

Videokonferenzen auf PC-Basis

Einige technische und wirtschaftliche Kriterien.

Günther Zandra

Verständlich, daß viele Geschäftsleute Aufwand und Zeit für ihre Vertriebsgespräche gerne reduzieren möchten. Der Gedanken, ein tête-à-tête mit dem Kontaktpartner über Bildschirm und Telefon abzuhalten, ist seit langem verlockend, scheiterte jedoch viele Jahre an Systemkosten in der Größenordnung von einer halben Million Schilling und der Notwendigkeit, Breitbandkanäle - zu teuren Gebühren - zur Verfügung zu haben. Diese Anfangsschwierigkeiten sind bereits Geschichte. Die Systempreise sind inzwischen stark abgesunken und mit dem Schmalband-ISDN ist - zumindest für eingeschränkte Bewegtbildübertragungsqualität - das Kanalproblem gelöst. In den USA gibt es bereits ein halbes Dutzend Lieferquellen für Videokonferenzen auf PC-Basis. Die schnell arbeitenden CPUs und raffinierte Datenkompression sowie die Preissenkungen der ICs haben zu Systemen geführt, die für ein paar Tausend Dollar erhältlich und unter Ausnutzung der beiden ISDN-Basiskanäle von je 64 kbit/s betreibbar sind.

Die Bildprobleme

Die relativ sehr niedrige Übertragungsdatenrate führt zu einem Eindruck „zerhackter“ Bilder, weil die Anzahl der übertragenen Videorahmen pro Sekunde (frames per second, fps) weit unter der üblichen Rate des Fernsehens liegt. Dies stört nicht bei Gesichtsbildern, wohl aber bei jähen Gesten mit der Hand. Ein anderes Problem ist die Erzielung des Lippen-synchronismus. Man kann es umgehen, wenn man das Bild als „Stummfilm“ laufen läßt und das Gespräch gleichzeitig telefonisch führt (was allerdings wie eine schlecht synchronisierter Film wirkt). Ein „whiteboarding“, grafische Zeichen auf weißem Hintergrund, oder auch anteilige Dokumentenübertragung können das Realzeitvideo zum Stillstand bringen, wenn die Daten über denselben Kanal wie das Bewegtbild oder den Ton laufen. Gerade für PC-basierende Videokonferenz ist der Austausch von Dateiinhalten ein wichtiges Mittel der Werbung für derartige Anlagen.

In der Praxis muß man unterscheiden, ob eine Anlage nur für ein LAN oder für Fernübertragung bestimmt ist. Für LANs gibt es spezielle „Paketierungen“ der Bild- und Tonsignale, die aber für WAN-Übertragungen, eben für ISDN vor allem, ungeeignet sind. Für den WAN-Betrieb besteht auch die Möglichkeit, vermittelte 64-kbit/s-Kanäle zu verwenden. Das öffentliche Telefonnetz ist aber ungeeignet, weil die schnellsten, praktisch verwendbaren Modems mit höchstens 28,8 kbit/s betrieben werden können. Das ist zu wenig an Bandbreite. Aber auch mit 64 kbit/s ist bisher kein wirklich zufriedenstellendes Arbeiten zustande gekommen. Man bildet daher aus beiden B-Kanälen ein Aggregat von 128 kbit/s. Einige Lieferquellen wenden bereits die Kompressionsnorm H.320 der ITU an, die recht gute Resultate liefert.

Komponenten der Videokonferenz

Man unterscheidet vier grundlegende Elemente von Videokonferenzprodukten auf PC-Basis: Die **Input/Output-Einrichtungen**, die **Systemsoftware**, einen **Video-Codec** und einen **ISDN-Terminal-Adapter**.

Input/Output-Einrichtungen

Die Eingangs-Ausgangs-Geräte enthalten eine Kamera, ein Mikrofon und die Tongeräte, z.B. Lautsprecher, Kopfhörer oder ein Ohrgerät. Manche Firmen verwenden auch gewöhnliche Telefone für den Tonoutput. Die Systemsoftware ist für das Betriebssystem verantwortlich. Sie initiiert Konferenzen und enthält gewöhnlich auch einen Beitrag für Datei-Übermittlung und die Konferenz begleitendes anteiliges Abbilden von Dokumenten, die als Konferenzunterlage dienen.

Codec

Der Codec ist der Mechanismus, der den komprimierten Ablauf (mit Kompression und Dekompression der Video- und Audio-Daten) ermöglicht. Ein Codec kann in Hardware implementiert werden, aber auch in Software oder in einem Gemisch beider Arten. Reine SW-Lösungen sind selten. SW-Codex sind zwar billiger als HW-Ausführungen, erfordern aber schnellere, leistungsfähigere Plattformen (zumindest 66-MHz-486er-PCs).

ISDN-Terminal-Adapter

Die vierte Komponente, ein ISDN-Terminal-Adapter, bildet den Auslaß zum ISDN-WAN. Hier ist strikt auf die Anpassung an die vom Netzbetreiber geforderte Spezifikation zu achten. In den USA kann man bei Nichterfüllung dieser Bedingung durch die Lieferquelle um ca. 250 \$ eine NT-1-Einheit erwerben.

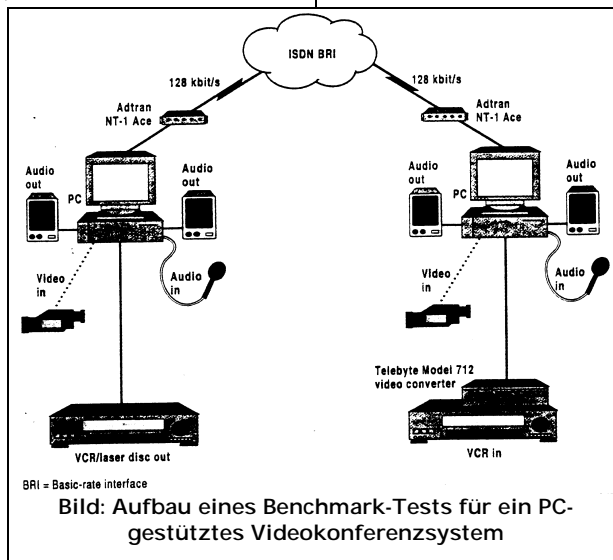
Tests an einem Videokonferenzsystem

Von vornherein muß man beim Anpacken des Testproblems darüber im klaren sein, daß ein guter Teil einer Bewertung subjektiver Natur sein muß, vor allem, was die Gesamtqualität der Video- und Tonwiedergabe betrifft. Besonders schwierig ist ein Prüfablauf, wenn sein Resultat als „benchmark“-Test gewertet werden soll. Es gibt aber, wie eine amerikanische Testorganisation festgehalten hat, zumindest zwei Schlüsseltests, die quantitative Bewertungsdaten liefern.

Der erste derartige Test besteht in der Zählung der Videorahmen pro Sekunde (frames per second, fps), die das System generiert. Der zweite erfaßt die Latenzzeitunterschiede zwischen Bild- und Tonübertragung.

Hinsichtlich qualitativer Beurteilung gilt für den Ton:

- seine Integrität,
- die Stetigkeit der Lautstärke,
- Verzerrungen plus Rauschen und vierdens



- die „Klarheit“.

Auch für die Videobewertung gibt es vier Qualitätscharakteristika:

- Bildfarbe,
- Kontrast,
- Helligkeit und
- Bildschärfe.

Eine amerikanische Testreihe hat zur Objektivierung der Untersuchung statt des Kamerabildes und Mikrophontons eine Videoaufzeichnung eines Films verwendet, der nur mittlere Bewegtheit aufwies.

Wenn man PC-gestützte Videokonferenzen betreibt, versteht es sich von selbst, daß auch „Dokumentaustausch“, allgemeiner ausgedrückt, wechselseitige Teilnahme an Applikationen, geprüft werden muß. Das wirft für die Anwender eine Reihe von Problemen auf, z.B. die Frage, ob die Hintergrundbeleuchtung für die komplette „Video-Session“ ausreicht, ob Hintergrundbewegung die Datenübertragung verschlechtert, oder die Frage, wieviel Bewegung dürfen Benutzer machen, ohne die Qualität zu beeinflussen. Normalerweise sollte das System unter typischer Bürobeleuchtung funktionieren (die bei PC-Betrieb relativ düster ist). Natürlich sollten Gesichtsbewegungen und Gesten von Teilnehmern gut übertragen werden.

Das Bild zeigt das von einem neutralen amerikanischen Labor verwendete Testsystem, mit dem Benchmark-Vergleich zwischen 5 US-Firmen, die sich diesem Wettbewerb stellten. Dabei wurde jeweils die von der

Lieferquelle verlangte PC-Plattform verwendet. Der für das Basisrate-ISDN erforderliche Network Terminal Adapter NT-1 war ein Produkt der US-Firma Adtran Inc. Für den Aufzeichnungs-Input wurden sowohl ein Videorecorder als auch ein Laser-Disc-Player verwendet. Jeder Videorahmen des Testbandes und der Bildplatte wurde mit einer Zeitmarke und einer Rahmennummer codiert. Dieses codierte Video wurde im Videoinput des Sendecodecs eingespeist, wobei eine konstante Aufzeichnungsrate von 30 fps verwendet wurde. Die Prüflinge hatten entweder bereits eine NTSC-Video-Inputbuchse oder es mußte ein Converter („VGA-to-NTSC physical converter“ von Telebyte Technology Inc.) zwischengeschaltet werden. Dann wurde eine Musterperiode von 90 gesendeten Rahmen aufgenommen und jede fehlende Rahmennummer als „dropped“ registriert. Als weitere Verfeinerung wurde zwecks Erfassung des Einflusses beim „application sharing“ die übertragene Frame Rate bei gleichzeitigem Transfer einer Datei von 1 Mbyte Umfang gemessen. Zur Messung der Audio- und Video-Latenzzeit wurde ein elektronisches Metronom verwendet. Zur Audiomessung wurden hörbare „ticks“ über den Audioinput der sendenden Station und das Netz zur Empfangsstation übertragen. Dann wurde der Metronomtakt so verstellt, bis das hörbare Metronomsignal in Gleichtakt zum hörbaren Signal im Ausgang des Empfängers geriet. An diesem Punkt war der Metronom-Output exakt um eine Periode im Vorlauf zum Signal der Empfangsstation. Dann wurden die „Metronom-Beats“, z.B. 104 pro Minute, durch 60 dividiert, was die Latenzzeit in Sekunden, also in diesem Beispiel 0,58 Sekunden, ergab.

Zur Berechnung der Videolatenzzeit wurde ähnlich vorgegangen, und zwar wurden vom Metronom generierte Lichtimpulse verwendet. Nachdem nun beide Latenzzeiten ermittelt waren, konnte ihre Differenz bestimmt werden. Aus ihr resultiert die bekannte Lippensynchronisation.

Für die subjektive Qualitätsbestimmung wurde bei jedem Test die gleiche (gleich gekleidete) Person und ein festgelegter Text verwendet. Beurteilt wurde nach 10 Punkten, jeder Test auch mit einer vorbestimmten Gewichtung. Hauptgewicht lag neben der Tonqualität auf dem Auflösungsvermögen, der Güte des Bewegtbildes und der Genauigkeit der Farbproduktion. Störend bei diesem Test war, daß die Systeme für Bild verschiedene Fenstergrößen aufwiesen. Die Lieferfirmen mußten daher ihre Fenstergröße auf etwa 1/4 des Bildschirms einstellen, um einen Vergleich zu ermöglichen.

Kritik am derzeitigen Stand der Technik

Die 5 geprüften Fabrikate stammten durchwegs von weltbekannten Firmen, z.B. von Pictoretel, dem Sieger des Wettbewerbs, von Intel und Northern Telecom. Daher muß das sachliche Bewertungsergebnis als Kritik am gegenwärtigen Stand der Technik aufgefaßt werden.

Die Prüfer waren zunächst über die niedrigen fps-Werte enttäuscht, weil sie ein Optimum in der Nähe von 24 fps (die TV-Bildrate in den USA ist bekanntlich 30 fps) erwartet hatten. Kein einziges der getesteten Systeme erreichte wenigstens 15 fps. Die deutlich darunter liegenden Werte ergaben natürlich die erwarteten „choppy“ Bilder. Die fps-Zahlen schwankten zwischen 9 (beim Pictoretel-Gerät PCS 50) und 12,7 einer anderen Firma. Dabei war dieser Unterschied quer über alle Testarten kaum subjektiv bemerkbar.

Alle geprüften Systeme funktionierten annehmbar, wenn es sich um Übertragung des Kopfes eines Teilnehmers handelte. Plötzliche Handgesten erschienen irgendwie langsam und ruckweise. Die Tests bestätigten auch den bekannten Kunstgriff in den Codecs, zur Einsparung an Bandbreite nur die Änderungen am Bildschirm zu übertragen. Das führt zu dem scheinbar paradoxen Ergebnis, daß bei einem bewegungslosen Teilnehmer mehr fps als bei Bewegung übertragen werden. Auch die Anzahl der Farben in einem Bild beeinflußt die fps-Werte. Weniger Farbtöne bedeuten weniger „updates“ für den Codec und somit mehr fps. Als Phänomen wurde auch beobachtet, daß die Codecs dann, wenn sich die Farben von Pixel zu Pixel wenig ändern, zur Pixelgruppierung tendieren. Man nennt diesen Effekt „tiling“ (von tile=Ziegel). Er trägt zusätzlich zur Bandbreitenreduktion bei. Deutlich zeigte sich auch, daß die Produkte höhere fps-Werte liefern, wenn der Teilnehmer vor einem stabilen, monocoloren Hintergrund positioniert ist.

Schwierigkeiten gab es auch bei der VCR-Filmclip-Übertragung. Die Beibehaltung der Sync-Signale war (infolge Banddehnungen und Leistungsschwankungen) weit schwieriger als beim Disc Player. Mehrfach kamen die Codecs außer Tritt. Um dies zu vermeiden, mußte man einen VCR mit Zeitbasiskorrektur verwenden.

Die an den Prüflingen gemessenen Latenzzeiten, die von der Art der Kompression und Dekompression in den Codecs abhängen, ergaben beträchtliche Unterschiede. Dabei ist gerade die Latenzzeit für guten interaktiven Verkehr ausschlaggebend. Verzögerungen über 0,75s vereiteln visuelle Kommunikation, weil der Unterschied zum direkten persönlichen Gespräch zu groß wird. Der schlechteste Prüfling ergab eine Latenzzeit von 0,59s, der beste bloß 0,27s. Auch die Tonver-

zögerung erinnert an den gleichen Effekt bei Satellitenübertragung. Der Prüfling von Northern Telecom verzichtete auf die eingebettete Tonübertragung und verwendete für die Tonkommunikation das normale Telefon. Das Videobild gleicht hier, wie schon zuvor erwähnt, einem Stummfilm. Eines der Systeme zeigte überhaupt keinen Latenzunterschied zwischen Bild und Ton, weil es einen der beiden 64-kbit/s-Kanäle für die Tonkommunikation verwendete. Dabei umgeht das Tonsignal den Codec, was aber bedeutet, daß das codierte Videosignal nur eine sehr kleine Latenz aufweisen darf.

In den qualitativen Tests lieferte das Gerät von Pictoretel bei kleinstem fps-Wert das schärfste Bild und die beste Farbproduktion. Wieder wurde bestätigt: Je höher die Frame Rate, desto geringer die „overall quality“. Das Intel-System benutzte einen ausgeprägten Tiling-Effekt bei einem fps-Wert von 11,6 im Durchschnitt. Dieser bewährt sich besonders bei farbtintensiven Stellen der Prüfclips. Deutlich bemerkbar war auch der Einfluß der Beleuchtung. Die verbreiteten Leuchtstoffröhren in Büros führen zu einer Überbetonung der hellen und dunklen Stellen. Auch bei den Audiotests war das Gerät PCS 50 durch die Klarheit der Sprache führend, die Verständlichkeit lag deutlich über jener von Telefonleitungen. Der Verkehr erfolgt in Vollduplexbetrieb.

Ein Hauptgewicht der Untersuchung lag auf der Fähigkeit jedes Systems, Zweifweg-FileTransfers, kooperative Applikationen (samt „Whiteboarding“) gleichzeitig mit der Videokonferenzsitzung zu übertragen. In der Beurteilung erhielt dieser Test das Gewicht von 10 %. (20 % wurden für live video, live audio und Synchronisation als Gewichtung vorgesehen, die beiden Latenzzeiten erhielten nur je 5 % Gewicht). Zwei der geprüften Systeme hatten arge Probleme mit der Bildqualität beim Austausch der Applikationsdaten. Bei File-Transfers sank ihr fps-Wert auf 5,3 ab. Das war aber keine Überraschung, weil beide Produkte Video und Daten über einen gemeinsamen 64-kbit/s-ISDN-Kanal übertrugen. Weit zufriedener arbeiteten zwei andere Systeme (darunter das PCS 50), die Ton und Daten auf einem Kanal mischen und das Videosignal auf dem anderen übertragen. Deshalb blieben auch die fps-Werte unverändert. Im System PCS 50 wird ein Betrieb namens „application sharing“ derart durchgeführt, daß zunächst eine Station das Bild einer Applikation an ihren Partner sendet. Ändert sich das Bild, bzw. wird es „updated“, dann wird dieses neue Bild dem Partner gesendet. Dieser kann hierbei die Applikation steuern, indem Kommandos an den sendenden Partner geschickt werden. Somit können beide Partner auf ihrem Arbeitstisch anteilig eine gemeinsame Applikation bearbeiten. Bei dem hier geschilderten Test wurden 3 Applikationen von Microsoft verwendet: Word, Excel und der Windows Calculator. Ein Problem tauchte allerdings auf: Bei bestimmten Applikationen kann ein Partner unter Umständen zu viel Kontrolle über den PC seines Partners gewinnen.

Ökonomische Überlegungen

Ältere österreichische Ingenieure hatten seit ca. 15 Jahren Gelegenheit, den unglaublichen Aufschwung der kleinen Firma Pictoretel zu verfolgen, die mit ein paar überaus begabten Leuten und unter Einsatz von Risikokapital das Wagnis unternahm, eine wirklich zufriedenstellende Videokonferenzstation zu entwickeln. Für die Datenverarbeitung war ursprünglich ein Kasten von ca. 70 cm Höhe erforderlich, der neben dem Schreibtisch aufgestellt werden mußte. Kein Wunder, daß die ersten Richtpreise in abschreckender Höhe lagen.

Inzwischen wurden durch Einsatz raffiniert ausgelegter ASICs und passender Software Gestehungskosten erzielt, die den Erwerb solcher Systeme in den Bereich des Möglichen rückten, ganz besonders in Regionen, die bereits mit ISDN operieren. Kritische amerikanische Stimmen meinen zwar, daß Videokonferenzen nun wohl in Sichtweite sind, daß man vorläufig aber mit dem breiten Ersatz der üblichen Dienstreisen durch diese noch junge Technik zuwarten sollte. Diese vorsichtige Haltung ist aber zum Teil darauf zurückzuführen, daß ISDN in den USA, im Gegensatz zu Europa, noch in den Kinderschuhen steckt.

Für unsere Betriebswirte müßte es eine interessante und herausfordernde Aufgabe sein, eine diesbezügliche Kosten-Nutzen-Rechnung aufzustellen. Man könnte z.B. als Basis ein Unternehmen mit 3 oder 4 Standorten annehmen, die über Schmalband-ISDN in Verbindung stehen, könnte die Materialinvestition samt Installationskosten und Logistik mit einer Amortisation in z.B. 5 Jahren in Gegenrechnung zu den Reisekosten von Geschäftsleuten der 4 Standorte stellen und auch deren Zeiteinbußen berücksichtigen, die aus ihrer häufigen Abwesenheit vom Dienstort entstehen.

Man könnte aber auch ganz andere Überlegungen anstellen. Etwa bei der Neuentwicklung eines heiklen Produkts, wenn Produktionsstätte und Entwicklungslabor voneinander entfernt sind, oder ähnlich, wenn ein Servicezentrum und die Qualitätssicherungsabteilung an verschiedenen Orten angesiedelt sind. In beiden Fällen ließe sich bei sorgfältiger Beobachtung über einen längeren Zeitraum sicher ein gravierender, Bearbeitungszeit senkender und damit ökonomischer Vorteil der Videokonferenzen feststellen lassen, besonders wenn sie auf Basis von leistungsstarken PCs durchgeführt werden. □