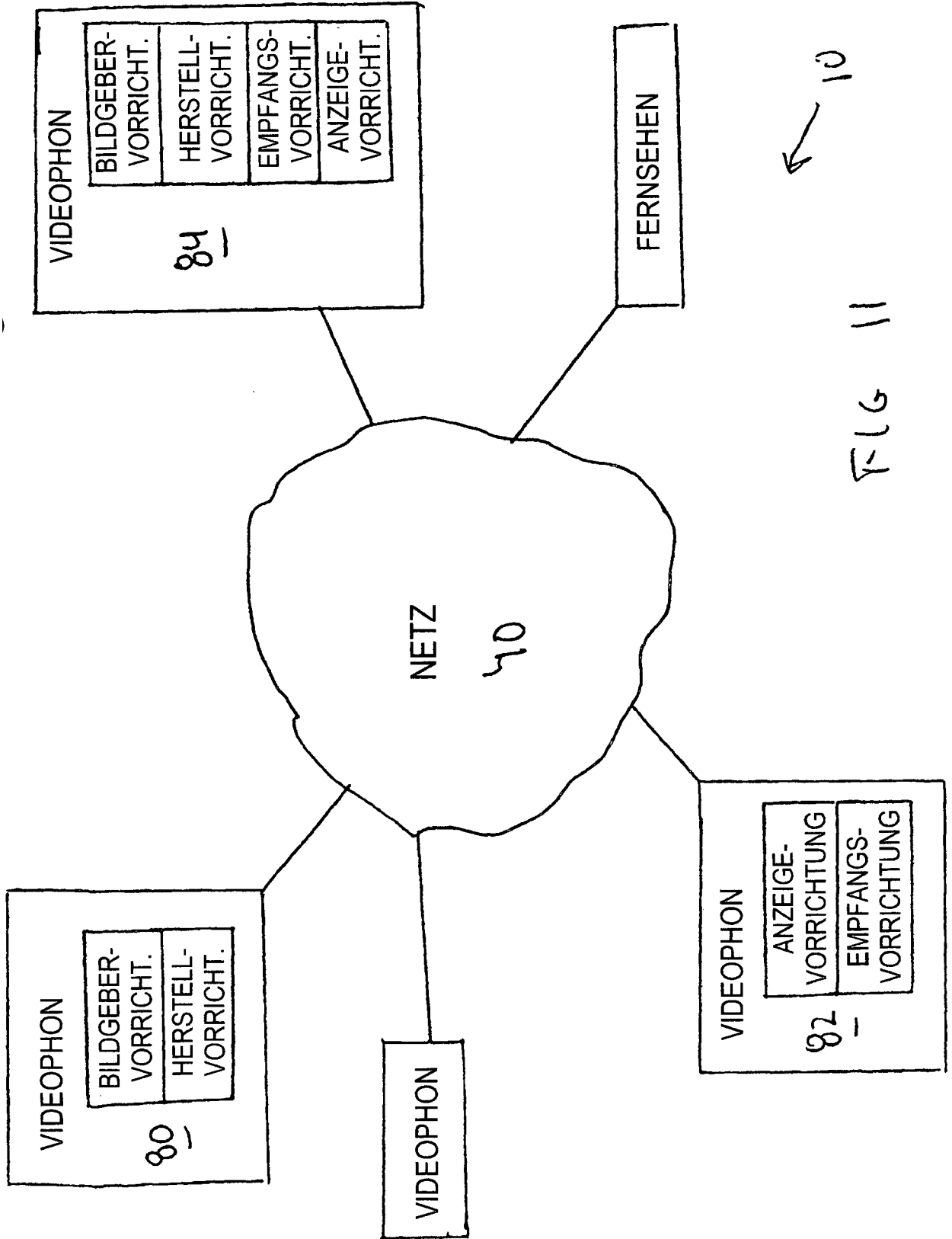


FIG 12

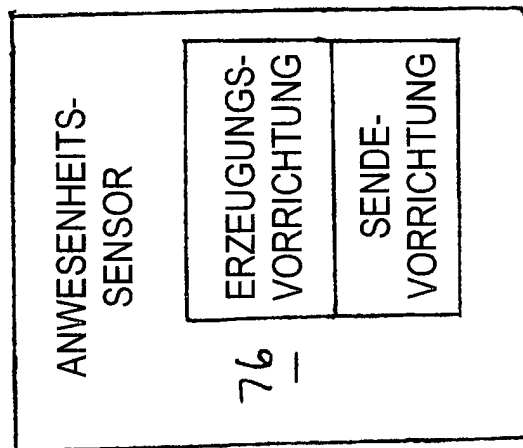
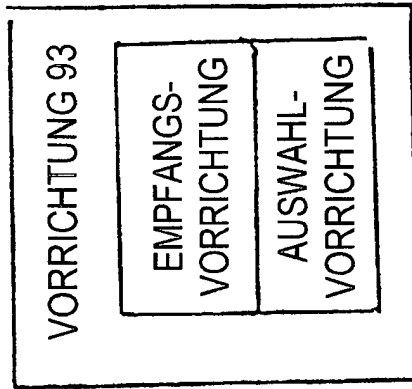
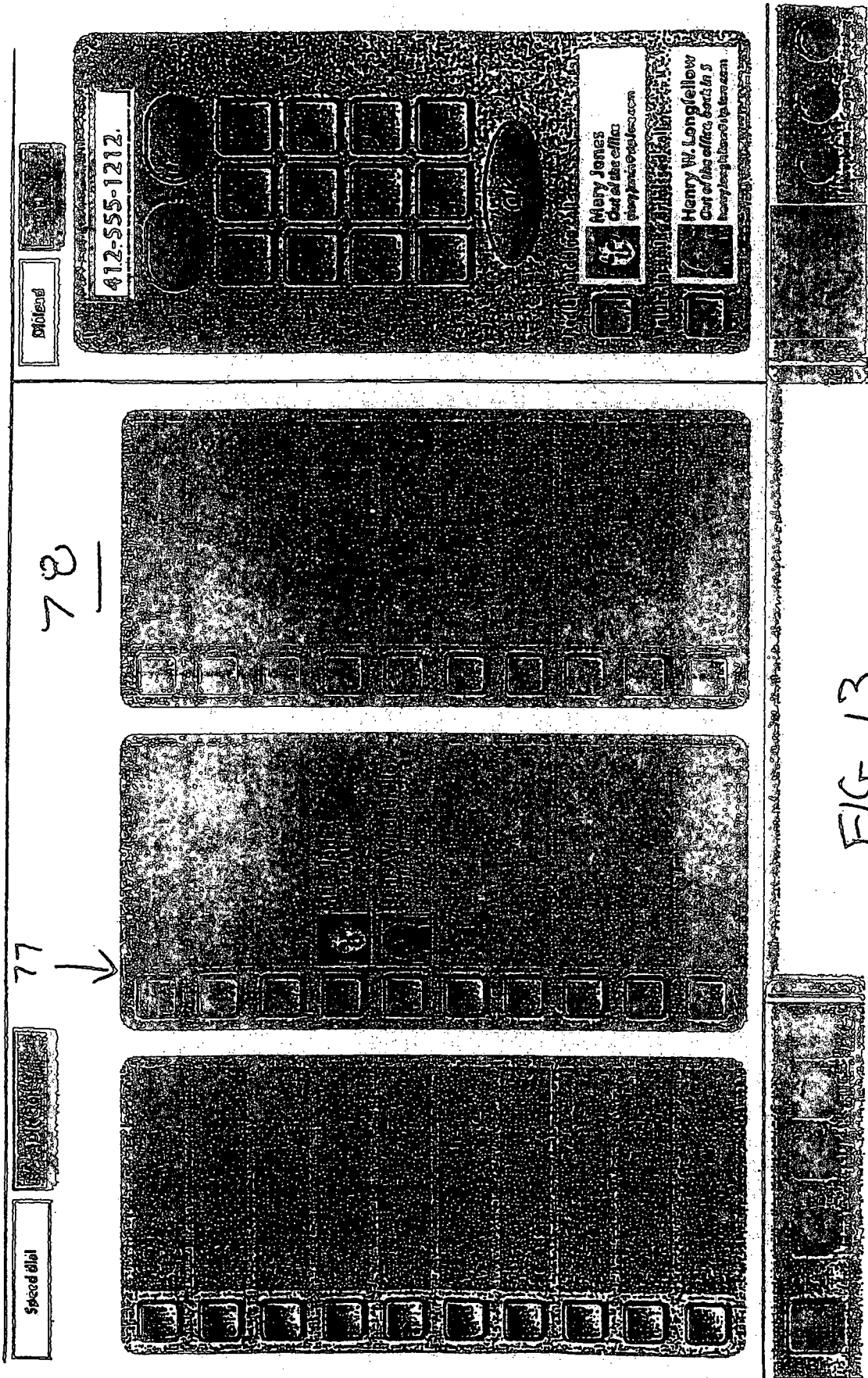


FIG 14





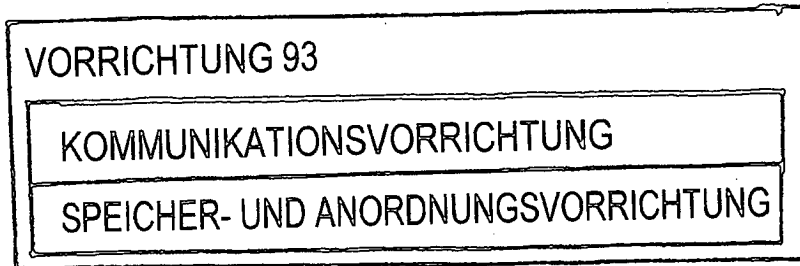


FIG 17

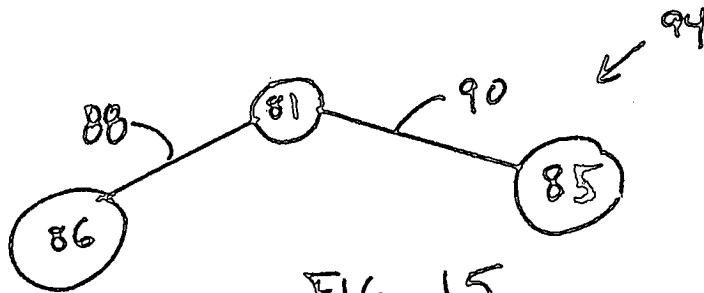


FIG 15

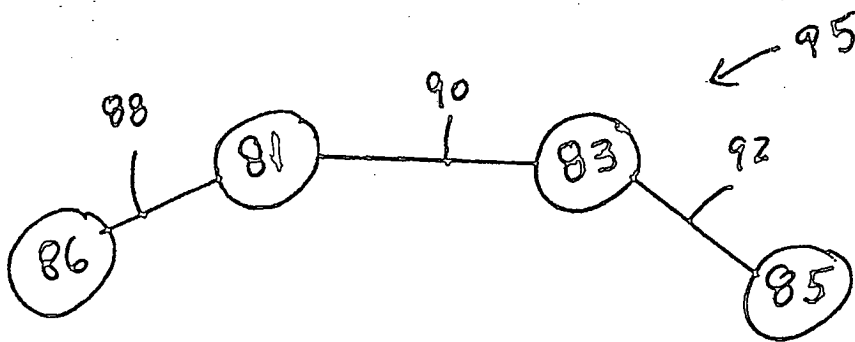


FIG 16