

Diss. ETH No. 15101

# Qualitätsaspekte multimodaler Kommunikation

Subjektive und objektive Messungen

ABHANDLUNG  
zur Erlangung des Titels  
DOKTOR DER TECHNISCHEN WISSENSCHAFTEN

der  
EIDGENÖSSISCHEN TECHNISCHEN  
HOCHSCHULE ZÜRICH

vorgelegt von  
Alexander Martin Braun  
Dipl. El.-Ing. ETH

geboren am 26. Juni 1973  
Deutscher Staatsbürger

Angenommen auf Antrag von  
Prof. Dr. Albert Kündig, Referent  
Prof. Dr. Dr. Helmut Krueger, Korreferent  
Prof. Dr. Burkhard Stiller, Korreferent

Zürich 2003

TIK-Schriftenreihe Nr. 50

Copyright 2003 by Alexander M. Braun

All rights reserved.

# Vorwort

---

Die vorliegende Arbeit entstand während meiner Tätigkeit am *Institut für Technische Informatik und Kommunikationsnetze (TIK)* am Departement für Informationstechnologie und Elektrotechnik der ETH Zürich.

Die Arbeit ist Teil des umfassenden QED-Projekts (Qualitätsmerkmale multimedialer Echtzeit-Dialogkommunikation), einer Kooperation des TIK und des Instituts für Arbeitshygiene und Ergonomie (IHA). Für die gute Zusammenarbeit und die wertvollen Gespräche danke ich meinem Projektkollegen Hans-Jörg Zuberbühler.

Mein besonderer Dank gilt Herrn Prof. A. Kündig, der mir das Umfeld, die Möglichkeiten und die Zeit bot, mich der in dieser Arbeit behandelten Problematik zu widmen. Speziell danke ich ihm für die individuelle Betreuung, den nötigen Freiraum und die sorgfältige Durchsicht der Arbeit. Ich danke den Herren Prof. H. Krueger und Prof. B. Stiller für die Übernahme des Korreferats und für ihre konstruktiven Vorschläge in der Endphase.

Weiter danke ich meinen Kollegen am Institut Patrik Estermann, Oliver Lamparter, Philipp Scherler, Bernhard Stauffer und Urs Röthlisberger für ihre jahrelange Unterstützung und das kollegiale Umfeld.

Mein Dank geht auch an die Studierenden, die im Rahmen des QED-Projekts Semester- und Diplomarbeiten durchgeführt haben, namentlich Niccolò Borer, Mathias Kienholz, Glenn Oberholzer, Philip Renggli und Patrik Scherer.

Ein besonderes Dankeschön gilt meinen Eltern und Grosseltern, die mir meine Ausbildung und damit diese Arbeit ermöglicht haben. Besonders danke ich meinem Vater für das Korrekturlesen der Arbeit.

Schliesslich danke ich meiner Frau Christine für Ihre unermüdliche Geduld und die Unterstützung.



# *Kurzfassung*

---

Qualitätsaspekte bei moderner Multimediakommunikation werden zunehmend wichtiger. Bei der Entwicklung von Netzwerken und Anwendungen muss vermehrt berücksichtigt werden, dass letztlich der Benutzer darüber entscheidet, ob er mit der gebotenen Qualität zufrieden ist oder nicht. Für benutzerbasiertes Netzwerkengineering fehlen jedoch weitgehend Modelle und Erkenntnisse der sich wandelnden Kommunikationsumgebung. Einzig Untersuchungen am University College of London beschäftigen sich intensiv mit diesem Gebiet.

Innerhalb der vorliegenden Arbeit wird einerseits der Bereich aufgezeigt, in dem zukünftig vermehrt geforscht werden muss, andererseits wird untersucht, wieviel Zweiwegverzögerung von Benutzern einer Videokonferenz maximal toleriert wird. Zu diesem Zweck wurden sowohl subjektive Versuche mit Testpersonen durchgeführt, als auch objektive Parameter gemessen, die Rückschlüsse auf die Zufriedenheit ohne direkte Befragung der Probanden erlauben.

Es hat sich gezeigt, dass die Benutzer in Abhängigkeit der Randbedingungen erstaunlich grosse Zweiwegverzögerungen im Vergleich zu bisher durchgeführten Studien in Kauf nehmen. Je nach Szenario liegen diese Werte zwischen 800 und 2800 ms. Der grösste Einfluss liegt bei der zu lösenden Aufgabe, aber auch technisches Vorwissen, die Tageszeit, die Anzahl der Gesprächsteilnehmer und die persönliche Verfassung der Probanden spielen unter anderem eine Rolle.

Des weiteren konnte nachgewiesen werden, dass anhand von Parametern wie der absoluten Sprechzeit und Anteilen von Double- und Tripletalk eine Aussage über die Zufriedenheit der Benutzer möglich ist. Aus den gewonnenen Erkenntnissen wurde ein Modell erstellt, das die Abhängigkeiten von Wahrnehmungs- und Akzeptanzschwelle und den miteinbezogenen unabhängigen Variablen veranschaulicht.

Weitere Untersuchungen mit angepassten Szenarien sind nötig, um schliesslich ein umfassendes Modell und einen Katalog mit verschiedenen Kommunikationssituationen und den entsprechenden Qualitätsparametern zu erhalten.

# *Abstract*

---

Quality aspects of multimedia communication are getting more and more relevant. When developing networks and applications, it has to be taken into account that it is finally the user who decides if the quality provided is acceptable or not. However for user-based network engineering there are models and knowledge of the changing communication environment missing. There is only one group at the University College of London that is seriously engaged in that area.

Within this thesis, on one hand the field is pointed out in which research has to be done. On the other hand the maximum of two-way delay is being explored that is tolerated by users of a video conference. To this end subjective tests have been conducted and objective parameters have been measured, that permit conclusions about user satisfaction without directly asking them.

It has been shown that users take into account really big values of two-way delay compared to other studies depending on the acilliary conditions. According to the scenarios these values vary between 800 and 2800 ms. The task being solved has the biggest influence at the threshold of acceptance, but technical knowledge, time of day, the number of participants and personal constitution also play a role.

In addition it could be verified that it is possible to combine parameters as absolute talk time, double and triple talk to ratings of user satisfaction. Out of measurement and findings a model has been developed that visualizes the dependencies of perception thresholds and acceptance threshold and the independent variables involved.

Further testing with adapted scenarios has to be performed to eventually arrive at the enclosing model and a catalogue of different communication situations and the correspondent quality parameters.

# *Inhaltsverzeichnis*

---

<b>Vorwort</b> .....	<b>iii</b>
<b>Kurzfassung</b> .....	<b>v</b>
<b>Abstract</b> .....	<b>vi</b>
<b>Inhaltsverzeichnis</b> .....	<b>vii</b>
<b>1 Einleitung</b> .....	<b>1</b>
1.1 Motivation .....	2
1.2 Problemstellung .....	3
1.3 Aufbau der Dissertation .....	7
<b>2 Verwandte Untersuchungen</b> .....	<b>9</b>
2.1 Konzepte und Modelle .....	9
2.2 Bewertungsverfahren .....	12
2.3 Untersuchungen in der klassischen Psychophysik .....	13
2.4 Untersuchungen zum Qualitätsempfinden bei Dialogsituationen ..	18
2.5 Schlussfolgerung .....	23
<b>3 Trends in der Telekommunikation</b> .....	<b>25</b>
3.1 Basistechnologien .....	25
3.2 Funktionalität und Mehrwert .....	26
3.3 Transportdienste .....	27
3.4 Einstellung und Verhalten der Benutzer .....	27
3.5 Trends bei den Endgeräten .....	28
3.6 Trends bei den Endverbraucherdiensten im Heimbereich .....	30
3.7 Trends bei den Endverbraucherdiensten im Geschäftsbereich .....	32

---

3.8 Schlussfolgerungen . . . . .	33
<b>4 Taxonomie der Kommunikation . . . . .</b>	<b>35</b>
4.1 Einleitung . . . . .	35
4.2 Einführung eines Modells . . . . .	36
4.3 Zeitliche Kommunikationsmuster . . . . .	47
4.4 Schlussfolgerungen . . . . .	50
<b>5 Vorgehen und Arbeitshypothesen . . . . .</b>	<b>51</b>
5.1 Abgrenzen des Forschungsgebiets. . . . .	51
5.2 Auswahl der Versuchsplattform . . . . .	65
5.3 Arbeitshypothesen und Fragestellungen. . . . .	66
<b>6 Durchführung der Versuche . . . . .</b>	<b>69</b>
6.1 Versuchsreihe 1 . . . . .	69
6.2 Versuchsreihe 2 . . . . .	84
6.3 Versuchsreihe 3 . . . . .	94
6.4 Versuchsreihe 4 . . . . .	106
6.5 Schlussfolgerungen . . . . .	112
<b>7 Messungen der Sprechaktivitäten . . . . .</b>	<b>115</b>
7.1 Methode . . . . .	116
7.2 Ergebnisse . . . . .	117
7.3 Diskussion. . . . .	119
<b>8 Schlussbetrachtungen und Ausblick . . . . .</b>	<b>123</b>
8.1 Modell der Benutzerzufriedenheit . . . . .	123
8.2 Empfehlungen für Netzwerkplaner und -gestalter . . . . .	125
8.3 Offene Fragen und Ausblick . . . . .	125
<b>Abkürzungen . . . . .</b>	<b>127</b>
<b>Referenzen . . . . .</b>	<b>129</b>
<b>Glossar . . . . .</b>	<b>139</b>
<b>Lebenslauf . . . . .</b>	<b>143</b>



# *Kapitel 1*

## *Einleitung*

---

Über Jahrzehnte hinweg wurde die Kommunikationstechnik mit unterschiedlichen Zielsetzungen ('Sprache' versus 'Daten') weiterentwickelt. Dies führte zu Kommunikationsnetzen, die auf unterschiedliche Anwendungen zugeschnitten waren und sich daher nur bedingt oder überhaupt nicht für andere Zwecke eigneten. Der Begriff *Telekommunikation*<sup>1</sup> wurde bis vor einigen Jahren überwiegend für die Anwendung *Telefonie* verwendet. Die Übertragung von Daten spielte hingegen eine untergeordnete Rolle. Insbesondere durch den enormen Erfolg des *Internets*<sup>2</sup> in den letzten zehn Jahren ist eine Entwicklung in Gang gesetzt worden, die die unterschiedlichen Netze vermehrt zusammenwachsen lässt. Über das Telefon- bzw. ISDN-Netz werden Daten verschickt und das Internet wird zur Übertragung von Sprache und Bewegtbildern genutzt. Diese Entwicklung hat gleichzeitig die Entstehung neuer Anwendungen gefördert, die bis anhin eine untergeordnete Rolle gespielt hatten, sogenannte *Multimediaanwendungen*. Man spricht in der heutigen Zeit deshalb auch von *Multimediatelekommunikation*<sup>3</sup>. Diese erweitert die Möglichkeiten der klassischen Telekommunikation, indem beispielsweise neben der Sprache auch Bewegtbilder des Gesprächspartners, ein Mauszeiger, Dokumente oder ein sogenannter *Shared Workspace* übertragen werden. Es stellt sich die Frage, ob bisherige Normen, Empfehlungen, Modelle und Untersuchungen hinsichtlich der Dienstgüte (*Quality of Service, QoS*<sup>2</sup>) auch für diese neuen Kommunikationsformen und -dienste gültig sind, oder ob diese nicht vielmehr wesentlich erweitert und angepasst werden müssen.

---

<sup>1</sup> Informationsaustausch mit Hilfe von Nachrichtentechnik und Übertragungsnetzen.

<sup>2</sup> siehe Glossar

<sup>3</sup> Telekommunikation unter Verwendung mehrerer Medien: Sprache, Bild, Shared Workspace etc.

## 1.1 Motivation

Die Gründe für den vermehrten Einsatz von Multimediaanwendungen und neuen Dienstleistungen sind vielschichtig. Zum einen werden immer mehr Arbeitsprozesse informatisiert<sup>1</sup> und vermehrt Nachrichteninhalte digitalisiert. Dies führt zu einem wachsenden Angebot an Information, das verfügbar und abrufbar ist. Des Weiteren steigt durch Globalisierung und Mobilität das Interesse, auf Informationen jederzeit und von überall her zugreifen zu können. Zum anderen müssen sich die enormen Investitionen in die Transportinfrastruktur während der frühen Neunzigerjahre bezahlt machen, und die entsprechenden Unternehmen suchen nach Diensten und Inhalten, die sie ihren Kunden verkaufen können. Weitere Gründe wie ein sich sozusagen selbsterfüllendes *Moore's Law*<sup>2</sup> [49] oder diverse Hypes und Modeerscheinungen der modernen Gesellschaft tragen ihr Übriges zu dieser Entwicklung bei.

Netzwerke sind zu einem wichtigen und wirtschaftlich wesentlichen Rückgrat der Gesellschaft in den Industrieländern geworden. Mit der Abhängigkeit von Netzwerken und Netzwerkdiensten steigt der Bedarf an Mechanismen zur Qualitätssicherstellung. Momentan bieten Netzwerke wie das Internet *Best Effort*-Dienste<sup>3</sup> an. Mit dieser Methode können keine Garantien über eine bestimmte Dienstqualität abgegeben werden. Die einzige Möglichkeit, einen genügenden Durchsatz bei annehmbaren Verzögerungen sicherzustellen, besteht darin, genügend Bandbreite zur Verfügung zu stellen, was jedoch zu einer schlechten Nutzung der Ressourcen führt.

Um diese Unzulänglichkeiten zu beheben, werden verschiedene Lösungsansätze verfolgt. Die Hauptidee besteht darin, objektiv messbare, technische Parameter heranzuziehen, die dazu verwendet werden können, den Zustand eines Netzwerkes zu beschreiben sowie verschiedenste QoS-Anforderungen zu garantieren. Für paketorientierte Netze werden momentan vor allem zwei Lösungsansätze diskutiert: *Integrated Services* (IntServ) und *Differentiated Services* (DiffServ). Das Konzept von IntServ [65] basiert auf der Reservation von Ressourcen in den einzelnen Netzwerkelementen mit dem *Resource Reservation Protocol* (RSVP) [101]. Dieser Ansatz bietet ein Maximum an QoS-Unterstüt-

---

<sup>1</sup> Unter Informatisierung versteht man die Technisierung des Wissens auf der Basis von Computern [94]

<sup>2</sup> Im deutschsprachigen Raum auch als Moorsches Gesetz bekannt. Es sagt aus, dass sich die Leistung von Computern alle 18 Monate verdoppelt.

<sup>3</sup> siehe Glossar und [80]

zung, gleichzeitig ist aber ein hoher Rechenleistungsaufwand in jedem Knoten notwendig, was Skalierungsprobleme mit sich bringt. Bei DiffServ [66] wird QoS hingegen nicht pro Datenstrom definiert, vielmehr werden mehrere Datenströme mit ähnlichen QoS-Anforderungen zu Klassen zusammengefügt.

Bevor jedoch eine bestimmte QoS innerhalb der Netzwerke garantiert werden kann, muss diese nach Ansicht des Autors definiert werden. Es ist letztlich der Endbenutzer, der bei einer bestimmten Anwendung eine bestimmte Dienstgüte erwartet, die von verschiedenen Kriterien abhängt. Für eine optimale Dimensionierung ist es deshalb wichtig, dass einerseits diejenigen Qualitätsparameter optimiert werden, die im speziellen Fall wichtig sind, andererseits sollten bei unwichtigen Parametern keine Ressourcen verschwendet werden. Dazu braucht es passende Modelle, die sowohl die Erwartungen und Anforderungen der Benutzer als auch die möglichen Kommunikationssituationen und -anwendungen miteinbeziehen und Schlussfolgerungen für die notwendige QoS erlauben.

Aufgrund derartiger Überlegungen wurde in der Fachgruppe Systemtechnik am Institut für Technische Informatik und Kommunikationsnetze der Eidgenössischen Technischen Hochschule Zürich das Projekt QED (Qualitätsmerkmale der Echtzeit Dialogkommunikation) initiiert. Die vorliegende Arbeit ist ein Teil dieses Projekts. Nachfolgend wird darauf eingegangen, welche Ziele in dieser Arbeit verfolgt werden, in welchem Rahmen Untersuchungen gemacht wurden und insbesondere auch, welche Aspekte nicht berücksichtigt bzw. bewusst ausser Acht gelassen wurden.

## *1.2 Problemstellung*

Die traditionellen Telekommunikationsnetzwerke konnten im wesentlichen mit folgenden Eigenschaften charakterisiert werden: Nahezu konstante (und bei drahtgebundener Übertragung niedrige) Verzögerung, sehr kleine Fehlerraten im Festnetz, Netzwerkdienste mit konstanter Bandbreite. Im Gegensatz dazu werden Endsysteme in modernen paketorientierten Netzwerken (wie dem Internet) mit Merkmalen wie variabler Verzögerung (Jitter), variablen Bitraten und möglichem Zell- bzw. Paketverlust konfrontiert. Diese Phänomene können je nach Art der gewählten Quellenkodierung und Kompression zu Beeinträchtigungen führen, die vom Benutzer wahrgenommen werden. Darüber hinaus sind die Benutzer immer häufiger mit der Frage konfrontiert, welche Qualitäts- und Leistungskennwerte – abhängig von der Anwendung – für

einen bestimmten Preis angemessen sind. Grundsätzlich sind inskünftig drei verschiedene Zeitpunkte denkbar, wann QoS-Parameter festgelegt werden. Erstens kann dies in einer Art Qualitätsvertrag geschehen, wenn der Kunde beim Netzwerkanbieter einen entsprechenden Anschluss mietet. Zweitens kann die entsprechende Qualität zu Beginn jeder Verbindung festgelegt werden, und schliesslich ist es auch denkbar, die Qualität dynamisch während der Verbindung an die Erfordernisse der Applikation anzupassen.

Wenn man bedenkt, dass sowohl die Netzwerktechnologien als auch die angebotenen Netzwerkdienste permanent weiterentwickelt werden, müsste man eigentlich davon ausgehen können, dass diese Entwicklungen aufgrund von allgemein anerkannten und bestens erforschten Grundlagen hinsichtlich der Anforderungen der Benutzer vorangetrieben werden. Umso erstaunlicher ist es, dass es im Bereich der Benutzeranforderungen bzw. der Qualitätsansprüche an moderne Multimediakommunikation praktisch keine systematischen Untersuchungen gibt.

Zwar sind im Falle von reiner Sprachkommunikation viele Untersuchungen gemacht worden um herauszufinden, welche Qualität der Benutzer erwartet. Stellvertretend sei hier die Empfehlung der ITU genannt [40], die für reine Sprachkommunikation eine maximale Einwegverzögerung von 150 ms empfiehlt. Auch wurden im Bereich der Psychophysik verschiedenste Studien durchgeführt, in denen es um die Wahrnehmung von Bild- und Toninformationen geht. In beiden Fällen geht es jedoch nur um Einweg- und nicht um Dialogsituationen. Gerade bei der Dialogkommunikation treten aber besonders kritische Situationen auf. Selbst wichtige Studien wie von Steinmetz in [79] beschrieben, die sich mit der Frage beschäftigen, wie wichtig bei der Verwendung von mehreren Multimediadatenströmen die zeitliche Synchronisation untereinander ist, betrachten reine Einwegkommunikation. Einzig Untersuchungen einer Forschungsgruppe am University College of London [12], [87] bewegen sich innerhalb des aufgezeigten Gebiets.

Dementsprechend ist es momentan sehr schwierig, die Planung und Dimensionierung von Netzen auf saubere Verkehrsvorhersagen und echte Benutzerbedürfnisse abzustützen. Um die vorhandenen Lücken zu schliessen, wurde das bereits in Abschnitt 1.1 beschriebene Projekt QED initiiert. Die vorliegende Arbeit ist im Kontext von QED durchgeführt worden. In dessen Rahmen wird versucht, anhand systematischer Untersuchungen, subjektiver Tests und objektiver Messungen Rahmenbedingungen für benutzerbasierte QoS-Parameter zu definieren. Die Arbeit befasst sich vor allem mit der Frage, wie die von Netzwerken und Endgeräten auferlegten Beschränkungen und die durch sie einge-

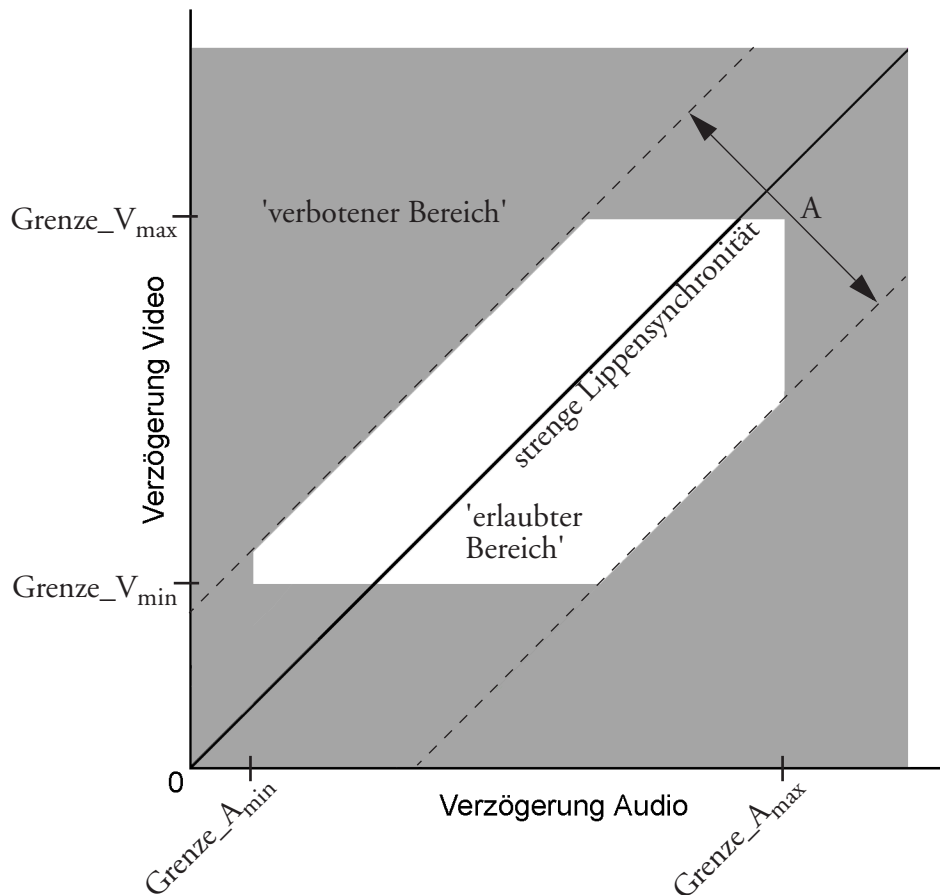
fügten Störungen (wie z.B. Verzögerungen oder begrenzte Auflösung) sich auf die Interaktion von zwei oder mehreren Teilnehmern in einer Telekommunikationssituation auswirken.

Mit den Ergebnissen der Untersuchungen soll letztlich ein qualitätsbasiertes Netzwerkengineering ermöglicht werden. Wünschenswert ist es dabei, alle nur erdenklichen Kommunikationsformen einzubeziehen. Selbst wenn man Untersuchungen auf den Bereich der interpersonellen Echtzeitkommunikationsformen begrenzt, sind der zu erforschende Raum derart gross, die Qualitätsparameter derart vieldimensional und die Abhängigkeiten derart komplex, dass diese den Rahmen einer einzelnen Dissertation sprengen würden. Trotzdem ist es sinnvoll, mit der vorliegenden Auswahl einen Grundstein zu legen, auf dem aufbauend weitere Forschungsarbeiten das Rahmenwerk nach und nach vervollständigen können.

Besonders interessant sind bei diesen Anwendungen Betrachtungen der Verzögerungen, die in technischen Systemen und bei der Übertragung von Multimediatdatenströmen entstehen, weil die Qualität einer Multimediakommunikation immer einen Kompromiss zwischen der zur Verfügung stehenden Bandbreite und dem Aufwand der Signalverarbeitung darstellt, der sich schliesslich in einer mehr oder weniger grossen Verzögerung beim Benutzer bemerkbar macht. Abbildung 1.1 zeigt schematisch für eine Übertragung von Audio und Video, in welchem zweidimensionalen Zeitfenster sich die Verzögerungen aufhalten können bzw. dürfen, damit sie sich innerhalb des *erlaubten Bereichs*<sup>1</sup> befinden. Dieser Bereich wird nach unten hin von den systembedingten Verzögerungen begrenzt, die durch Laufzeit, Kodierung, Pufferung usw. hervorgerufen wird ( $Grenze_{V_{min}}$ ,  $Grenze_{A_{min}}$ ). Da die Komplexität der Signalverarbeitung von Video grösser ist als von Audio, ist die untere Grenze absolut gesehen für Videosignale kleiner. Nach oben hin wird der Bereich von der Toleranzgrenze der Benutzer begrenzt. Diese Grenze ist individuell pro Benutzer und auch abhängig von diversen Faktoren, insbesondere der zu lösenden Aufgabe ( $Grenze_{V_{max}}$ ,  $Grenze_{A_{max}}$ ).

---

<sup>1</sup> Der *erlaubte Bereich* in diesem Zusammenhang ist ein Bereich, in dem sich Kombinationen von Verzögerungswerten von Video- und Audioströmen bewegen müssen, weil sie einerseits technisch bedingt eine Untergrenze besitzen, und weil sie andererseits oberhalb einer bestimmten Grenze vom Benutzer nicht mehr akzeptiert werden. Zusätzlich müssen bestimmte Bedingungen bezüglich der Lippensynchronität erfüllt sein, nämlich dass die Differenz der beiden Datenströme zueinander einen bestimmten Toleranzwert seitens der Benutzer nicht überschreiten.



A: Bereich aufgespannt durch die Toleranz des Benutzers, dass der Ton dem entsprechenden Bild folgen oder vorangehen darf. Der Bereich ist gegenüber der strengen Lippsynchronität asymmetrisch aufgeteilt, da dies der natürlichen Erfahrung des Menschen entspricht (Lichtgeschwindigkeit ist schneller als Schallgeschwindigkeit)

*Abbildung 1.1: Verzögerungsbereiche von Video und Audio*

Ganz klar einzugrenzen ist das Forschungsgebiet der vorliegenden Arbeit auch dahingehend, dass die Wechselwirkungen zwischen einem angebotenen Dienst, der entsprechenden QoS und dem Wert, den ein Benutzer dieser Kombination zuweist, nicht untersucht werden kann. Sehr interessant ist diesbezüglich unter anderem die Frage, wie sich die Schwelle, ab der eine Qualitätsverschlechterung als störend beurteilt wird, verschiebt, wenn für eine bestimmte Qualität eine bestimmte Gebühr erhoben wird.

### 1.3 *Aufbau der Dissertation*

In **Kapitel 2** werden verwandte Untersuchungen vorgestellt bzw. Ergebnisse aus der Literaturrecherche zusammengefasst.

In **Kapitel 3** folgt eine Übersicht über Trends in der Multimediakommunikation aus einem universitären Blickwinkel.

In **Kapitel 4** werden zwei zueinander orthogonale Ansätze einer *Taxonomie der Kommunikation*<sup>1</sup> vorgestellt. Diese dienen als Koordinatensystem für die Versuche sowie als Diskussionsgrundlage für die Schlussfolgerungen.

**Kapitel 5** zeigt das Vorgehen bezüglich der durchgeführten Versuche auf. Es wird auf die Auswahl der Versuchsplattform eingegangen und die Arbeitshypothesen werden vorgestellt.

In **Kapitel 6** werden die Versuche genau beschrieben und es wird auf die Ziele, Methoden und Ergebnisse eingegangen.

Nachdem sich **Kapitel 7** mit der Frage beschäftigt hat, ob subjektive Versuche durch objektive Messungen zu ersetzen sind und welche Ziele auf diesem Gebiet erreicht wurden, folgen Diskussion und Schlussfolgerungen in **Kapitel 8**.

---

<sup>1</sup> siehe Glossar





# *Kapitel 2*

## *Verwandte Untersuchungen*

---

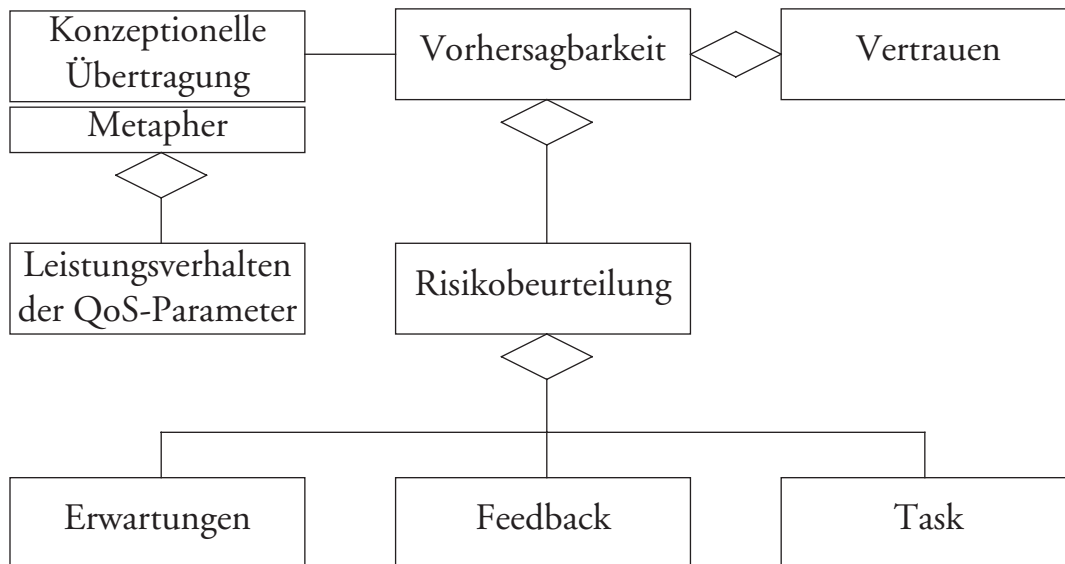
Um die Benutzeransprüche bezüglich QoS und interpersoneller Multimedia-kommunikation zu erforschen, muss zuerst ein Rahmen definiert werden. Dazu gehört ein konzeptionelles Modell, das aufzeigt, welche Faktoren und Qualitätsparameter einen Einfluss auf die Anforderungen haben und wie diese untereinander zusammenhängen. Es muss auch berücksichtigt werden, dass – anders als beispielsweise bei der Telefonie – mindestens zwei Medien Verwendung finden. Zweitens muss festgelegt werden, wie die Erfassung der (subjektiven) Benutzeransprüche geschehen soll. Und schliesslich müssen Szenarien (Kommunikationssituationen) festgesetzt werden, die gewisse Randbedingungen (realistische Anwendungen, sinnvolle Aufgaben, realitätsnahe Umgebung etc.) erfüllen müssen. Nachfolgend wird anhand einer Literaturrecherche der aktuelle Forschungsstand aufgezeigt.

### *2.1 Konzepte und Modelle*

Bei den Modellen gibt es bis auf eine Ausnahme nur solche, die für reine Sprachübertragung (z.B. E-Model [39] oder QoS-Schema von Möller [58]) erstellt wurden oder aber solche, die für Verteilmedien wie das Fernsehen gelten.

Eine Forschungsgruppe am *University College of London (UCL)*, die Untersuchungen im Bereich Benutzerwahrnehmung und Dienstqualität anstellt, und die sich das Ziel gesetzt hat, der Netzwerkgemeinschaft ein Rahmenwerk für die Abbildung von Benutzeranforderungen auf QoS-Parameter des Netzwerks zu erstellen, hat ein konzeptionelles Modell mit Überlegungen und Faktoren auf höchster Ebene, die massgeblich daran beteiligt sind, dass ein Benutzer ein QoS-Schema akzeptiert, geschaffen [10], [14], [15], [16], [86], [91]. Dieses

Modell [13] wurde durch Studien im Bereich von Erwartungen und Verhalten der Benutzer betreffend Dienstqualität erstellt und im Laufe der letzten Jahre verfeinert. Es gibt eine Übersicht über die Abläufe, die an der Qualitätsbeurteilung durch die Benutzer beteiligt sind (Abbildung 2.2).



*Abbildung 2.2: Konzeptionelles Modell von Bouch/Sasse*

Bouch/Sasse postulieren, dass einem bestimmten Dienst durch den Benutzer ein Wert zugeordnet wird, sobald der Benutzer für die Einhaltung einer bestimmten Dienstqualität etwas bezahlen muss. Der Schlüsselfaktor ist das Vertrauen des Benutzers, dass das jeweilige Kostenschema entsprechend dem aktuellen QoS-Zustand einen fairen Wert aufweist [11]. Was die Frage des Vertrauens in einen Dienst betrifft, wird die Modellierbarkeit in erster Linie durch die vorhersagbare Qualität erreicht [12]. Vorhersagbarkeit bedeutet in diesem Fall die Sicherheit, dass die beantragte und benötigte Dienstgüte auch in Zukunft zur Verfügung stehen wird. Es hat sich gezeigt, dass Benutzer einer tieferen, aber stabil bleibenden Qualität einer höheren, aber wechselnden Qualität den Vorzug geben. Dieses interessante Ergebnis bedeutet aber nicht, dass die Netzwerkbandbreite verkleinert werden kann, sondern dass Verkehrsspitzen in Multimediadatenströmen ausgemittelt werden sollten. Auf einer höheren Ebene wird die Vorhersagbarkeit dadurch erreicht, dass die Wünsche der Benutzer bezüglich ihres konzeptionellen Metaphers, der von ihnen gewählten Strategie und dem Risiko, das sie für eine bestimmte Kommunikation vorschlagen, berücksichtigt werden.

Laut Bouch erstellen sich (vorallem neue) Benutzer konzeptionelle Metaphern, um die Multimediatasks mit solchen aus der realen Welt in Verbindung zu bringen. Die Qualitätsansprüche, die in der realen Umgebung erwartet werden, werden anschliessend dazu verwendet, Qualitätsgrade und Vorhersagbarkeiten im Internet zu vergleichen und zu beurteilen. Beispielsweise wird die Zustellung von E-Mails mit der von Briefpost verglichen und da es bei letzterer keine Ankunftsbestätigung gibt, wird auch für E-Mails keine erwartet. Ein anderes Beispiel zeigt, dass für die Telefonie über IP nur die Qualität der bekannten ISDN-Telefonie erwartet wird, nicht jedoch HIFI-Qualität. Andererseits gelten für Internetradio höhere Ansprüche; diese werden mit UKW-Rundfunk verglichen. Für Videokonferenzen könnte dies bedeuten, dass die Benutzer beim Bild eine fernsehähnliche Qualität und beim Ton eine dem Telefon vergleichbare erwarten, oder sie könnten sogar Parallelen zu einem Gespräch von Angesicht zu Angesicht ziehen.

Die von Benutzern erstellten Strategien hängen einerseits von der zu lösenden Aufgabe, andererseits von den Erwartungen ab, die sich aus dem vom System erhaltenen Feedback ergeben [11]. Die Aufgabe (Task) und die dabei verwendete Semantik scheinen wichtige Faktoren bei der Beurteilung der subjektiv wahrgenommenen Qualität zu sein [14], [87]. Untersuchungen zeigen, dass die Teilnehmer eines Videokonferenztests ihre Zufriedenheit mit dem System unter den gleichen QoS-Bedingungen, aber abhängig von der Aufgabe unterschiedlich bewerteten [Kapitel 6], [55].

Obwohl das oben vorgestellte Modell speziell für QoS im Internet zugeschnitten worden ist, wurde es innerhalb dieser Arbeit als Referenzmodell für die Untersuchungen herangezogen. Basierend auf den Erfahrungen aus bisher bekannten Untersuchungen und den Ergebnissen der vorliegenden Arbeit wird ein eigenes, verfeinertes Modell erstellt werden, das jedoch nicht so viele Parameter mit einbeziehen wird. Der Ansicht des Autors nach bedarf es grundlegenderer Untersuchungen als die oben beschriebenen, um all die Zusammenhänge aussagekräftig in ein Modell einbinden zu können. Wie bereits an anderer Stelle erwähnt, wurde beispielsweise der Einfluss von unterschiedlich hohen Kosten für verschiedene Qualitätsstufen nicht untersucht, da dies erst auf einer höheren Ebene sinnvoll erscheint. Weitere konzeptionelle Modelle im Bereich des vorliegenden Forschungsgebietes wurden bisher nicht veröffentlicht.

## 2.2 *Bewertungsverfahren*

Es gibt verschiedene Verfahren, mit denen subjektiv wahrgenommene Qualitätseindrücke (Audio) gemessen werden. Die meisten davon basieren auf der fünfstufigen Skala des *Mean Opinion Score* (MOS, ITU Rec P.800). Bei diesem Verfahren werden einer gemischten Gruppe von Menschen unter kontrollierten Bedingungen Sprach-Beispiele vorgespielt. Die Qualität der Beispiele wird unter verschiedenen Aspekten nach dem Schulnotenprinzip beurteilt. Die Kriterien sind das Verstehen eines Sprechers (der Listening-Test), ein Gesprächstest (der Conversation-Test), die nötige Anstrengung, um etwas zu verstehen (der Listening-Effort) und die empfundene Lautstärke (die Loudness-Preference). Die Auswertung der Noten ergibt eine Kennzahl, die die Meinung einer Testperson widerspiegelt, der so genannte Mean Opinion Score, kurz MOS. Der MOS kann Werte zwischen 1,0 für schlecht und 5,0 für sehr gut ergeben, wobei ein Wert von 4,0 der vom herkömmlichen Telefon gewohnten Qualität entspricht. Nachteile dieses Tests sind der hohe Zeitaufwand, die nicht gegebene Reproduzierbarkeit der Ergebnisse und die organisatorischen Schwierigkeiten, würde man den Test regelmässig wiederholen wollen.

Darüberhinaus gibt es Verfahren, die das Signal vor der Übertragung mit dem Signal danach vergleichen und anhand von mathematischen Modellen eine 'objektive' Bewertung vornehmen, für Hörtests beispielsweise PSQM (Perceptual Speech Quality Measure, ITU Rec P.861) oder PESQ (Perceptual Evaluation of Speech Quality, ITU Rec P.862). Im PSQM-Verfahren werden Sprach-Samples jeweils vor und nach dem Passieren des Testaufbaus rechnerisch miteinander verglichen. Komplexe Formeln berücksichtigen hierbei nicht nur die physikalischen, sondern auch die psycho-akustischen Aspekte des menschlichen Gehörs. Das Ergebnis ist wiederum eine Kennzahl, die die Abweichung des empfangenen Samples vom Original ausdrückt, und zwar so, wie ein Mensch dies empfinden würde. Dabei ist 0 der beste Wert und steht für keine Verschlechterung, 6,5 ist der schlechteste mögliche Wert und bedeutet eine starke Verfremdung der Sprache. Die PSQM-Kennzahlen können leicht in den MOS umgerechnet werden.

Für die Beurteilung der Bildqualität werden Verfahren wie RMS (Root Mean Square), MTF (Modulation Transfer Function Area), SQRI (Square Root Integral) oder JND (Just Notable Differences) verwendet. Eine gute Übersicht der genannten Verfahren findet sich in [18].

Ein Verfahren für Multimediakommunikation, das für das Gebiet der vorliegenden Arbeit anwendbar wäre, ist dem Autor nicht bekannt.

## 2.3 *Untersuchungen in der klassischen Psychophysik*

Dieser Abschnitt gibt einen Überblick über bekannte Untersuchungen betreffend menschlicher Wahrnehmung von auditiven und visuellen Stimuli im Hinblick auf Störungen bei multimedialer Kommunikation. Unter *klassischer Psychophysik* wird hier das Untersuchen von Einwegsituationen verstanden. Interessant für die vorliegende Arbeit sind jedoch Zweiwegsituationen, die aber bisher sehr selten untersucht wurden.

### 2.3.1 *Verzögerungsdifferenzen zwischen auditiven und visuellen Stimuli*

Nachfolgend werden einige Untersuchungen sehr ausführlich besprochen, weil aufgezeigt werden soll, dass die Versuchsanlage einen grossen Einfluss auf die Resultate hat. Die Auswahl zeigt einen repräsentativen Querschnitt nach Meinung des Autors.

#### *Dixon und Spitz: Erkennen von auditiver und visueller Asynchronität [26]*

In der Studie von *Dixon* und *Spitz* wurden 18 englischsprachigen und zehn spanischsprachigen Testpersonen ein Film und dessen englischer Begleitton vorgeführt. Die verwendete Versuchsanordnung erlaubte dabei das Einstellen einer variablen Verzögerung zwischen Audio- und Videokanal. Die Testpersonen trugen Kopfhörer und schauten aus einem Meter Abstand auf einen Bildschirm mit 23 cm Diagonale. Ihnen wurde gesagt, dass sie während des Betrachtens der Szenen (ein Sprecher liest Prosatexte oder ein Hammer schlägt auf einen Amboss) einen Knopf drücken sollten, was dazu führen würde, dass Bild und Ton allmählich nicht mehr zueinander synchronisiert sein würden. Sobald sie diese Asynchronität bemerkten, sollten sie den Knopf loslassen.

Das aussagekräftigste Ergebnis über alle Versuche gesehen ( $p < 0,001$ ) war die grössere Empfindlichkeit betreffend der Asynchronität, wenn der Ton vor dem dazu passenden Bild abgespielt wurde als im umgekehrten Fall. Dieses Ergebnis galt für jede Testperson, für beide Ohren getrennt sowie gemeinsam als auch unabhängig von den beiden gezeigten Szenen (Mann, Hammer). Die zweite Variable, mit aussagekräftiger Auswirkung war der präsentierte Kontext: Alle Testpersonen stellten die Asynchronität in der „Hammerszene“ schneller fest als bei der Stimulierung durch Sprache ( $p < 0,002$ ).

*McGrath und Summerfield: Zwischenmodale Zeitbeziehungen und audiovisuelle Spracherkennung von normalhörenden Erwachsenen [56]*

In der Studie von *McGrath* und *Summerfield* mussten zwölf Testpersonen (normalhörend und normalsichtig bzw. mit korrigierter Sehkraft) herausfinden, ob der Beginn eines 120-Hz Tons vor oder nach dem Öffnen von gezeichnet nachgebildeten Lippen präsentiert wurde. Die Personen sassen 1.5m vor dem Bildschirm (10cm x 6,2cm) in einem abgeschirmten Raum. Die akustischen Signale wurden mit einem Spitzenwert von 73 dBA auf beiden Ohren wiedergegeben. Der Ablauf bestand aus drei Darbietungen und einer erzwungenen Entscheidung für die zwei präsentierten Möglichkeiten: Die erste Darbietung fand immer ohne Asynchronität statt, bei den beiden folgenden war eine ohne und eine mit. Die Aufgabe der Testperson war es nun, durch Knopfdrücken herauszufinden, bei welchem der beiden letzten Stimuli Bild und Ton nicht zueinander synchron waren.

Das Ergebnis zeigte, dass die Empfindung in Situationen signifikant grösser ( $p < 0,01$ ) war, in denen der Ton vor dem Öffnen der Lippen abgespielt wurde als umgekehrt.

*Lewkowicz: Wahrnehmung von zeitlicher audiovisueller Synchronität bei Kleinkindern [52]*

Das Ziel in *Lewkowicz*' Studie bestand in der Untersuchung der Entwicklung von Kindern bezüglich zeitlicher Erscheinungen. Zur Überprüfung der Richtigkeit seiner Vorgehensweise bei Kleinkindern führte er zehn Versuche mit Erwachsenen durch. Die Teilnehmer sassen 50 cm vor einem Display mit einer 63 cm Diagonalen. Sie beobachteten eine springende Platte und hörten einen Ton (63 Hz, 63 dBA), der immer dann auftrat, wenn die Platte sprang. Anschliessend wurde ihnen eine Reihe asynchroner Testversuche vergewahrt, bei denen der Ton entweder vor oder nach dem Sprung der Platte auftrat. Die Aufgabe der Teilnehmer war es, jedesmal zu entscheiden, ob Platte und Ton synchron auftraten oder nicht.

Die Ergebnisse erbrachten eine höhere Wahrnehmungsschwelle für die Asynchronität, wenn der Ton dem entsprechenden Bild nachfolgte als im umgekehrten Fall.

*Pandey, Kunov und Abel: Störeinflüsse durch Verzögerungen von Audiosignalen auf die Sprachwahrnehmung mit Lippenlesen [62]*

*Pandey* et al. untersuchten die Auswirkungen von Verzögerungen der Audiosignale auf die audiovisuelle Wahrnehmung von auf Video aufgezeichneten Sätzen bei zwölf normalhörenden Testpersonen. Das Testmaterial bestand aus Sätzen, vorgelesen von einem männlichen Sprecher, wobei teilweise ein Bild mit einem Schlüsselwort des jeweiligen Satzes gezeigt wurde. Den Teilnehmern wurde nur Audio, nur Bild oder beides gleichzeitig vorgeführt. Sprachsignale wurden mit einem S/N-Wert von 0 resp. 10 dB abgespielt. Die Störgeräusche wurden durch Gerede von vielen übereinander gelagerten Sprechern bei 60 dBA erzeugt.

Die Auswirkung der Audioverzögerung auf die Sprachwahrnehmung mit Lippenlesen war beim schlechteren S/N-Wert störender, jedoch für Verzögerungswerte von 80 bis 120ms nicht signifikant.

*Steinmetz [79]*

Das Forschungsziel bei *Steinmetz* bestand im Erlangen von ausführlichem Wissen über die Synchronisationsanforderungen bei der Benutzerschnittstelle von multimodalen Kommunikationsapplikationen. Es wurden zwei Fälle betrachtet: Die Anforderungen bei Lippen- sowie bei Zeiger-Synchronisation<sup>1</sup>. Bei der Lippensynchronisation wurde zwischen Ansichten von Kopf, Schulter und ganzem Körper unterschieden. Bei der Ansicht des ganzen Körpers haben Beobachter eine höhere Toleranz betreffend Asynchronität gegenüber der Kopfansicht. Situationen, in denen Audio vor Video präsentiert wird, werden als leicht störender als im umgekehrten Fall befunden.

Diese Untersuchungen sind vermutlich diejenigen, die die grössten Ähnlichkeiten haben dürften mit unseren eigenen, sowohl was die Methodik als auch die zu erwartenden Ergebnisse angeht.

---

<sup>1</sup> Es wurde untersucht, wie Benutzer auf die Asynchronität zwischen einem Mauszeiger, der in einem Shared Workspace auf Objekte zeigt, und der dazugehörigen Spracherklärung reagieren.

### 2.3.2 Diskussion und Vergleich

Studie	Situation	Grenzwert $t_{VA}$ [ms] (Video vor Audio)	Grenzwert $t_{AV}$ [ms] (Audio vor Video)	% $t_{AV}/t_{VA}$
Dixon/Spitz	Lippensynchronisation, Hammer & Amboss	258	131	51
		189	75	40
McGrath/ Summerfield	gezeichnete Lippen	138	79	57
Lewkowicz	springende Platte Erwachsene Kleinkinder	112	65	58
		450	350	78
Pandey/ Kunov/Abel	Lippenlesen mit Maskierungsrauschen	Einbussen bei der Performance bei 80 - 120 ms	wurde nicht gemessen	-
Steinmetz	Lippen-, Zeigersynchronisation	ca. 240 ca. 750	ca. 160 ca. 500	67 66

Tabelle 2.1: Grenzwerte für Asynchronität von Audio und Video

Das markanteste und signifikanteste Ergebnis bei allen Untersuchungen ist die menschliche Fähigkeit, bei asynchronen Multimediaströmen die Situation *Ton vor Bild* ca. doppelt so früh zu erkennen wie *Ton nach Bild*. Dieses Ergebnis ist durchaus plausibel, da es die Erfahrung der Menschen in einer natürlichen Umgebung widerspiegelt, in der die Lichtgeschwindigkeit deutlich höher ist als die Schallgeschwindigkeit und somit bei Naturereignissen wie Gewittern tatsächlich zuerst das Bild (Blitz) und dann der entsprechende Ton (Donner) wahrgenommen wird. Somit haben die Menschen „gelernt“, solche Situationen eher zu tolerieren und sind dadurch auch weniger empfindlich. Diese Erklärung wird durch eine weitere Tatsache untermauert: Es gibt einige Anzeichen dafür, dass die Toleranz gegenüber Asynchronitäten im Fall der Lippensynchronität grösser ist als beispielsweise beim Beobachten eines aufschlagenden Gegenstandes. Die Übersicht in Tabelle 2.1 zeigt deutlich, dass die gerade noch wahrnehmbaren Differenzen zwischen Bild und Ton in Abhängigkeit der verwendeten unterschiedlichen Versuchsanordnungen deutlich variieren. Da die Datenbasis aber relativ klein ist, müssen die aufgeführten Mittelwerte mit Vorsicht genossen werden. Aus diesem Grund wird eine leicht schwächere Aussage



vorgeschlagen, die anstatt der Schwelle der Wahrnehmung eher die Schwelle angibt, ab der die Asynchronitäten eine Beeinträchtigung bewirken: Bis hin zu einer Differenz von 160 ms bei der Wahrnehmung von Bild vor Ton (90 ms bei der Wahrnehmung von Ton vor Bild), kann davon ausgegangen werden, dass die Wahrnehmung des Benutzers nicht gestört wird.

Die untere Schwelle (90 ms) stimmt annähernd mit den Ergebnissen von *Steinmetz* überein. Im Vergleich zu den anderen Studien untersuchte er ein Gebiet, das zumindest technisch gesehen recht ähnlich zu dem des QED-Projekts ist. Leider unterlässt er Angaben über Versuchsbedingungen wie Bildschirmgröße oder Betrachtungsabstand. Ausserdem fehlen detaillierte Angaben über die Standardabweichung. Dies macht die Beurteilung der Bedeutung von *Steinmetz* Ergebnissen schwierig. Wie er jedoch selbst sagt, sind diese als „Ergebnisse einer ersten Stufe, die als allgemeine Richtlinie dienen sollen“ zu verstehen. Insofern bieten sie eine geeignete Basis für weitere Untersuchungen.

Wie bereits erwähnt, wird vermutet, dass Parameter der Versuchsanordnung wie Bildschirmgröße, Betrachtungsabstand und Bildauflösung bei der Bewertung von wahrgenommener Asynchronität eine wichtige Rolle spielen. Der grösste Einfluss auf die Wahrnehmung wird nach Ansicht des Autors vom Task<sup>1</sup> der jeweiligen Untersuchung ausgeübt. Soweit bisher ersichtlich, wurden in allen betrachteten Studien ähnliche Werte dieser Parameter verwendet: Bildgrößen von 23 bis 63 cm und Betrachtungsabstände von 50 bis 150 cm. Soweit dem Autor bekannt ist, sind entsprechende Studien mit kleineren Bildschirmen (wie beispielsweise solche von kommenden UMTS-Mobiltelefonen) bisher nicht durchgeführt worden; trotzdem geben *Steinmetz* Ergebnisse Hinweise darauf, dass kleinere Bilder (Unterschiede zwischen der Kopfansicht und der des ganzen Körpers) zu einer höheren Toleranz gegenüber Asynchronität von Bild und Ton führen.

---

<sup>1</sup> Task wird in dieser Arbeit verwendet für die Aufgabe, die Probanden bei Versuchen durchzuführen haben. Dabei eingeschlossen sind alle Parameter, die diese Aufgabe exakt beschreiben, wie Anzahl der Teilnehmer, Vorwissen, Muttersprache, Dauer des Experiments etc.

## 2.4 Untersuchungen zum Qualitätsempfinden bei Dialogsituationen

Während es bei vielen Anwendungen im Bereich der Kommunikation zwischen Mensch und Maschine (z.B. beim Browsen im Web) recht schwierig sein kann, das Benutzerverhalten vorherzusagen und zu modellieren, ist dies bei zwischenmenschlicher Kommunikation in Echtzeit<sup>1</sup> nicht der Fall. Hier erwarten die Benutzer vom Netzwerk ein Verhalten, das im wesentlichen die gleichen Merkmale besitzt wie die „Luft-Schnittstelle“ bei einem Gespräch, die alle Kommunikationskanäle zulässt wie Audio (man hört sich), Video (man sieht sich) sowie das Benutzen von gemeinsamen Hilfsmitteln und Informationen (Notizblock, Wandtafel, gemeinsam zugängliche Dokumente etc.). Letztendlich erwarten die Benutzer, dass die Telekommunikationsdienste eine Unterhaltung von Angesicht zu Angesicht nachbilden, inklusive der Übertragung aller wichtigen Aspekte der physikalischen Umgebung auf der Gegenseite (beispielsweise Hintergrundgeräusche). Darüber hinaus sind die zeitlichen Strukturen (z.B. Aktivitätsmuster der Gesprächsteilnehmer) in Verbindung mit einer typischen Situation oftmals denen eines Gesprächs von Angesicht zu Angesicht ähnlich. Bei einigen Parametern können sogar von vornherein bereits von „klassischen“ Telekommunikationsdiensten her bekannte Ergebnisse verwendet werden (beispielsweise ist die Gesprächsdauer bei einer Desktop-Videokonferenz vergleichbar lang mit der Dauer eines klassischen Telefongesprächs).

### 2.4.1 Absolute Verzögerung und ihre Auswirkung auf Echtzeit-Dialogkommunikation

Bei der Übertragung von Multimediatechniken von A nach B ist die Einfügung einer mehr oder weniger grossen Verzögerung unausweichlich. Angefangen von der physikalischen bedingten Signallaufzeit (ca. 300'000 km/s bei einer Funkstrecke und 200'000 km/s bei leitergebundener Übertragung) über die Dauer der jeweiligen Wandlung im Endgerät bis hin zur Kompression/Dekompression und der Paketisierungsverzögerung ergeben sich Gesamtverzögerungen, die beispielsweise im Internet (IP v4) keine garantierte obere Grenze besitzen, wohl aber z.B. in einem ATM-Netzwerk. Bei reinen Audioübertragungen kann diese Verzögerung mit der entsprechenden Wahl von Endgeräten und Übertragungstechniken bis hin zu kontinentalen Distanzen verhältnismässig klein gehalten werden (Referenz auf Tabelle). Aber allein schon durch das Hinzufü-

<sup>1</sup> Für derartige Situationen wird im folgenden der Begriff *Dialog* verwendet. Entgegen der landläufigen Meinung bezeichnet *Dialog* nicht nur ein Zweiergespräch, sondern auch Situationen mit mehr als zwei Kommunikationspartnern (siehe auch Glossar).

gen von Videodaten wird diese infolge der höheren Komplexität des Videosignals unweigerlich grösser.

Ein nicht zu vernachlässigendes Problem bei der Durchführung von Wahrnehmungstests betreffend Verzögerungen sind die beiden Effekte, dass die Versuchspersonen einerseits auftretende Verzögerungen fälschlicherweise dem Gesprächspartner zuweisen und nicht der technischen Übertragung, andererseits sich bei Erkennen der Verzögerung relativ rasch an die neue Situation anpassen, indem sie beispielsweise eine diszipliniertere Unterhaltung führen. Anhand der Abbildungen 2.3 und 2.4 lässt sich dieses Phänomen veranschaulichen.

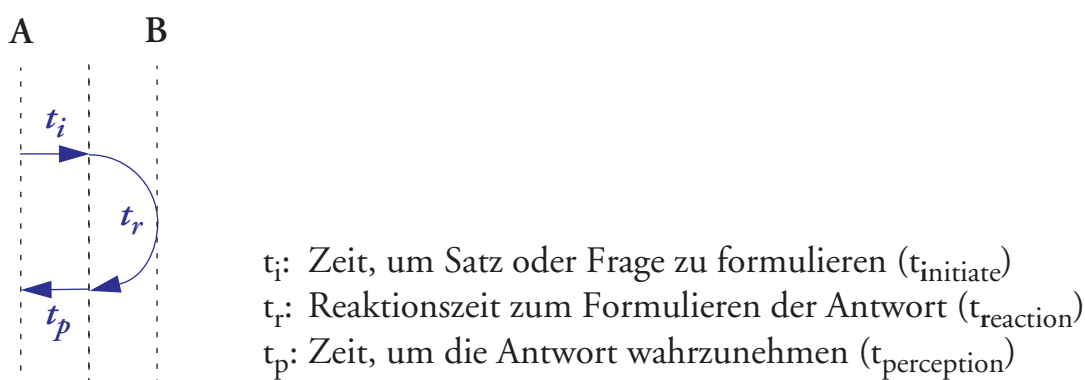


Abbildung 2.3: Gesamtverzögerung bei einem Gespräch von Angesicht zu Angesicht

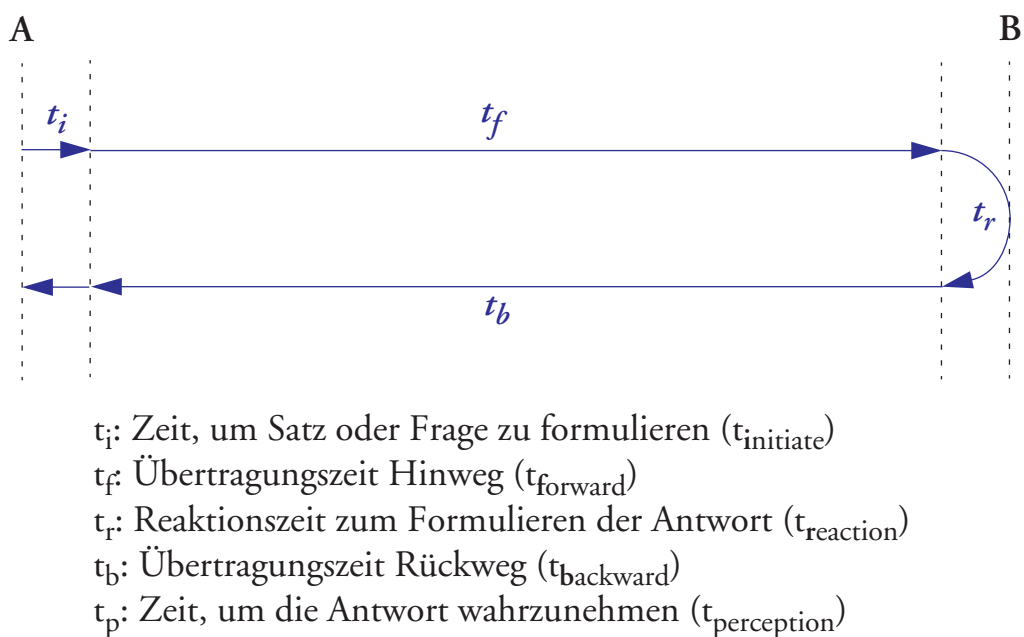


Abbildung 2.4: Gesamtverzögerung bei Videokonferenzen

Üblicherweise vergleichen Menschen eine neue Situation mit einer bereits bekannten. Im Falle einer Videokonferenz ist dies ein Gespräch von Angesicht zu Angesicht. Im letzteren Fall sind die Zeiten  $t_f$  und  $t_b$  beide Null, somit rechnet der unerfahrene Benutzer A bei einer Videokonferenz ebenfalls mit einer solchen Situation. Wenn nun die Antwort von B verzögert eintrifft, wird die Verzögerung fälschlicherweise der tatsächlichen Zeit  $t_r$  hinzugerechnet. Tabelle 2.2 Gibt eine Übersicht, wie sich die Zeiten  $t_f$  und  $t_b$  typischerweise zusammensetzen.

	$t_{\min}$ [ms]	$t_{\max}$ [ms]	$t_{\text{typ}}$ [ms]
Eingabe (Kamera)	25	35	30
A/D-Wandlung	<1	5	1
Komprimierung	<1	250	45
Paketisierung	0.5	25	10
Laufzeit	<1	100	40
Node-Processing	<1	300	1
Traffic-Shaping	0	50	10
Reassemblierung der Pakete	0.5	25	10
Ausgleichen des Jitters	<1	50	25
Dekomprimierung	<1	150	25
D/A-Wandlung	<1	5	1
Ausgabe (Bildschirm, Lautsprecher)	15	30	25
<b>Gesamte Einwegverzögerung</b>	<b>~41</b>	<b>1025</b>	<b>223</b>

*Tabelle 2.2: Typische Verzögerungsquellen bei Videokonferenzen (Bild)*

Bis heute wurden zu dieser „oberen Grenze“ der Verzögerung, die für multimediale Übertragungen noch „erlaubt“ ist (da sich der Benutzer bis dahin nicht gestört fühlt oder die Verzögerung gar nicht merkt), viele verschiedene Aussagen gemacht. Nachfolgend sind einige ausgewählte Ergebnisse aufgelistet, wie sie in der Literatur zu finden sind:

*Helder [37]*

Diese Veröffentlichung beschreibt eine der ersten systematischen Untersuchungen über absolute Verzögerung. Sie wurde durch die Satellitenübertragung mit der von Natur aus grösseren Verzögerung durch den längeren Übertragungsweg ausgelöst. Der experimentelle Ansatz ist recht ungewöhnlich, da „echte“ Telefonverbindungen über das transatlantische Kabelnetz verwendet wurden, bei denen ein künstlicher Verzögerungswert eingefügt wurde. Nach dem Gespräch schaltete sich ein Telefonist ein und befragte die Gesprächsteilnehmer nach der von ihnen wahrgenommenen Qualität. Dieser Ansatz kann aus zwei Gründen nicht vollständig überzeugen:

- Die experimentelle Umgebung erlaubt eine grosse Anzahl an möglichen Abläufen
- Die Auftraggeber (Bell Laboratories) dürften höchstwahrscheinlich nicht ganz unvoreingenommen gewesen sein, da sie sicherlich kein Interesse daran gehabt haben dürften, schwerwiegende Argumente gegen die Satellitenkommunikation herauszufinden.
- Helder verwendet zur Auswertung lediglich eine psychometrische Funktion, aus der ersichtlich ist, dass die ersten Untersuchungen einen Wert um 700 ms ergaben, die zweiten sind für eine Auswertung untauglich. Wahrscheinlich aus diesem Grund und weil der Schwerpunkt seiner Untersuchungen auf der Tauglichkeit neuer Filter lag, gibt er keine genauen Werte an.

*Yamaguchi, Wada, Yamamoto [100]*

In diesem Artikel wird ein System für die Videotelefonie beschrieben. Die Übertragungsrates beträgt 64 kbit/s (ISDN) und es wird behauptet, dass die Verzögerung des Audiosignals bedenklicher sei als die des Videosignals. Die durch die Benutzer noch tolerierbare Grenze wird mit 360 ms angegeben. Es werden keine weiteren Bewertungen für die Qualitätsgrenzen für Audio wie *nicht erkennbar, erkennbar aber nicht störend, störend* verwendet.

*Chen, Walrand, Messerschmitt [22]*

In dieser Veröffentlichung wird für Netzwerke mit paketisierter Übermittlung von Sprachinformation ein spezielles Protokoll vorgeschlagen, das die Ende-zu-Ende Verzögerung minimieren soll. Die Autoren postulieren, dass der Bereich von 100 bis 600 ms allgemein für die Übertragung von Audio wie in der Telefonie von den Benutzern akzeptiert werde.

*Gonsalves, Tobagi [35]*

Hier wird versucht, ein netzwerkunabhängiges Rahmenwerk für die Evaluation von Sprach/Daten-Netzwerken zu formulieren. Es wird ausgesagt, dass eine maximale Verzögerung von 100-200 ms innerhalb eines lokalen Netzwerks nicht überschritten werden soll.

*Scherer, Braun, Estermann [69]*

Subjektive Untersuchungen zeigten, dass es nicht möglich ist, für die Echtzeit-dialogkommunikation absolute und allgemeingültige maximale Verzögerungswerte für Audio und Video zu definieren. Die Benutzerzufriedenheit betreffend dieser Qualitätsparameter hängt noch von vielen weiteren Faktoren ab. In Situationen, in denen eine Person den Dialog dominiert und daraus praktisch einen Monolog macht, ist ein Verzögerungswert von 600 ms für Audio beispielsweise überhaupt kein Problem für die Benutzer. Im Vergleich zum selben Wert ist dieser in einer hitzigen Diskussion jedoch absolut störend und inakzeptabel. Weitere wichtige Faktoren für die Benutzerzufriedenheit sind u.a. Bildauflösung, Synchronisation zwischen Audio- und Videostreamen, die Erfahrung der Benutzer mit technischen Systemen, die Aufgabe, die die Benutzer zu lösen haben bzw. der Grund der Konferenz, die sozialen Beziehungen zwischen den Gesprächsteilnehmern, Zweck und Umfeld des eingesetzten Konferenzsystems (zu Hause oder im Geschäft), persönliche Empfindlichkeit der Benutzer auf bestimmte Qualitätsparameter usw.

Studie	QoS Parameter	Kritischer Wert für Ende-zu-Ende Verzögerung
Helder	Verzögerung Audio	700 ms
Yamaguchi, Wada, Yamamoto	Verzögerung Audio	360 ms
Chen, Walrand, Messerschmitt	Verzögerung Audio	600 ms
Gonsalves, Tobagi	Verzögerung Audio	100-200 ms max. 400 ms
Scherer, Braun, Estermann	Verzögerung Audio und Video	100...600 ms Situationsabhängig

*Tabelle 2.3: Maximale absolute Verzögerungswerte für Audio (und Video)*

Aus obestehender Tabelle ist ersichtlich, dass es keinen übereinstimmenden Wert für die absolut zulässige Verzögerung von Audio und Video bei Multimedialkommunikation gibt. Diese Tatsache wird bei praktisch keiner der referenzierten Studien in Betracht gezogen und es wird auch nicht beachtet, dass weitere Faktoren (Task, Versuchsaufbau etc.) eine Rolle spielen.

Nach Meinung des Autors sind die meisten der bisher gemachten und oben aufgeführten Untersuchungen auf dem Gebiet dieser Arbeit zu pauschal und zu wenig differenziert. Da vermutet wird, dass es keine fixe Wahrnehmungs- und Akzeptanzschwelle für Qualitätsparameter gibt, sondern dass diese von vielen Faktoren, insbesondere vom Task der Multimedialkommunikation, abhängen, wurde entschieden, eigene Versuche durchzuführen, die in ein konzeptionelles Rahmenwerk eingebunden sein sollen und eine Klassierung von Kommunikationsszenarien erlauben soll.

## *2.5 Schlussfolgerung*

Die Problematik nach der gerade noch tolerierbaren Verzögerung wird auf verschiedene Arten angegangen und in den meisten Fällen sind die Studien zu unterschiedlichen Ergebnissen gekommen. Betrachtet man diese Untersuchungen genauer, stellt man fest, dass die Testbedingungen äusserst unterschiedlich und die Aufgabenstellungen und Anforderungen an die Benutzer jedesmal verschieden waren. Die grosse Varianz der Resultate hat somit insbesondere zwei Hauptursachen: Die Umgebung (das Test-Setup) und der Task.

Für die Untersuchungen innerhalb dieser Arbeit bedeutete dies, dass das Test-Setup möglichst konstant gehalten und die Einflüsse der Tasks auf die Benutzeranforderungen hinsichtlich Verzögerungen genauer untersucht werden musste. Obwohl die Versuche unter Laborbedingungen durchgeführt wurden, wurde viel Wert auf eine natürliche Versuchsumgebung gelegt. Dies deshalb, damit die Ergebnisse möglichst ohne grosse Interpretation in die „echte Welt“ übertragen werden können.





# *Kapitel 3*

## *Trends in der Telekommunikation*

---

Die Vorhersage der zukünftigen Entwicklung von Telekommunikationsdiensten und deren Wechselwirkung mit der jeweils unterliegenden Netzwerkinfrastruktur ist eine der schwierigsten und spekulativsten Aufgaben für Trendforscher geworden. Dennoch wird nachfolgend versucht, einige denkbare Entwicklungen zu skizzieren. Die im Folgenden gemachten Aussagen erheben keinerlei Anspruch auf Vollständigkeit. Vielmehr werden einige exemplarische Trends beschrieben, um daraus Schlussfolgerungen hinsichtlich Qualitätsanforderungen der Benutzer zu ziehen. Es wird insbesondere vermutet, dass die zunehmende Vielfalt an Diensten, die vermehrten Möglichkeiten der gegenseitigen Substitution sowie die Erschliessung neuer Anwenderkreise (zum Beispiel Nutzung von Diensten, die früher nur im professionellen Bereich eingesetzt wurden, durch Kinder und Jugendliche) auch zu Änderungen bei den Qualitätserwartungen führen. Dazu kommen gänzlich neue Kommunikationsformen (vor allem multimodaler Art), für die früher durchgeführte subjektive Untersuchungen kaum mehr Gültigkeit haben.

### *3.1 Basistechnologien*

Die Leistung wichtiger Hardwaretechnologien für die Vermittlung (mikroelektronische und zukünftig auch optische Bauteile) und die Übertragung (wie Glasfasern, Signalverarbeitung und EHF-Elektronik) verbessert sich nach wie vor mit erstaunlichem Tempo, in der Mikroelektronik beispielhaft beschrieben durch das bekannte *Moorsche Gesetz*. In der Regel erlauben neue Basistechnologien, deren Entwicklung in führenden Forschungslaboratorien gerade aufgezeigt wird, die Vorhersage des Stands der Technik mit industrieller Massenproduktion in fünf bis fünfzehn Jahren. Somit kann mit grosser Sicher-

heit angenommen werden, dass die typischen Leistungssteigerungen der beiden vergangenen Jahrzehnte weitere zehn bis fünfzehn Jahre anhalten werden.

### 3.2 *Funktionalität und Mehrwert*

Diese Betrachtungsweise der Entwicklung wichtiger Hardwaretechnologien kann allerdings dazu verleiten, bei der Produkteentwicklung von einer konstanten Steigerung im Sinne von „kleiner, billiger, schneller, ...“ auszugehen. Der Autor ist jedoch der Ansicht, dass eine derartige „lineare“ Hochrechnung ein Trugschluss ist, hauptsächlich da die Funktionsweise moderner Produkte und die Werte, die Benutzer mit ihnen verbinden, immer häufiger von der *Software* abhängen. Der Begriff Software soll hier sehr weit gefasst verstanden werden und beispielsweise die Gerätesoftware von Mobiltelefonen und PDAs, die Überwachungssoftware von Routern, Switches und Netzwerk-Management-Systemen sowie alle Programme und Inhalte in Verbindung mit Mehrwertdiensten beinhalten. Des Weiteren sollte beachtet werden, dass Kommunikationsnetzwerke und die von ihnen zur Verfügung gestellten Basisdienste (z.B. Datentransport) nicht länger getrennt voneinander angesehen werden können. Seit die Technologie die Einbettung von leistungsfähigen Mikroprozessoren und Kommunikationsfunktionen in eine gewaltige Auswahl von bisher „unintelligenten“ Systemen und Geräten (im Sinne von *Pervasive Computing*) erlaubt, können daraus komplett neue Produktcharakteristiken entstehen, denen die Benutzer bei genügend tiefen Kosten zum „Durchbruch“ im Massenmarkt verhelfen.

Naturgemäß beginnen sich bei derartigen Entwicklungen bisher getrennte Geschäftsbereiche vermehrt zu überlappen (z.B. bieten Netzbetreiber zusätzlich auch Inhalte an), oder es entstehen neue Arten von Geschäftstätigkeiten (z.B. vertrauenswürdige Finanzintermediäre im elektronischen Handel). Letzten Endes können daraus im Unterschied zu heute grundlegend verschiedene Geschäfts- und Wirtschaftsmodelle entstehen, mit wesentlich neuen Aspekten, wie sie zum Beispiel erscheinen, wenn die Netzwerkkosten im Vergleich zu den gesamten Produktionskosten betrachtet oder im Rahmen der sogenannten TCO (total cost of ownership) gesehen werden.

### 3.3 *Transportdienste*

Eine der Schlüsselfragen in diesem Zusammenhang bezieht sich auf die Definition von *grundlegenden Transportdiensten* und deren Stellung im weltweiten Geschäftsumfeld, beispielhaft aufgezeigt durch die folgenden beiden Extremsituationen:

- (a) Die reine Übertragung von Rohdaten- oder Paketströmen kann als wenig gewinnbringendes, einheitliches Grundgeschäft angesehen werden, bei dem lediglich einige wenige weltweit agierenden Unternehmen mitmischen. Eine Variante dieses Modells kann sogar von einem Wiedererstarken von nationalen oder regionalen Betreibern ausgehen, falls das entsprechende Übertragungsnetz wie ein Strassen- oder Schienennetz angesehen wird, vergleichbar den in einem deregulierten Strommarkt aufkommenden sogenannten „Netzwerkfirmen“. Oberhalb des Grundübertragungsdienstes bieten eine breite Auswahl von untereinander im Wettbewerb stehenden Firmen Mehrwertdienste an, sowohl auf globaler als auch regionaler Ebene, wobei sie sich bei den Kosten, der Qualität, dem Inhalt und der Abhängigkeit voneinander abgrenzen.
- (b) Der Markt für Anwendungsdienste ist isoliert; beispielsweise existierten sehr spezielle Anwendungen mit einem Minimalbedarf an Datenaustausch. Die Anbieterfirmen dieser Dienste sind vertikal integriert, was bedeutet, dass sie den gesamten Protokollstapel abdecken und ihre eigenen Übertragungsnetze betreiben.

Am wahrscheinlichsten sieht der Autor das Entstehen einer Mischung aus beiden Modellen, wobei lokale oder regionale Übertragungsnetze wie unter (a) betrieben, und verschiedene untereinander im Wettbewerb stehende lokale oder regionale Portale zwischen Endverbraucher und Dienstanbieter vermitteln werden, während sie für ihren Übertragungsbedarf über längere Distanzen den harten Wettbewerb unter einigen wenigen weltweiten Netzbetreibern ausnutzen. Solch ein hybrides Modell weist Ähnlichkeiten mit der Luftfahrtindustrie auf.

### 3.4 *Einstellung und Verhalten der Benutzer*

Eine der gefährlichsten Fallen bei der Vorhersage zukünftiger technischer Entwicklungen ist die stillschweigende Annahme, dass sich grundlegendes menschliches Verhalten sowie menschliche Vorlieben im Laufe der Zeit nicht

änderten. Dies ist oft mit fest verwurzelten Vorstellungen über den Wert oder die Akzeptanz bestimmter Dienste oder Geräte verbunden, obwohl dazu völlig unterschiedliche Ansichten entstehen können, wenn neue Generationen ihre eigene „Weltsicht“ und ihre eigene Interpretation von Grundwerten wie Verantwortlichkeit des Einzelnen etc. entwickeln. Da die Herstellungskosten in der Massenproduktion in absehbarer Zukunft höchstwahrscheinlich weiter sinken werden (siehe oben), werden IT-basierte Anwendungen und Geräte vermehrt Gegenstand von Modetrends im Heimbereich werden. Insofern ist es durchaus denkbar, dass Anforderungen an die Lebensdauer, die Übertragungsqualität oder die benutzerfreundliche Bedienung hinter neuen Werten wie einer attraktiven Form oder bestimmten modischen Erscheinungsbildern zurückstehen werden. Dies kann z. B. bedeuten, dass die Benutzer eine eingeschränkte Qualität akzeptieren, da ihnen andere Merkmale wie modisches Design oder Mobilität wichtiger erscheinen. Solch eine Entwicklung kann bisherige QoS-Masstäbe ähnlich auf den Kopf stellen wie seinerzeit der unerwartete Erfolg von SMS, wo sich die Benutzer an die Texteingabe mit einem extrem unpraktischen Interface gewöhnt haben.

Nachfolgend werden im Hinblick auf die Diskussion von Qualitätsaspekten einige Trends in der Telekommunikation präsentiert, wie sie vom Autor eingeschätzt werden bzw. in der Literatur zu finden sind [34], [70], [78], [92], [93], [95], [96], [97], [98], [99]. Dabei ist zu beachten, dass diese Einschätzung in erster Linie für Industrieländer gilt.

### *3.5 Trends bei den Endgeräten*

Bei den Endgeräten sind zwei Haupttrends zu beobachten: Einerseits werden Funktionen von bisher mehreren Geräten zu neuen, noch leistungsfähigeren zusammengeführt; andererseits gibt es neue Geräte, die speziell in einem eng definierten Bereich eingesetzt werden, in diesem aber alle nur denkbaren Ansprüche erfüllen (beispielsweise ergonomische und miniaturisierte Bluetooth-Headsets). Namentlich folgende Entwicklungen sind zu erwarten:

- Zusammenwachsen von Unterhaltungselektronik und PC/Internet Technologie, besonders im Zusammenhang mit Diensten wie Video On Demand oder Interaktivität bei speziellen Fernsehsendungen.
- Verbindungen von Mobiltelefonen mit anderen Geräten wie PDAs, GPS-Empfängern, MP3-Playern, Digitalkameras usw.)

- Spezielle Videotelefone oder Videofone, wie sie früher immer wieder angekündigt wurden, wird es kaum geben. Videotelefonie im Heimbereich wird via PCs oder einem TV-Gerät realisiert werden, im Geschäftsbereich ebenfalls via PC (Einzelplatz) oder speziellen Multimediaräumen (für Sitzungen usw.). Ausserdem sind neue Endgeräte wie UMTS-Mobiltelefone geeignet, diese Technologie in einem neuen Umfeld sinnvoll einzusetzen.
- Mobiltelefone, die heute schon die Funktionalität einer Digitalkamera beherrschen sowie Standbilder oder kleine Videosequenzen per MMS übertragen können, werden mit neuen Technologien wie UMTS auch echte Videotelefonie erlauben. Dies nicht nur, wie im klassischen Sinne gedacht, mit der Darstellung des Gesprächspartners, sondern auch durch das Zeigen der Umgebung und anderen Personen oder Gegenständen.
- Endgeräte werden immer „intelligenter“, d.h. sie werden bis zu einem gewissen Grad Aufgaben selbständig übernehmen, die bisher von einem Benutzer zuerst gestartet werden mussten. Durch moderne Kommunikationsmöglichkeiten bleibt der Aktionsradius aber nicht wie bisher klein, sondern es ist eine sinnvolle Vernetzung mit anderen Geräten/Dienstleistungen möglich.

Bei den genannten Beispielen fällt auf, dass sowohl Endgeräte als auch Endbenutzer untereinander vermehrt kommunizieren. Ausserdem wird die Kommunikation multimedialer und multimodaler im Vergleich zu bisher. Dies bedeutet, dass sich die Benutzer nicht nur an neue Bedienkonzepte und kompliziertere Kostenstrukturen gewöhnen müssen, sondern auch dass sie mit neuartigen Formen der Multimediakommunikation konfrontiert werden. Bisher wissen weder die Gerätehersteller noch die Diensteanbieter, was ihre Kunden in Zukunft für Ansprüche hinsichtlich QoS haben werden. Sicher ist jedoch, dass diese mit bisherigen Modellen und Annahmen nicht nachgebildet werden können. Der Erfolg oder Misserfolg von neuen Applikationen wird auch wesentlich davon abhängen, ob die Benutzer Vertrauen in neue Formen von Multimediakommunikation haben werden oder nicht. Wenn es beispielsweise nicht gelingt, die Erwartungen der Benutzer hinsichtlich QoS zu erfüllen, werden sie dem Dienst sehr skeptisch gegenüberstehen. Ein möglicher Weg hin zu Modellen, die die QoS-Anforderungen aus Benutzersicht erfassen, sind Versuche mit Testpersonen, wie sie in Kapitel 6 beschrieben sind.

### *3.6 Trends bei den Endverbraucherdiensten im Heimbereich*

Im Heimbereich wird es vor allem Dienste geben, die für die Anbieter durch einen grossen Kreis an potentiell erreichbaren Kunden attraktiv erscheinen, ohne zu hohen Initialkosten pro Haushalt zu führen. Andererseits müssen zum Teil auch dort Dienste eingeführt werden, wo dies gerade nicht der Fall ist und bereits getätigte und zum Teil immense Investitionskosten amortisiert werden müssen. Ein Beispiel ist die UMTS-Technologie in Deutschland und anderen europäischen Staaten, wo Telekommunikationsunternehmen viel Geld für Lizenzen und den Aufbau von Netzen ausgegeben haben und immer noch ausgeben, bisher aber noch keine konkreten Anwendungen oder gar Endgeräte zur Verfügung stehen. Vermehrt werden sich auch Dienstleistungen aus dem Geschäftsbereich im Heimbereich durchsetzen, weil sie mit der Zeit billiger werden und durch die bei der geschäftlichen Nutzung gewonnenen Erfahrungen immer mehr an Attraktivität auch für die Freizeit gewinnen. Nachfolgend werden einige Beispiele aufgezählt:

- Vor allem im Bereich der Aus- und Weiterbildung werden Dienste wie „Distance Learning“ eingeführt werden. Dabei sitzt der Studierende oder Weiterzubildende zu Hause am PC und verfolgt Vorlesungen interaktiv am Monitor und kann über einen Rückkanal auch Fragen stellen. Aufgaben und Übungen werden online gelöst mit der Möglichkeit von Kontakten zu Assistenten via Videokonferenz oder Chat/Discussionboard.
- Die immer wieder angekündigte bzw. totgesagte Videotelefonie bzw. Videokonferenz wird in neuen Kontexten endlich sinnvoll eingeführt werden können. Es wird sich herausstellen, dass diese Technik durchaus sinnvoll anzuwenden ist, wenn sie nicht als Ersatz zu „echten“ Treffen und Gesprächen sondern als zusätzlicher Kommunikationskanal angesehen wird. Wie oben bereits erwähnt, wird es mit UMTS-Mobiltelefonen möglich sein, mit der eingebauten Kamera sowohl den Gesprächspartner zu zeigen (beispielsweise zu Beginn des Gesprächs), danach aber auch andere Ausschnitte wie die Ansicht eines Modells oder Geräts, eine Landschaft oder Umgebung, weitere Personen/Freunde etc. Somit beschränkt sich der Videokanal nicht auf eine mehr oder weniger statische Ansicht des Gesprächspartners, sondern es können sinnvoll Meinungen über etwas eingeholt werden, ohne dass sich die andere Person am selben Ort befinden muss. Beim Kauf eines Kleidungsstücks kann gefragt werden, ob die eine oder andere Farbe besser passe, bei der Buchung einer Reise kann der Partner befragt werden, ob ihm die Abbildung des Hotels A besser gefällt als die des Hotels B etc.

- Es werden in naher Zukunft auch Dienste im Bereich des *E-Government* angeboten werden. Diese reichen von der Bestellung von Pässen, Ausweisen, Urkunden etc. über das Abstimmen bei Wahlen bis hin zu Gerichtsverfahren, bei denen die Beteiligten nicht mehr in einem Gerichtssaal anwesend sein müssen.
- Ebenfalls durchsetzen werden sich sogenannte Agentendienstleistungen. Dabei werden Aufgaben nicht mehr komplett selbst erledigt, sondern ein intelligenter elektronischer Agent wird mit den entsprechenden Informationen versehen, bevor dieser sich selbständig an die Lösung der Aufgabe macht. Mit dieser Methode können z.B. Suchaufträge im Internet, Online-shopping von bestimmten Lebensmitteln oder das Programmieren des Videorecorders durchgeführt werden.
- Ein ganz neuer Markt wird durch Kleider mit eingebauten Computer und Kommunikationskomponenten (sogenannte *Wearables*) geschaffen. Dabei ist z.B. ein Mobiltelefon mit Lautsprecher und Mikrofon komplett in die Kleidung integriert, so dass keine Freisprecheinrichtung mehr nötig ist. Die Tastatur befindet sich beispielsweise auf dem Ärmel und das Display wird in eine Sonnenbrille integriert.
- Mit der Anbindung an schnellere Netze wird auch echtes Video On Demand möglich werden, bei dem Filme und Sendungen interaktiv bestellt und angesehen werden können. Dabei werden in eine Richtung primär Videobilder, in die andere vor allem Steuerbefehle übertragen.
- Ein ganz neuer Bereich eröffnet sich in der Telemedizin. So wird es z.B. nicht mehr nötig sein, bei bestimmten Beschwerden den Arzt persönlich aufzusuchen. Die Diagnose und Beratung erfolgt via Videokonferenz, das Rezept wird elektronisch auf das Mobiltelefon übertragen, das in der Apotheke als Abholnachweis dient, oder aber das Medikament wird direkt von einem Onlineshop versandt.
- Mit der Möglichkeit, praktisch überall Zugang zu einem Netzwerk zu haben, bieten sich auch für Dienste wie „Virtual Home Environment“ (VHE) neue Möglichkeiten. Dabei ermöglicht das VHE dem Benutzer, seine Wohnumgebung (soweit dies Geräte und Einrichtungen betrifft) überall und terminalunabhängig nutzen zu können. Somit können auch bei Abwesenheit kritische Ereignisse im Haus wie das Ausfallen des Tiefkühlers nach einem Stromausfall oder ein geplatzter Waschmaschinenschlauch festgestellt und entsprechende Massnahmen ergriffen werden. Sozusagen symmetrisch dazu existiert als ähnlicher Dienst das „Virtual Office“, indem die elektronische Arbeitsumgebung mit nach Hause oder in den Urlaub genommen werden kann.

Bei allen genannten Diensten spielt als wichtige Komponente die interpersonelle und multimodale Kommunikation vermehrt eine Rolle; sie kann sogar ein Schlüsselfaktor bei der Akzeptanz sein. Es werden beispielsweise Fragen auftauchen, wie ein Assistent eine oder mehrere Videokonferenzen in einer Fragestunde verwaltet, ob die gebotene QoS den Ansprüchen der Studierenden genügt oder ob beispielsweise das Bild zu schlecht oder die Verzögerungen zu gross für eine natürliche Diskussion sind. Sollte dies der Fall sein, kann es passieren, dass solche Konsultationen von zu Hause aus unterbleiben und auf eine andere Art und Weise kompensiert werden.

### *3.7 Trends bei den Endverbraucherdiensten im Geschäftsbereich*

Wie einleitend beschrieben, wird bei den Dienstleistungen im Geschäftsbereich eher auf Qualität, Zuverlässigkeit und Sicherheit Wert gelegt werden. Davon abgesehen ist aber eine genaue Abtrennung der beiden Bereiche nicht immer möglich. Es kann also durchaus sein, dass einige der unter 3.6 genannten Trends auch hier gelten bzw. umgekehrt.

- **Teleworking:** Mitarbeiter einer Firma oder Freischaffende haben keinen speziellen Arbeitsplatz mehr zur Verfügung, vielmehr teilen sie sich diesen mit anderen Mitarbeitern oder arbeiten zusätzlich noch von zu Hause aus. Dabei übernimmt eine breitbandige Anbindung an das Firmennetzwerk alle nötigen Kommunikationsaktivitäten wie Internet-/Intranetzzugang, Videoconferencing, Austausch von Mitteilungen unter den Mitarbeitern etc. Teleworker sollen sowohl firmenintern als auch rechtlich als gleichgestellte Mitarbeiter angesehen werden, denen keine Nachteile durch den fehlenden firmeninternen Arbeitsplatz entstehen.
- Echte Videokonferenzen, Videostreaming etc. werden sowohl für interne Weiterbildungen und Mitarbeiterschulungen als auch für Meetings bzw. Mitteilungen von Vorgesetzten an Untergebene verwendet werden. Es wird sich herausstellen müssen, inwiefern diese neuen Formen die bisherigen komplett ersetzen bzw. lediglich ergänzen können.
- Im Bereich der Medizin gibt es verschiedene Ansätze wie Telediagnostik, Telehistologie und auch Telechirurgie.
- Der vermehrte Einsatz von Techniken wie Spracherkennung und/oder künstlicher Intelligenz führt zu Dienstleistungen wie Callcentern, in denen nur noch teilweise Menschen arbeiten. Bei interaktiven Diensten wird auf



IVR- (Interactive Voice Response) oder multimodale Technologien gesetzt werden, da sich andere Methoden wie beispielsweise das Vorlesen einer Auswahl und Bestätigen durch Tastendruck am Telefon kaum durchsetzen dürften. Sinnvoll wäre beispielsweise eine Kombination eines sprachgesteuerten Agenten, der gleichzeitig via einem Display und nicht nur akustisch dem Benutzer Feedback geben kann.

Bei vielen wenn nicht allen der oben genannten Applikationen sowie auch denjenigen unter Abschnitt 3.6 spielt aus Sicht des Autors für die Akzeptanz das gegenseitige Vertrauen der beteiligten Personen und auch derer Vertrauen in die involvierte Hard- und Software sowie Daten eine grosse Rolle. Dieses Vertrauen kann sicher wesentlich gesteigert werden, wenn neuartige Applikationen einen direkten Kontakt in Echtzeit erlauben. Es sollte deshalb für jede Situation bekannt sein, welche QoS-Ansprüche die Benutzer haben, insbesondere wieviel Verzögerung die Benutzer bei interpersoneller Kommunikation mit Hilfe dieser neuartigen Applikationen noch tolerieren.

### *3.8 Schlussfolgerungen*

Bei einer genauen Betrachtung der Trends stellt sich heraus, dass die Dienstgüte eine grosse, wenn nicht sogar die grösste Rolle bei der Einführung von neuen Endgeräten und Dienstleistungen spielen wird. Insbesondere werden QoS-Fragestellungen dort relevant, wo eine multimodale Kommunikation zwischen zwei oder mehreren Personen oder auch zwischen Mensch und Maschine entsteht.

Geht man (wie in den durchgeführten Untersuchungen im Rahmen dieser Arbeit) davon aus, dass die Ansprüche an QoS in erster Linie vom Umfeld und den entsprechenden Tasks bestimmt werden, stellen sich die beiden folgenden Fragen:

- (1) Welches sind die repräsentativen Tasks, mit denen die Benutzer inskünftig konfrontiert werden?
- (2) Welche Qualitätsansprüche werden sie in den entsprechenden Situationen haben?

Wie sich schon in Abschnitt 2.4 herausgestellt hat, wird die absolute Verzögerung bei interpersoneller Multimediakommunikation für den wichtigsten

QoS-Parameter gehalten. Es gilt also konkret herauszufinden, welche Ansprüche die Benutzer bezüglich Verzögerungen bei repräsentativen Kommunikationsanwendungen stellen. Zudem ist es wünschenswert, als Ergebnisse von spezifischen Untersuchungen einen ganzen Katalog zu erhalten, aus dem diese Ansprüche abgeleitet werden können. Dahinter muss ein konzeptionelles Modell stehen, das die wichtigen Bereiche und Abhängigkeiten der Multimediatelekommunikation abdeckt. Nur mit diesem Ansatz ist letztlich ein qualitätsbasiertes Netzwerkmanagement möglich, das die Benutzeranforderungen befriedigt.

Da jeder Dienst auch etwas kostet, muss das verwendete Netz entsprechende Möglichkeiten zur Verfügung stellen, die Kosten entsprechend zu erfassen und dem jeweiligen Nutzer zuzuordnen. Darüberhinaus sind dynamische Situationen denkbar, in denen ein Dienst umso teurer wird, je mehr Nutzer diesen gleichzeitig anfordern oder aber je grösser die Auslastung des Netzwerkes ist. Bedenkt man, dass gleichzeitig noch Faktoren wie Mobilität, unterschiedliche Netzbetreiber oder sich ändernde Umgebungsbedingungen und Qualitätsanforderungen der Benutzer eine Rolle spielen, wird das konzeptionelle Modell, das hinter den Punkten stehen muss, sehr kompliziert.

# *Kapitel 4*

## *Taxonomie der Kommunikation*

---

### *4.1 Einleitung*

Als systematische Grundlage für spätere Diskussionen und Feststellungen wird in den Abschnitten 4.2 und 4.3 eine Taxonomie der Kommunikation vorgestellt. Während in Abschnitt 4.2 ein *funktionelles*, hierarchisch strukturiertes Modell präsentiert wird, nimmt der Abschnitt 4.3 mit der Analyse der *zeitlichen Strukturen* eine dazu orthogonale Sichtweise ein. Dazu kommen im Sinne der dritten Sichtweise noch die *räumlichen Strukturen*, zum Beispiel Betrachtungen über die Häufigkeit der Kommunikation in Abhängigkeit der örtlichen Position der Kommunikationsteilnehmer. Dieser dritte Aspekt ist aber für die Untersuchungen dieser Arbeit irrelevant und wird im folgenden ausgeklammert.

Zuerst soll an dieser Stelle jedoch genauer auf den Begriff *Telekommunikation* bzw. *Kommunikation* eingegangen werden. *Telekommunikation* bedeutet grundsätzlich nichts anderes als ein Informationsaustausch über Distanzen, die nicht mit den menschlichen artikulatorischen und sensorischen Organen überbrückt werden können und sich deshalb auf nachrichtentechnische Einrichtungen abstützen. Aus der Sicht eines Ingenieurs der Telekommunikation ist der Begriff Kommunikation traditionell auf einer technischen Ebene angesiedelt. Dabei geht es vor allem darum, Kommunikationshilfsmittel zur Verfügung zu stellen und diese im technischen Sinn laufend zu verbessern. Auf dem Gebiet der Übermittlungstechnik sind dies z.B. das Herstellen von schnellen, optischen Vermittlern oder die optimale Ausnutzung der theoretisch möglichen Bandbreite eines gegebenen Übertragungskanal.

Eine komplett andere Sicht der Kommunikation hingegen zeigt sich im Bereich der Philosophie, wo man sich unter anderem mit dem Idealtypus menschlicher Kommunikation auseinandersetzt. Beispielsweise widmete sich

Habermas in seiner Argumentationstheorie [36] der Frage, wie Kommunikation auf der Ebene von Sprechakten und den jeweils erhobenen Geltungsbedingungen aussieht.

Aus der Sicht des Autors, die er durch die Literaturrecherche (Kapitel 2) bestätigt sieht, hat sich Forschung und Entwicklung im Bereich Telekommunikation zu lange vorwiegend im traditionellen, technischen Bereich aufgehalten, ohne die in Abschnitt 1.1 erwähnten Veränderungen der Gesellschaft zu berücksichtigen – Veränderungen, die ihrerseits zum Teil ihre Ursache in der Technisierung des Umfeldes haben. Anstatt sich damit auseinanderzusetzen, wie die Kommunikationsbedürfnisse der Benutzer im Zeitalter von Globalisierung, Mobilität, Internet und Vernetzung wirklich aussehen, wird z. B. weiterhin daran gearbeitet, den Durchsatz von Übertragungs- und Vermittlungssystemen zu maximieren. Etwas überspitzt lässt sich sagen, dass eine bessere Übertragung im technischen Sinn nicht automatisch zu einer besseren Kommunikation im Sinne von Habermas führt.

Es stellt sich für den Autor deshalb die Frage, ob die Modelle, Methoden und Verfahren, mit denen vor 50 Jahren an diese Problematik herangegangen wurde, heute immer noch gültig sind, oder ob diese nicht wesentlich umfassender zu gestalten sind unter der Berücksichtigung neuer Faktoren und Einflüsse, die früher keine Rolle gespielt haben.

## 4.2 Einführung eines Modells

Als eine Art *Koordinatensystem*, das die präzise Beschreibung eines bestimmten Telekommunikations-Szenarios erlaubt, wird zuerst eine Klassifizierung (Taxonomie) der Kommunikation erstellt. Solche Beschreibungen unterstützen eine zielgerichtete Spezifikation der geplanten subjektiven Tests, und bei der Analyse der veröffentlichten Forschungsergebnisse sind sie ebenfalls nützlich.

Obschon im folgenden ein Modell mit übereinanderliegenden Funktionsschichten, ähnlich wie bei den weithin bekannten Netzwerk-Referenzmodellen (z.B. das 7-schichtige OSI Referenzmodell oder der 5-schichtige Internet Protokollstapel) verwendet wird, konzentriert sich der Autor auf die Anwendung per se (im folgenden auch als *Kommunikations-Szenario* bezeichnet). Die Modelle können als zusätzliche Schichten angesehen werden, die oberhalb der Anwendungsschicht der Netzwerk-Modelle angesiedelt werden können. Dieser Ansatz liegt in der Tatsache begründet, dass die Erwartungen der Benutzer an die Qualität zu einem grossen Teil vom Profil eines *Kommunikations-Szenarios*

(gegeben durch die Merkmale der eigentlichen Anwendung) abhängen. Während z.B. für eine bestimmte Sprachanwendung die Natürlichkeit der übertragenen Sprache und die Möglichkeit, den Sprecher zu erkennen, als alles überragendes Kriterium angesehen wird, kann für eine andere Anwendung gute Verständlichkeit zu niedrigen Kosten gefordert werden. Bei der Planung von subjektiven Experimenten ist es somit wichtig, den *Kontext der Anwendung* klar festzulegen sowie die Experimente entsprechend zu gestalten.<sup>1</sup>

Es gibt keine allgemein anerkannte Taxonomie für Kommunikation. Um die Untersuchungen durchzuführen, die für das QED-Projekt notwendig sind, *muss Kommunikation als mehrdimensionales Problem aufgefasst werden*. Die folgenden vier Abschnitte beschreiben die vier Aspekte Kodierung, Modalität, Orientierung und sozialer Kontext, welche eine passende Vorgehensweise für die Untersuchung von Kommunikationssituationen darstellen. Es wird gezeigt, dass diese vier Aspekte - mit einigen Erweiterungen - als Ebenen, wie in Abbildung 4.5 dargestellt - angeordnet werden können.

Kommunikation	Mögliche Attribute
Sozialer Kontext	formal, informal, ...
Orientierung	personenbezogen, nicht personenbezogen
Kodierung	verbal, nonverbal, ...
Modalität	visuell, auditiv, ...

**Abbildung 4.5:** Kommunikationsebenen

*Man beachte, dass die Angaben in den Kästchen typische Beispiele darstellen<sup>2</sup>*

Obwohl sich viele Kommunikations-Szenarien auf eine breite Auswahl von unterliegenden Netzwerklösungen abstützen können, schränken bestimmte technische Eigenschaften dieser Netzwerke die Palette der nützlichen Anwendungen ein oder begrenzen deren Leistungsverhalten oder Funktionalität. Zum

<sup>1</sup> Ein ähnlicher Ansatz wurde in [7] gewählt.

<sup>2</sup> Für die Beschreibung eines Kommunikations-Szenarios wird pro Ebene je ein Attribut verwendet. Theoretisch sind alle möglichen Kombinationen erlaubt, in der Praxis sind einige jedoch nicht sinnvoll.

Beispiel kann sich die Implementierung einer Telekonferenz wesentlich schwieriger gestalten, wenn die Vermittlungsschicht keine Multicast-Dienste zur Verfügung stellt; oder Besonderheiten bei der Wellenausbreitung (z.B. Eigenschaften des Übertragungsmediums unterhalb der physikalischen Schicht) können die Leistungsmerkmale eines mobilen Kommunikationssystems beeinflussen.

Das QED-Projekt befasst sich mit allen vier oben erwähnten Ebenen (Abbildung 4.5). *Höhere Ebene* bedeutet in diesem Zusammenhang, dass wir die Grenze der Betrachtung des Menschen als einfachen Reizempfänger überschreiten hin zu einer ganzheitlicheren Sichtweise, bei der die sogenannte *Kommunikationssituation* (auf der Ebene des sozialen Kontextes) eine ausserordentlich wichtige Rolle spielt. *Kommunikationssituation* beinhaltet die Beschreibung der Anzahl Teilnehmer in einer bestimmten Session, die Art der Teilnehmer (z.B. eine *Person* oder eine *Maschine*) und die Beschaffenheit ihres Informationsaustauschs (symmetrisch oder asymmetrisch). Typische Situationen reichen von einseitiger Punkt-zu-Punkt Kommunikation bis hin zu Szenarien von Mehrpunkt-Konferenzen; zusätzlich müssen Diskussionen zwischen Personen von Dialogen zwischen Mensch und Maschine unterschieden werden.

Wie bereits oben dargelegt, bestimmt unser in Ebenen angeordnetes Beschreibungsschema die Wahl des passenden Testverfahrens: Die beiden unteren Schichten können in erster Linie in Experimenten zur Einweg-Kommunikation untersucht werden, bei denen Methoden der klassischen Psychophysik zur Anwendung gelangen [4], [33]. Betrachtungen der oberen beiden Schichten führen gewöhnlich zu Experimenten mit Dialogszenarien (zweiweg oder mehrweg).

Auf den oberen Schichten sind mannigfache Situationen denkbar, die üblicherweise von der Art der Anwendung oder den von den Kommunikationspartnern zu bewältigenden Aufgaben abhängen. Somit muss letztlich sogar die Mentalität und der Charakter der beteiligten Personen bei der Planung der Experimente oder der Interpretation der Resultate in Betracht gezogen werden. Erstaunlicherweise hat sich herausgestellt, dass die Untersuchung der meisten Szenarien mit mehreren Teilnehmern (entsprechend einem Dialog oder Gespräch) ein Schritt in „terra incognita“ ist; beispielsweise existieren keine allgemein anerkannten Forschungsansätze, und in den meisten Fällen mangelt es an Verfahren, Klassifizierungen und sogar an sauberen Definitionen der betrachteten Entitäten. Dies trifft insbesondere auf die Psychophysik zu, in der traditionellerweise einerseits viele Fragestellungen zu *Einweg-Situationen*

erforscht wurden (wie z.B. Maskierungseffekte in der Akustik), während andererseits Untersuchungen zu Dialogszenarien äusserst selten zu finden sind. Nichtsdestotrotz ist der Autor davon überzeugt, dass die Untersuchung von Kommunikationsmustern im Zeitbereich als wichtiger Schlüssel dienen kann, um die Ergebnisse von subjektiven Tests auf entsprechende objektive Parameter abzubilden. In Kapitel 4.3 wird eine kurze Einleitung in diese Materie geben.

#### 4.2.1 Modalität

Die grundlegendste Voraussetzung für die Teilnahme an jeglicher Kommunikation ist das Empfinden und Wahrnehmen des übertragenen und Information beinhaltenden Signals. Dies schliesst die menschlichen Sinnesorgane ein, die fähig sind, Licht, Schall, Geruch, Geschmack und Berührungen wahrzunehmen, was jeweils einem entsprechenden Modus entspricht (oft wird alternativ auch der Ausdruck Wahrnehmungskanal verwendet). Obwohl zukünftige Entwicklungen in der Telekommunikation auch haptische (Tastsinn) und olfaktorische (Geruchssinn) Modi in einem speziellen Zusammenhang ansprechen könnten (z.B. in der Telemedizin), beschränken wir uns momentan auf den auditiven und/oder visuellen Kanal.

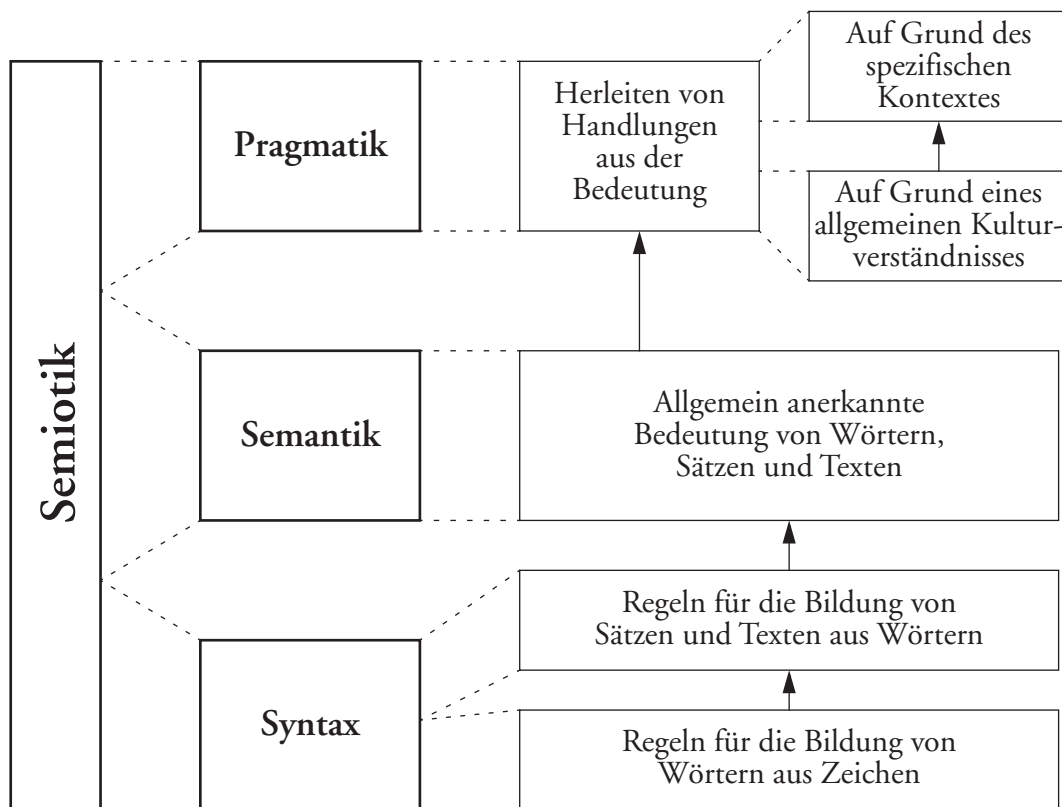
Wenn auditive und visuelle Kanäle gleichzeitig in ein Kommunikations-Szenario einbezogen sind, spricht man in der Regel von *multimedialer Kommunikation*. Vermehrt wird auch der alternative Ausdruck *multimodale Kommunikation* verwendet, wobei „multi“ nicht nur einfach „Ton und Bild“ bedeutet; vielmehr soll damit ausgedrückt werden, dass mehrere verschiedene Kommunikationsformen (in der Bedeutung von *Untermodi*) innerhalb des auditiven und/oder visuellen Kanals enthalten sein können. Beispielsweise ist der visuelle Kanal beteiligt, wenn das Videosignal ein Portrait der Kommunikationspartner zeigt, er ist aber auch mit von der Partie, wenn Texte und Graphiken im Rahmen eines sogenannten *Shared Workspace* ausgetauscht werden. Letztere Anwendung beinhaltet noch einen weiteren, unterstützenden Kommunikationsmodus in der Form eines separaten Kanals, der eine Maus oder einen Joystick gleichzeitig mit einem lokalen und übertragenen Zeiger koppelt. Diese Beispiele führen bereits schon zur zweiten Schicht unseres Kommunikationsmodells, da die Untermodi sich selbst durch verschiedene Arten der *Kodierung* unterscheiden.

### 4.2.2 Kodierung

Information kann auf viele verschiedene Arten kodiert und übertragen werden, und wie bereits im vorherigen Abschnitt angedeutet, können verschiedene Formen der Kommunikation mit speziellen Kodierungsarten simultan verwendet werden. Konzeptionell kann man sogar ein Sprachsignal in einen verbalen Teil und die sogenannte prosodische Information zerlegen, wobei prosodische Merkmale wie Tonhöhe, Klang, Lautstärke, Betonung und Artikulationsrhythmus Informationen beinhalten, die es erlauben, einen gesprochenen Satz richtig zu interpretieren (z.B. um eine Frage von einer Aussage zu unterscheiden), und den Sprecher zu identifizieren.

Eines der hervorstechendsten Merkmale auf der Ebene der Kodierung ist die Unterscheidung der Art der Information zwischen *verbal* und *non-verbal* wie im Folgenden beschrieben. Man spricht von *verbaler Kodierung*, falls geschriebene und/oder gesprochene Sprachen benutzt werden und falls Regelwerke (Grammatiken oder Lexika) existieren, durch die die Richtigkeit eines Textes oder einer Äußerung und deren Meinung klar festgelegt ist. Folglich kann die verbale Kodierung selbst auch wieder auf verschiedenen Ebenen betrachtet werden (Abbildung 4.6). Dabei sollte beachtet werden, dass es umso schwieriger wird, Grammatiken und Lexika als formale und allgemein anerkannte Basis zu erstellen, je höher man sich in den Ebenen aus Abbildung 4.6 bewegt. Die richtige Auslegung von textuellen bzw. sprachlichen Mitteilungen ist nämlich auch von Zusammenhängen auf pragmatischer und zum Teil auch auf semantischer Ebene abhängig, wobei „Zusammenhänge“ für Aspekte wie Orientierung und sozialem Kontext steht, wie sie in den Kapiteln 4.2.3 und 4.2.4 weiter unten besprochen werden.





*Abbildung 4.6: Auf Ebenen angeordnete Ansicht der verbalen Information<sup>1</sup>  
Man beachte, dass die Angaben in den Kästchen typische Beispiele darstellen*

Oft wird *non-verbale Kodierung* mit vielen verschiedenen Arten von bildhafter Darstellung in Verbindung gebracht (z.B. per Video übertragene Gesten, ergänzende Diagramme und Bilder zu einer mündlichen Präsentation, Piktogramme, Symbole etc.). Wie bereits beschrieben, verkörpern in einem Sprachsignal enthaltene prosodische Merkmale ebenfalls non-verbal kodierte Information. Des Weiteren sollte beachtet werden, dass diverse Formen von Hintergrundinformationen (sowohl auditiv als auch bildhaft) wichtig sind für den Kontext eines Kommunikationsszenarios. Beispielsweise wirkt ein Telefonanruf betreffend einer Zugsverspätung wesentlich glaubhafter, wenn der Angerufene im Hintergrund Bahnhofsgerausche vernimmt. Ein Spezialfall betrifft das Umgebungsgeräusch durch Gespräche anderer Personen, weil dessen Charakteristik der des eigentlichen Nutzsignals stark ähnelt. Somit ist es technisch sehr schwierig, das Nutzsignal vom Störsignal zu trennen (diese Problematik tritt beispielsweise ein, falls bei der Quellenkodierung mit niedrigen Bitraten aktiv/inaktiv-Muster (sogenannte *on/off-patterns*) ausgenutzt werden sollen).

<sup>1</sup> Diese Figur stellt eine prägnante Zusammenfassung [48] dar, wie sie z.B. in [25] zu finden ist.

Die Funktionsweise einiger sehr gut bekannter Methoden, die in subjektiven Tests von Sprachübertragungssystemen verwendet werden, kann direkt mit der Struktur in Abbildung 4.6 in Beziehung gebracht werden. Diese Tests befassen sich je nachdem mit der Verständlichkeit von isolierten Silben, Wörtern oder ganzen Sätzen. Offenkundig sind dabei isolierte Silben besonders heikel (Störungen führen hier zur höchsten Fehlerrate), da die Testpersonen in diesem Fall keine Kontextinformation hinzuziehen können, wenn es darum geht, das Gehörte „zu verstehen“. Es zeigt sich nämlich, dass die meisten Arten zwischenmenschlicher Kommunikation erhebliche Redundanzen aufweisen. Dies sollte jedoch nicht als ineffizient betrachtet werden; vielmehr ist dies ein wirkungsvoller Ansatz, Kommunikationssituationen mit starken Störungen erfolgreich zu meistern.

Ohne den Betrachtungen in Kapitel 5 vorgreifen zu wollen, soll bereits die Hypothese erwähnt werden, dass ein beträchtlicher Teil moderner Telekommunikations-Dienstleistungen nicht an Problemen leidet, bei denen Störungen herkömmlicher Art die Verständlichkeit beeinträchtigen. Vielmehr wird die Qualität anhand der Übertragungstreue non-verbaler Merkmale wie dem präzisen Erkennen des Gesichtsausdrucks oder der Natürlichkeit der Sprache gemessen; Parametern also, welche im weitesten Sinne Aufschluss geben über die „Stimmung“ des Gesprächspartners. Ausserdem wird auf das Vorhandensein oder Fehlen von Nebengeräuschen als einer Quelle von Information über den Kontext der Kommunikation Wert gelegt. Auf der anderen Seite kann bei mobilen Benutzern ein widersprüchliches Phänomen beobachtet werden: Viele Benutzer bewerten die Möglichkeit, praktisch überall kommunizieren zu können, mindestens als ebenso wichtig wie die herkömmlichen Qualitätsmerkmale des momentan genutzten Dienstes. Zwei Beispiele sollen diese Beobachtung veranschaulichen: Sehr oft wird bei Mobilfunkbenutzern eine Sprachqualität weit unter dem Niveau des gewohnten Telefonstandards toleriert; oder Versender von Kurzmitteilungen (SMS) geben sich mit einer extrem unhandlichen Benutzerschnittstelle zufrieden. Dies will allerdings nicht heissen, dass verbale Qualitätsmerkmale wie das Verstehen von geschriebener oder gesprochener Sprache keine Rolle mehr spielen würden!

### *4.2.3 Orientierung*

Wie oben bereits kurz angedeutet, beinhalten die höheren Ebenen unseres Modells einen breitgefächerten Merkmalsbereich. Damit einhergehend haben sich aus den Studien von Aspekten der höheren Kommunikationsebenen viele verschiedene Theorien entwickelt. Im folgenden werden einige dieser Konzepte

kurz angeschnitten. Dabei ist zu beachten, dass die meisten Ansätze für Kommunikation im Geschäftsbereich festgelegt wurden, wahrscheinlich deshalb, weil dieser Bereich als besonders einflussreich angesehen wird, wenn es darum geht, Qualitätsanforderungen zu etablieren.

Die Ebene der *Orientierung* (Abb. 4.5) ist stark mit der *Pragmatik-Ebene* aus Abbildung 4.6 verwandt, da sie gewissermassen die Zielsetzung eines Kommunikationsszenario beschreibt sowie die damit verbundene Auffassung, die die Teilnehmer über die taktischen Voraussetzungen haben sollten, die nötig sind, um die zu behandelnde Thematik zu verstehen. In einigen Fällen könnten sich diese Voraussetzungen in einem recht eng definierten Bereich befinden (wenn sie beispielsweise besondere technische Kenntnisse beinhalten, die es braucht, um eine bestimmte Aufgabe zu lösen), während in anderen Fällen eher eine gemeinsame Weltanschauung für eine erfolgreiche Diskussion notwendig sein könnte.

Der untenstehende Abriss ist notwendigerweise unvollständig und sollte als Indiz dafür angesehen werden, dass es überaus wichtig ist, einen bestimmten Versuchsaufbau genau zu definieren, u.a. in dem man die Terminologien und die gewonnen Erkenntnisse aus den genannten Theorien verwendet.

#### 4.2.3.1 Die Kategorien von Bales

Bereits in der 50er Jahren führte Bales [5] eine Reihe von Experimenten durch, in denen Testpersonen „künstliche“ Meetings abhielten. Er untersuchte die Beschaffenheit der stattfindenden Wechselwirkungen. Aus diesen Experimenten erarbeitete er vier Hauptkategorien: *positive Reaktionen*, *negative Reaktionen*, *Problemlösungsansätze* und *Fragen*. In der CSG (Communication Studies Group) reduzierte Short [74] diese vier Kategorien auf zwei: Die *positiven und negativen Reaktionen* von Bales wurden als *personenbezogen* klassifiziert, die beiden anderen Kategorien *problemlösende Bestrebungen* und *Fragen* als *nicht personenbezogen*. Die CSG sieht die personenbezogenen Reaktionen als entscheidende Kriterien für das Verständnis der Kommunikation via Telekonferenz an.

#### 4.2.3.2 Der SYMLOG-Raum

Viele von Bales et al. [6] geleitete Untersuchung weisen darauf hin, dass mindestens drei bipolare Grundcharakteristiken notwendig sind, um die Kommunikation in kleinen Gruppen zu beschreiben. Diese drei Dimensionen erzeugen

den sogenannten SYMLOG-Raum (Systematic Multiple Level Observation of Groups):

*Dominanz* gegenüber *Unterwürfigkeit*,

*Freundlichkeit* gegenüber *Unfreundlichkeit*,

*Akzeptanz* gegenüber *Ablehnung von Autorität*.

Diese Charakteristiken wurden in standardisierte Fragebögen eingearbeitet und häufig in verschiedenen Kulturen angewandt. Dadurch sollen sie genügend zuverlässig sein, höhere Ebenen der Kommunikation zu charakterisieren.

#### 4.2.3.3 *Der Unterschied zwischen Inhalt und Beziehung*

Watzlawick [88], [89] unterscheidet zwischen dem *inhaltlichen* Kontext einer Nachricht und demjenigen, der den *Beziehungskontext* enthält, wodurch er eine direkte Verbindung zu den Begriffen *Bericht* und *Befehl* herstellt, die durch Bateson [68] eingeführt wurden. In [89] weist Watzlawick treffend auf die Übereinstimmung dieser Begriffe mit den in der Informatik verwendeten Ausdrücken *Daten* und *Steuerung (control)* hin. Da Steuerinformation aussagt, was mit den vorliegenden Daten geschehen soll, kann sie als „Information über die Information“<sup>1</sup> (oder *Metainformation*) betrachtet werden. Diese Erkenntnis wird durch das folgende Axiom [89] beschrieben: „Jegliche Kommunikation besteht aus einem inhaltlichen und einem beziehungs-mässigen Aspekt, so dass der letztere den ersteren bestimmt und somit eine Metakommunikation darstellt“. Bei der zwischenmenschlichen Kommunikation kann der Austausch von Kontrollinformation als „Übergabe einer Vorschrift“ betrachtet werden, die von den Kommunikationspartnern anschliessend umgesetzt werden soll. In dieser Hinsicht sind sich die Ansichten von Watzlawick und die von Bales erstellten Kategorien ziemlich ähnlich und wir können unsere Taxonomie möglicherweise dahingehend vereinfachen, indem wir einen sich an einer Aufgabe orientierenden Ansatz als *nicht personenbezogen* verstehen, wohingegen Beziehungen als *personenbezogen* angesehen würden.

---

<sup>1</sup> Es wird hier eine Sicht eingenommen, wie sie in der Informatik seit langem gepflegt wird: Mit *Information* werden nicht nur Sachverhalte beschrieben (Information im engeren Sinne), sondern auch Handlungsanweisungen und Regeln (herkömmlich z. B. ein Kochrezept, in der Informatik ein Programm sowohl in Quellform und auch als ausführbarer Code).

## 4.2.4 Sozialer Kontext

### 4.2.4.1 Einleitung

Eines der bekannten Experimente in der Physik bei der Behandlung der Fluid-dynamik ist das der sogenannten *kommunizierenden Röhren*: Sind zwei Behälter beliebiger Form an ihrem tiefsten Punkt mit einer Röhre miteinander verbunden, wandert eine Flüssigkeit, die in einen der Behälter gefüllt wird, durch die Verbindung in den anderen und nach einiger Zeit (in Abhängigkeit der Visko-sität der Flüssigkeit) erreicht sie in beiden Behältern den gleichen Stand. Ver-gleicht man die Kommunikation mit diesem Phänomen, kann man sagen, dass sie ein Vorgang ist, bei dem zwei (oder mehrere) Teilnehmer durch den Aus-tausch von Information und Metainformation (siehe 4.2.3.3) versuchen, einen Informationsgleichstand zu erreichen (unter Berücksichtigung der Randbedin-gungen, die durch den spezifischen Kontext gegeben sind)<sup>1</sup>.

Während in der Fluidodynamik der Rückkopplungsmechanismus für das Errei-chen eines statischen Gleichgewichtszustandes vor einem makroskopischen Betrachter verborgen bleibt, sind Rückmeldungen (*Feedback*) in der Kommu-nikation unverzichtbare Bestandteile. Daher hängt die Art und Weise unseres Verhaltens stark von den Rückmeldungsmöglichkeiten ab, die ein Kommuni-kations-Szenario erlaubt. So bestehen beispielsweise selbst für sogenannte *Ver-teilmitteln* wie Hörfunk oder Fernsehen Rückmeldekanäle in Form von Beschwerden via Telefon, Briefen an die Redaktion oder indirekt, indem der Abonnent seinen Vertrag kündigt. In den genannten Fällen sind Rückkanäle eher willkürlicher Art; es können grosse Verzögerungen auftreten, und norma-lerweise ist ihre Bandbreite begrenzt. In einer solch asymmetrischen Situation wird der Informationsanbieter wesentlich vorsichtiger in seiner Wahl des Aus-drucks sein, wohingegen zwei direkte Gesprächspartner mit unverzüglichen Feedbackmöglichkeiten auf eine rasche Einigung hoffen dürfen, auch wenn sie sich spontan (und damit unter Umständen ungenauer) äussern.

In den nächsten Abschnitten wird der soziale Kontext hinsichtlich folgender beider Aspekte beschrieben:

- *Quantifizierbarkeit*, z.B. in Bezug auf die Anzahl der Teilnehmer und die Symmetrie der verbindenden Kanäle.

---

<sup>1</sup> Es geht bei diesem Vergleich nicht darum, dass die Information während des Ausgleichs gesamt-haft zunimmt, sondern darum, dass in beiden Situationen ein Gleichgewichtszustand entsteht, was in beiden Fällen nur durch Rückkopplungsmechanismen ermöglicht wird.

- *Grad der Formalität*, z.B. in Abhängigkeit vom Grad der Existenz formeller Regeln oder einer Vorschrift für den Informationsaustausch.

#### 4.2.4.2 *Kommunikationssituation*

*Kommunikationssituation* im Sinne dieser Arbeit beinhaltet die Beschreibung der *Anzahl* Teilnehmer in einem bestimmten *Szenario*, den Typ der Teilnehmer (z.B. *Mensch* oder *Maschine*) sowie die Art (*asymmetrisch* oder *symmetrisch*) ihres Informationsaustauschs. Typische Situationen reichen von *unidirektionaler Punkt-zu-Punkt Kommunikation* bis hin zu *Mehrpunkt-Konferenzen*; dabei sind *Unterhaltungen zwischen Personen* und *Dialog zwischen Mensch und Maschine* auseinanderzuhalten.

#### 4.2.4.3 *Formelle und informelle Kommunikation*

Die Unterscheidung von *formeller* und *informeller* Kommunikation wird von verschiedenen Theoretikern vorgenommen. Smith [76] definiert formelle Kommunikationskanäle als „aus offiziellen Quellen stammend und offizielle Bestimmungen übertragend... Normalerweise bewegen sich formelle Mitteilungen innerhalb solcher Kanäle, womit diese Rechtmässigkeit und Glaubwürdigkeit erlangen.“ Andererseits sind informelle Kommunikationskanäle „nicht rational festgelegt. Sie entstehen durch Zufälle in der räumlichen Ordnung, durch Freundschaften...“ Sowohl formelle als auch informelle Kommunikation kann einem ansteigenden, abfallenden oder ebenen Pfad folgen (zu höherer, niedrigerer oder gleichwertiger Autorität).

Wiederum zeigte es sich, dass die vorhandene Literatur fast ausschliesslich auf geschäftliche Formen der Kommunikation eintritt. Als weitere Unterscheidung kann geschäftliche Kommunikation in *programmierte*, *geplante* und *orientierende Tätigkeiten* unterteilt werden:

Thorngren [84] unterscheidet drei geschäftliche Kommunikationsformen. *Programmierte* Kommunikationstätigkeiten sind routiniert, standardisiert und wiederholbar wie beispielsweise das Geben oder Erhalten von Befehlen oder das Ausführen von Käufen bzw. Verkäufen. Die betroffenen Personen sind für gewöhnlich gut miteinander vertraut. Im Gegensatz dazu sind *orientierende* Kommunikationstätigkeiten neuartig, unstrukturiert und komplex, beispielsweise das Erschliessen eines neuen Marktes oder das Entwickeln von Ideen für neue Produkte. Die beteiligten Personen hatten normalerweise noch nie miteinander zu tun. Zwischen diesen beiden Extremfällen liegen die *geplanten*

Tätigkeiten. Sie führen zur Entstehung und Umsetzung von Alternativen, die ihrerseits wieder Ergebnisse der *orientierenden* Tätigkeiten darstellen.

Weitere Theoretiker verwenden ähnliche Kategorien: Ansoff [2] unterscheidet zwischen strategischen, administrativen oder operativen Handlungen; das Konzept von Simon [75] besteht aus Intelligenz, Design und Auswahl.

### 4.3 *Zeitliche Kommunikationsmuster*

Selbstverständlich wurde schon vor recht langer Zeit festgestellt, dass Kommunikationsprozesse bei genügend langer Betrachtung charakteristische Strukturen erkennen lassen. Folgt man Betrachtungen von Watzlawick [89], kann ein Kommunikationsszenario in mehrere Interaktionen, und jede davon weiterhin in einzelne Nachrichten unterteilt werden. Obwohl Watzlawick letztere eine „einzelne Kommunikationseinheit“ nennt - er betrachtet die Nachricht als eine Art Atom auf seiner Beobachtungsebene (Pragmatik) - wissen wir, dass beispielsweise mündliche Nachrichten weiter aufgeteilt werden können in kleinere Einheiten wie Äusserungen und Phoneme, schriftliche in Sätze, Wörter und Buchstaben; in beiden Fällen gelten statistische und grammatikalische Regeln (in diesem Zusammenhang ist die Bedeutung von Abschnitt 4.2.2 und speziell auch Abbildung 4.6 zu beachten).

Des weiteren wird bei gesprochenen Nachrichten nicht nur beim aktiven Sprechen (wenn der Sprecher „Nachrichten-“ oder „Informations-Atome“ produziert, die Vokalen, Zisch- und Explosivlauten entsprechen) Information übertragen, da die Dauer der Pausen prosodische Information enthält (sowohl innerhalb als auch zwischen Sprechphasen, siehe Abschnitt 4.2.2). Diese kann z.B. bei der richtigen Interpretation der Interpunktion oder beim Herausfiltern von Merkmalen, die der Erkennung des Sprechers dienen, weiterhelfen. Anders ausgedrückt könnten wir Watzlawick zitieren [[89], S. 49]: „Aktivität oder Inaktivität, Worte oder Stille, alles hat einen Wert als Nachricht...“.

In Abschnitt 4.2.4.1 wurde bereits auf die Wichtigkeit von Feedback in der Kommunikation hingewiesen. Bei der Sprachübertragung besitzen Rückmeldungen generell eine grosse Bedeutung, dazu gehören insbesondere auch kurze (verbale oder non-verbale) Äusserungen durch den „Zuhörer“ B, währenddem A gerade spricht. Für diese im Englischen als *doubletalk* bezeichnete Situation gibt es demnach zwei Gründe:

- die eben beschriebenen erwünschten Rückmeldungen („ich höre immer noch zu“)
- das in der Regel weniger erwünschte „sich ins Wort fallen“.

Nicht unerwartet untermauerten bereits erste Versuchsreihen [69] die These, dass zeitliche Muster als wichtiger Schlüssel in den Untersuchungen der vorliegenden Arbeit dienen könnten. Diese Tests wurden mit Studierenden durchgeführt, die überhaupt nicht mit modernen CSCW-Werkzeugen vertraut waren. Als Folge davon waren sie alleine schon von der Aufgabe, eine hoch entwickelte Multimedia Workstation zu bedienen, derart eingeschüchtert, dass sie sich sehr diszipliniert unterhalten haben. Diszipliniert bedeutet hier, dass sehr geordnete Gesprächsabläufe stattgefunden haben in der Art von „A spricht, B hört zu“, gefolgt von „B spricht, A hört zu“, jeweils mit ausreichend langen Pausen dazwischen um sicherzugehen, dass eine Äusserung auch wirklich beendet war. Auf den ersten Blick überraschend, bei genauerer Betrachtung der Experimente jedoch folgerichtig erklärbar erscheint die Erkenntnis, dass bei dieser Sachlage selbst grössere Verzögerungen von wesentlich mehr als 100 ms als nicht störend empfunden wurden! Offensichtlich liegt der Grund dafür in der Tatsache, dass die Verzögerungen keine ernsthaften Auswirkungen auf die zeitlichen Muster dieser Art von „disziplinierter Unterhaltung“ hatten, da die Werte gewisser Zeitkonstanten oberhalb der Verzögerungswerte lagen.

Der Schlüssel zu einem systematischeren Ansatz kann also durchaus in einer zweckmässigen Modellierung der zeitlichen Muster liegen, die mit einem Gespräch verbunden sind, sowie in der Schaffung eines Bezugs zwischen entsprechenden Parametern und bestimmten Anwendungstypen. Es ist dann z. B. möglich, aufgrund der Messung dieser Parameter Rückschlüsse auf die noch tolerierten Verzögerungswerte zu ziehen, mit anderen Worten sind in diesem Fall nicht mehr für jeden Anwendungstyp spezifische subjektive Versuchsreihen notwendig.

Es hat sich bereits herausgestellt, dass die bisher bekannten Modelle in ihrer Genauigkeit selbst für reine Sprachdialoge nicht ausreichen, die zu untersuchenden Belange vollständig abzudecken. In [57] wird beispielsweise ein Modell vorgestellt, das eine Unterhaltung in den folgenden vier Zuständen sieht: „A und B still“, „A spricht, B still“, „B spricht und A still“ und „A und B reden beide“. Dieses Modell wurde zur Analyse von DSI-Systemen<sup>1</sup> entwickelt (mit dem Ziel, den Gewinn durch statistisches Multiplexieren vorauszusagen).

---

<sup>1</sup> Digital Speech Interpolation



Dennoch werden nur Durchschnittswerte angegeben, unter anderem für die Dauer einer einzelnen Sprecherphase oder für Wahrscheinlichkeit und Dauer von sogenannten Double-Talk-Intervallen. Der Autor ist der Meinung, dass detailliertere Modelle notwendig sind, die zum Beispiel beim Sprechverhalten folgende Zustände (Tabelle 4.4) unterscheiden:<sup>1</sup>

Zustand	Sprecher A	Sprecher B
1	still	still
2	Quasiverbale Äusserungen	still
3	sprechend	still
4	sprechend	sprechend
5	still	Quasiverbale Äusserungen
6	still	sprechend
7	Quasiverbale Äusserungen	Quasiverbale Äusserungen

*Tabelle 4.4: Zustandsmodell des Sprechverhaltens (Dialogsituation)*

Quasiverbale Äusserungen spielen beim Führen von natürlichen Unterhaltungen (von Angesicht zu Angesicht) vermutlich eine wichtige Rolle hinsichtlich ihrer Feedback gebenden Funktion von „A“ nach „B“, während „B“ gerade spricht. Es sollte möglich sein, solche Dialoge mit stochastischen Prozessen zu modellieren, indem Übergangswahrscheinlichkeiten (oder Frequenzen) sowie Werte für die Dauer charakteristischer Intervalle (wie durchschnittliche Dauer einer Pause  $T_0$  oder durchschnittliche Dauer von Double-Talk  $T_4$ ) verwendet werden. Eine graphische Darstellung dieser Situation ist in Abbildung 4.7 unten dargestellt.

Mit diesem Ansatz sind zwei Schlüsselprobleme verbunden:

- Es sollte darauf hingearbeitet werden zu zeigen, dass eine geringe Anzahl von Parametern ausreicht, um Kommunikationssituationen ausreichend zu beschreiben, wie sie im „echten“ Leben vorkommen.
- Im Hinblick auf Online-Monitoring sollten Möglichkeiten gefunden werden, die oben genannten Parameter durch Analyse der einzelnen Signale (wie Sprache, Gesichtsausdruck, Handlungen zur Kontrolle eines fernübermittelten Zeigers etc.) zu messen. Zu diesem Zweck mag es einiges an intelligenter Verarbeitung benötigen, beispielsweise um das „Sprechen“ vom

<sup>1</sup> Es sollte beachtet werden, dass diese Beschreibung streng genommen nur dann sinnvoll ist, wenn keine Verzögerung vorhanden ist.

„Produzieren quasiverbaler Äusserungen“ unterscheiden und eine korrekte Segmentierung garantieren zu können, oder um aussagekräftige Ereignisse aus einem Videosignal abzuleiten, das den Gesichtsausdruck beinhaltet.

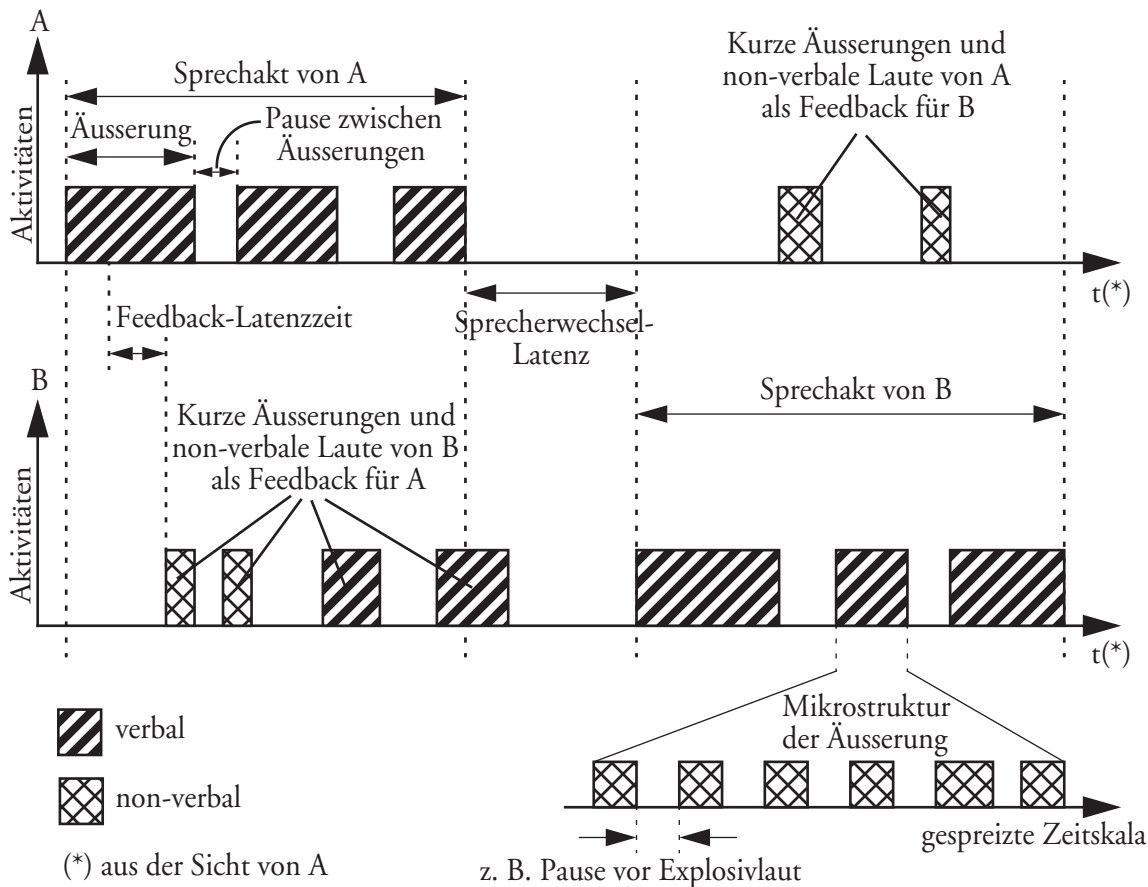


Abbildung 4.7: Dialogmuster

#### 4.4 Schlussfolgerungen

Die in diesem Kapitel angestellten Betrachtungen und die vorgestellten Modelle sollen ermöglichen, Kommunikationssituationen in verschiedene Kategorien zu unterteilen. Solche Kategorien fassen diejenigen sogenannten *Szenarien* oder *Setups* zusammen, die gemessen an ihren Anforderungen an bestimmte Qualitätsparameter als gleichwertig angesehen werden können. Somit können bei der Planung eines Netzwerks unter Berücksichtigung der entsprechenden Kategorie schon im voraus wichtige Parameter derart festgelegt werden, dass zukünftige Benutzer mit der QoS zufrieden sind: Qualitätsbasiertes Netzwerkengineering wird möglich. Dafür wird es jedoch notwendig sein, nicht nur einen Qualitätsparameter (in dieser Arbeit die Zweiwegverzögerung) als unabhängige Variable in einem umfassenden Modell zu betrachten, sondern in weiteren Untersuchungen andere Faktoren zu untersuchen, um so ein vollständiges Modell zu erhalten.

# *Kapitel 5*

## *Vorgehen und Arbeitshypothesen*

---

Nachdem in den vorigen Kapiteln vor allem allgemeine Betrachtungen zum Thema angestellt und Grundlagen geschaffen wurden, wird nachfolgend das Vorgehen im spezifischen Forschungsprojekt der vorliegenden Arbeit vorgestellt. Zum einen müssen aus Modellierungsgründen Vereinfachungen vorgenommen werden, vor allem aber ist es nicht möglich, den kompletten Raum aller denkbaren Kommunikationssituationen innerhalb einer Dissertation zu untersuchen. Zuerst wird deshalb der Bereich der möglichen Anwendungen definiert. Anschliessend wird auf die Dienstgüte-Parameter (QoS-Parameter) sowie deren genaue Definition eingegangen und die Auswahl eingeschränkt. Nachdem die Vergleichbarkeit der Studien (Kapitel 6) mit anderen Untersuchungen sichergestellt wird, werden die Anforderungen zusammengestellt, die an die Tasks (Aufgaben) für die subjektiven Versuche gestellt werden. Im Anschluss wird auf die Variablen eingegangen, die bei den Tests berücksichtigt werden, und es werden Konzepte und Modelle vorgestellt, wie sie andere Forschungsgruppen verwenden. Bevor schliesslich auf bestimmte Arbeitshypothesen eingegangen wird (Abschnitt 5.3), wird das verwendete Testverfahren genauer erläutert.

### *5.1 Abgrenzen des Forschungsgebiets*

Wie in Abschnitt 1.2 erläutert, beschränkt sich diese Arbeit auf die Betrachtung von Videokonferenzsituationen. Dabei werden sowohl geschäftliche als auch private Einsatzmöglichkeiten berücksichtigt. Das zu untersuchende Gebiet kann, wie in Abb. 5.8 gezeigt, dargestellt werden.

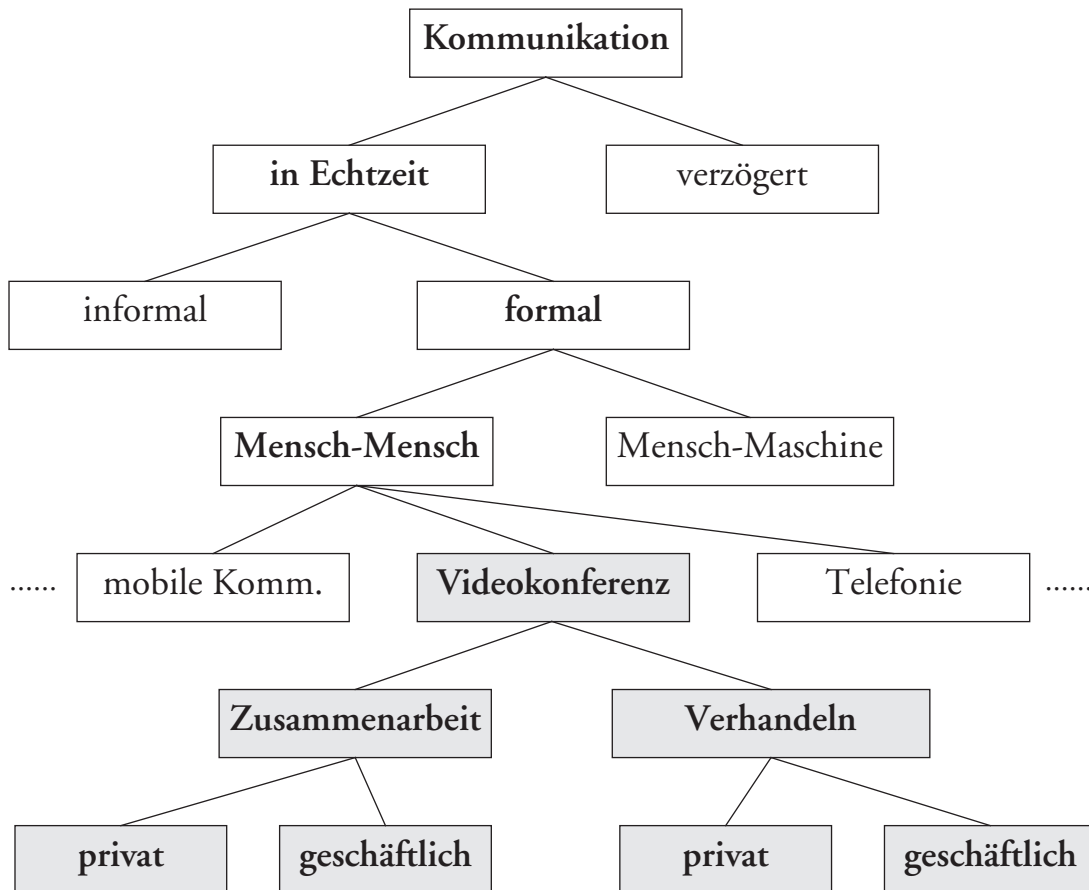


Abbildung 5.8: Eingrenzen des Gebiets bezüglich der Applikationen

Betrachtet man Videotelefonie oder Videokonferenzen nicht als Ersatz für ein Gespräch von Angesicht zu Angesicht, sind verschiedene Ansätze denkbar und teilweise auch in der Literatur diskutiert worden.

- Die *Hypothese der non-verbalen Kommunikation* sieht beispielsweise den Videokanal als Unterstützung zum reinen Audiokanal an und sieht in ihm einen Vermittler von Mehrwert (Darstellung des Gesprächspartners und seiner nonverbalen Kommunikation wie Gesichtsausdruck, Gesten etc.) [42], [45]. Speziell für diesen Fall scheint es sehr von der Situation abzuhängen, ob es einen zusätzlichen Nutzen von Video gegenüber nur Audio gibt [20], [83].
- Die *Verbindungs-Hypothese* hingegen geht davon aus, dass der Videokanal weniger als Unterstützung des Audiokanals verwendet wird, sondern das Zustandekommen einer 'opportunistic connection' erleichtert, indem die Kommunikationsverfügbarkeit des gewünschten Gesprächspartners beispielsweise durch das Betrachten eines Videobilds abgeschätzt wird, das die

Gesamtansicht eines Büros zeigt. Somit kann der Kommunikationsinitiator anhand der momentanen Tätigkeit des gewünschten Kommunikationspartners einschätzen, ob dieser gerade 'gestört' werden kann oder nicht [47].

- Andere Studien sehen die Möglichkeit für den Nutzen von Video in der Verwendung von 'Video als Zusatzinformation' (Hypothese *video as data*). Diesmal aber nicht in dem Sinne, dass das Gesicht oder eine Komplettansicht des Gesprächspartners übertragen wird, sondern dreidimensionale Gegenstände, die im Zusammenhang der Konferenz eine wichtige Rolle spielen. Somit wird ein dynamischer Arbeitsplatz zur Verfügung gestellt und eine gemeinsame physikalische Umgebung simuliert [81].

In den Untersuchungen in Kapitel 6 wird die Videokonferenz im Sinne von den oben genannten Punkten 1 (non-verbale Kommunikation) und 3 (video as data) verstanden.

### *5.1.1 QoS-Parameter*

Nachfolgend werden verschiedene Parameter vorgestellt, die bei Videokonferenzen eine wichtige Rolle spielen. Anschliessend wird dargelegt, welche Parameter wirklich untersucht werden und aus welchen Gründen.

#### *5.1.1.1 Synchronisation*

Multimediakommunikation im Sinne dieser Arbeit besteht aus zwei wesentlichen Kanälen: Audio und Video. Die meisten Studien, die sich mit Untersuchungen zu Multimediakonferenzen befassen, erkennen den Audiokanal als den wichtigeren Informationsträger der beiden Kanäle [54]. Das Weglassen des Audiosignals hat schwerwiegende Auswirkungen auf das Leistungsverhalten des Benutzers [90]. Die Benutzer scheinen eine kleine relative Verzögerung von Video gegenüber Audio einer grösseren absoluten Verzögerung bei gleichzeitiger Synchronität der beiden Medienströme vorzuziehen [83].

#### *5.1.1.2 Bandbreite*

Die bei einer Multimediaanwendung verwandte Bandbreite ist in der Regel proportional zur zeitlichen und räumlichen Auflösung des Bildsignals (z.B. Auflösung und Wiederholrate) sowie zur Tonqualität [31]. Dabei ist aber zu beachten, dass die gesamthaft zur Verfügung stehende Bandbreite sinnvoll auf die Video- und Audiodaten aufgeteilt werden muss. Unter Umständen ist sonst eine Videokonferenz mit hervorragender Bildqualität und miserabler Tonqualität bandbreitenintensiver als mit guter Bildqualität und hervorragender Ton-

qualität. Sowohl bei der Bild- als auch der Tonqualität besteht bei konstanter Bandbreite immer ein Kompromiss (trade-off) zwischen zeitlicher und räumlicher Auflösung.

Eine Studie, die sich mit Benutzerzufriedenheit in Abhängigkeit von Bildwiederholraten beschäftigt, hat gezeigt, dass die Wiederholrate bei unterschiedlichen Bildinhalten verschieden wichtig ist [3]. Es wird gefolgert, dass bei einer Bildgrösse von 160x120 Pixeln eine Rate von 15 Bildern/s gerade noch akzeptiert wird. 16 Bilder/s sind nötig für zufriedenstellende Lippsynchronisation [46]. Unabhängig von der tatsächlichen zeitlichen Korrelation haben Benutzer bei Raten unterhalb von 5 Bildern/s das Gefühl, dass Video und Audio asynchron seien [85].

Die Anforderungen an die Tonqualität hängt sehr stark von der Anwendung ab. Da Benutzer (und im speziellen Neulinge) die Qualität oft mit bekannten Anwendungen vergleichen [15], sollte Telefonqualität für normale Videokonferenzen ausreichen. Der Zusammenhang zwischen Bandbreite und Sprach- bzw. Tonqualität ist der besten ausgelotete Problemkreis innerhalb der Telekommunikation. In diesem Zusammenhang wurden unzählige Studien gemacht und insbesondere haben die Organisationen CCIF bzw. CCITT und die angeschlossenen Netzbetreiber viel Aufwand betrieben um nachzuweisen, dass die für die FDM-Telefonsysteme gewählte Bandbreite von 0.3 bis 3.4 kHz (mit einem Kanalaraster von 4 kHz) einen ökonomischen bzw. aus Benutzersicht vertretbaren Kompromiss darstellt. Ökonomisch, da Filter mit guard-bands von total 900 Hz einigermaßen leicht zu konstruieren waren und mit einem Trägerabstand von 4 kHz die damaligen physischen Medien gut ausgenutzt werden konnten. Aus Benutzersicht, weil trotz weitgehender Unterdrückung der Zischlaute noch eine akzeptabel hohe Silben-, Wort- und Satzverständlichkeit resultierte. Diese Überlegungen sind anschliessend aus Kompatibilitätsgründen auch für die PCM-Telefonie übernommen worden, wobei es galt, dafür zusätzlich die räumliche „Auflösung“ festzulegen. Man wählte 8 Bit/sample, damit Quantisierungsverzerrungen auch bei mehrmaliger A/D – D/A- Wandlung nicht hörbar werden [58].

### *5.1.1.3 Jitter*

Jitter ist die variable Verzögerung, die Datenelemente beim Durchlauf eines oder mehrerer Übertragungs- oder Verarbeitungssysteme erfahren [82]. Von Jitter spricht man also, wenn Pakete, die in gleichbleibenden Abständen losgeschickt worden sind, mit variablen Abständen ankommen. Jitter kann immer mittels zeitlicher Glättung auf Kosten der absoluten Verzögerung ausgeglichen

werden. Massgeblich für die Qualität ist einerseits der residuale Jitter (abhängig vom Quellenkodierungs- bzw. Dekodierungsverfahren) und die genannte absolute Verzögerung. Je nach Anforderungen muss ein passender Kompromiss eingegangen werden. Bei interaktiven Multimediaanwendungen wird von Bouch ein von Benutzern maximal tolerierter Wert von 200 ms angegeben [15].

#### *5.1.1.4 Verlust von Audiopaketen*

Bei der Übermittlung von Audiodaten wird in modernen Systemen zuerst eine Quellenkodierung der Daten vorgenommen und danach eine Kanalkodierung. Erstere dient dazu, beispielsweise anhand von psychoakustischen Modellen irrelevante Daten zu entfernen, damit der zu übertragende Datenstrom kleiner wird. Letztere fügt je nach zu verwendendem Übertragungskanal wieder Redundanz hinzu, damit Fehler anschliessend bis zu einem gewissen Grad wieder ausgemerzt werden können. Es versteht sich, dass die hinzugefügte Redundanz kleiner sein sollte als die zuvor entfernten Quelldaten. Je nach verwendeten Verfahren ist der resultierende Bitsrom mehr oder weniger anfällig auf Paketverluste.

#### *5.1.1.5 Absolute Verzögerung*

Untersuchungen des Einflusses von absoluter Verzögerung sowohl bei Video als auch bei Audiokonferenzen sind recht verbreitet. Absolute Verzögerung kann nie ganz eliminiert werden, zumindest derjenige Teil, der durch die physikalische Übertragung gegeben ist, bleibt immer vorhanden. Allerdings ist die reine Übertragungszeit im Vergleich zu den anderen Quellen meist vernachlässigbar klein. Nur bei transkontinentalen Verbindungen erreicht dieser Wert Grössenordnungen von bis zu 100 ms (ohne die Nutzung von Satellitenverbindungen). Viel gewichtiger sind hingegen systeminhärente und Kompressionsverzögerungen. Ein paketbasiertes Netzwerk beispielsweise sendet Daten erst dann los, wenn ein Paket voll ist, somit entsteht die sogenannte Paketisierungsverzögerung, die sich zur gesamten Verzögerungszeit addiert. Weitere Faktoren sind Wartezeiten in Pufferspeichern, Verarbeitungszeiten bei der Komprimierung und Dekomprimierung, A/D- und D/A-Wandlung etc. Eine Zusammenstellung der beteiligten Faktoren sowie typischer Werte wurde in Tabelle 2.2 dargestellt.

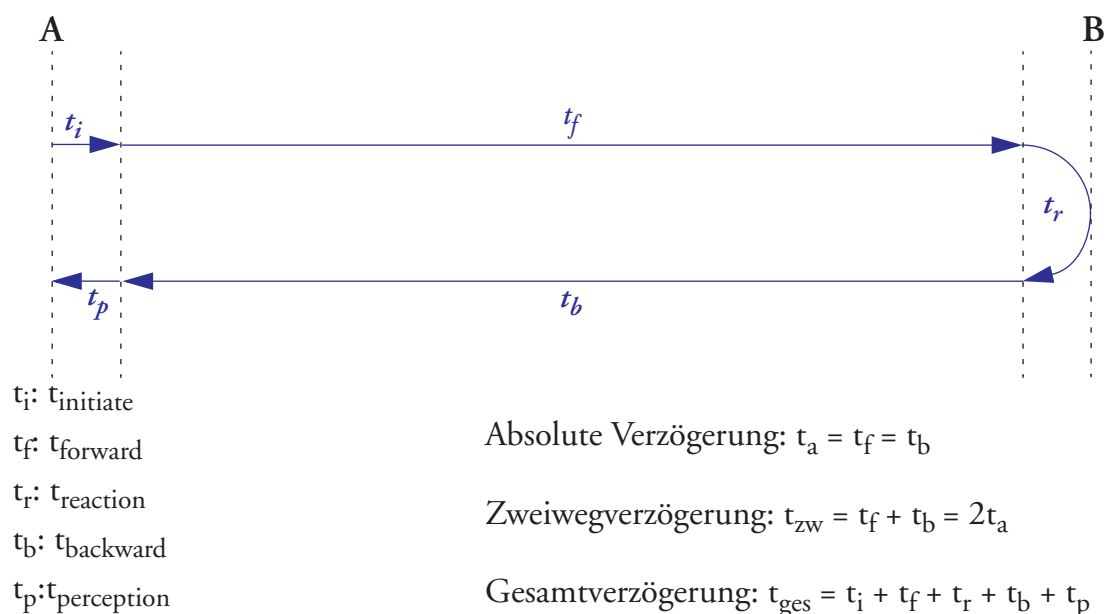
Bei der Auswahl der relevanten Qualitätsparameter hat sich diese Arbeit hauptsächlich auf die *absolute Verzögerung* des kompletten Multimediadatenstroms konzentriert. Die folgenden Gründe sprechen für diese Gewichtung:

- bei Verbindungen über grosse Distanzen sind alleine schon die Laufzeiten qualitätsrelevant
- immer öfter werden Quellencodierungsverfahren eingesetzt, die wegen der komplexen Algorithmen zu namhaften Verzögerungen führen
- Massnahmen zur Verhinderung, Maskierung oder Beseitigung von Störungen sind unsausweichlich mit weiteren Verzögerungen verbunden.

Zusätzlich wurde schon bei den Betrachtungen zu den Trends in der Telekommunikation festgestellt, dass insbesondere neue Multimediadienstleistungen und -applikationen auf das Vertrauen der Benutzer angewiesen sind und dieses sicherlich stark davon abhängt, ob eine Kommunikation in Echtzeit möglich ist oder nicht.

Obwohl in einem Vorversuch auch die Parameter *Bildwiederholfrequenz* und *relative Verzögerung* (Synchronisationsverzögerung zwischen auditivem und visuellem Datenstrom) betrachtet wurden [69], musste erkannt werden, dass der durch die Verwendung von mehreren unabhängigen Variablen sich enorm vergrössernde Untersuchungsraum für die weiteren, genaueren Tests zu komplex geworden wäre. Allerdings wird auf die Ergebnisse in Abschnitt 6.1 trotzdem eingegangen, da aus ihnen wichtige Erkenntnisse für die nachfolgenden Untersuchungen abgeleitet werden konnten.

Im folgenden werden die Begriffe *absolute Verzögerung*, *Zweiwegverzögerung* und *Gesamtverzögerung* definiert.



*Abbildung 5.9: Definition von Gesamtverzögerung und Zweiwegverzögerung*



Unter *absolute Verzögerung* wird die Zeitspanne verstanden, die durch technische Systeme sowie Übertragung von A nach B (Einwegverzögerung) entstehen. Diese Verzögerung setzt sich aus den in Tabelle 2.2 genannten Faktoren zusammen. Der Begriff *Zweiwegverzögerung* bezieht die Tatsache mit ein, dass ein Kommunikationspartner einer Videokonferenz im Vergleich zu einem Gespräch von Angesicht zu Angesicht die absolute Verzögerung doppelt wahrnimmt, nämlich einmal vom Hin- und einmal vom Rückweg. Die Gesamtverzögerung schliesst zusätzlich noch die Zeiten ein, die A benötigt, um seine Aussage zu formulieren, die B für die Wahrnehmung und Formulierung der Antwort benötigt sowie jene, die A wiederum für die Wahrnehmung braucht. In Kapitel 6 ist dementsprechend bei der Angabe von Verzögerungswerten immer die *Zweiwegverzögerung* gemeint.

### *5.1.2 Vergleich mit anderen Kommunikationssituationen*

Bei den Betrachtungen innerhalb dieser Arbeit wird davon ausgegangen, dass die natürliche Referenzsituation für eine Diskussion via Videokonferenz ein Gespräch von Angesicht zu Angesicht ist. Als weitere Vergleichssituation dient das Telefongespräch bzw. die Audiokonferenz, die in der interpersonellen Telekommunikation bei weitem die bekannteste und verbreitetste Art des Informationsaustausches ist und damit im Sinne der Thesen von Bouch und Sasse bewusst oder unbewusst als Qualitätsmassstab verwendet wird. Nicht zum Gegenstand unserer Untersuchungen gehört es, herauszufinden, wie eine Videokonferenz qualitativ gesehen gegenüber einem Live-Gespräch oder einem Telefonat abschneidet und warum. Für diese Fragestellungen wird auf Untersuchungen verwiesen, in denen solche Vergleiche mit unterschiedlichen Systemen (im Labor und Feldversuch, mit niedriger und hoher Qualität) und unter verschiedensten Gesichtspunkten untersucht wurden ([21], [24], [29], [30], [38], [60], [63], [71], [74], [83]).

Um die Ergebnisse nicht nur qualitativ, sondern auch quantitativ direkt mit anderen Studien vergleichen zu können, in denen lediglich Untersuchungen im Audibereich durchgeführt wurden, wurde ein Referenzversuch durchgeführt, der im Vergleich zur entsprechenden Videokonferenz keine Änderungen erfahren hat, ausser dass der Videokanal weggelassen wurde.

### *5.1.3 Anforderungen an Tasks*

Nachdem wir uns entschlossen hatten, als Testplattform ein Videokonferenzsystem zu verwenden, mussten wir uns Gedanken machen, welchen Zweck dieses

bei der Kommunikation zu erfüllen hatte bzw. welche Aufgabe unsere Teilnehmer während ihrer Konferenz zu lösen haben sollten. Speziell in diesem Fall waren wir auf die Zusammenarbeit mit unserem Partnerinstitut IHA angewiesen [103]. Folgende Aspekte wurden herausgearbeitet:

- Natürliche Umgebung, möglichst wenig Einfluss der „Laborbedingungen“.
- Simulieren von Kommunikation im privaten und im geschäftlichen Bereich mit zwei oder drei beteiligten Personen.
- Wiederholbarkeit der Untersuchungen zu einem späteren Zeitpunkt.
- Keine „Verkabelung“ der Probanden bzw. keine sonstigen in die persönlichen Entfaltungsmöglichkeiten eingreifenden Messapparaturen sollten verwendet werden.
- Aufgabe sollte möglichst ohne spezielles Vorwissen und unabhängig von Alter, Geschlecht etc. zu lösen sein (kein Messen von Nebeneffekten).
- Die Dauer eines Versuchs sollte nicht länger als 60-90 Minuten sein, da sonst der Einfluss durch Ermüdung, Unlust bzw. Gewöhnung das Resultat verfälschen kann.
- Der Task sollte realistisch sein, nicht hypothetischer Art.
- Beide bzw. alle drei Personen sollten symmetrisch in die Aufgabe eingebunden werden, so dass Dialoge und keine Monologe entstehen<sup>1</sup>.
- Die Interaktion bei den Tasks sollte möglichst hoch sein, da wir in diesem Fall kritischere Grenzwerte erwarteten (worst case).
- Es soll nach Möglichkeit die *Methode 2AFC (2-alternative-forced-choice)* [33] eingesetzt werden, bei der einer der beiden Verzögerungswerte in zufälliger Reihenfolge 0 ms aufweist. Damit wird ein eindeutiger Bezugswert definiert.

In Tabelle 5.5 unten sind die vier Versuchsphasen mit der jeweiligen Anzahl der Teilnehmer (#TN), der Art des Versuchs (VC: Videokonferenz; AC: Audiokonferenz; SW: Shared Workspace) sowie des gewählten Tasks in einer Übersicht zusammengestellt. Auf die einzelnen Tasks gehen wir in Kapitel 6 bei der Beschreibung der einzelnen Phasen genauer ein.

---

<sup>1</sup> Auf Situationen mit verschiedenartigen Rollen (Vorsitz, Protokollführer, ...) wurde bewusst verzichtet, da zu erwarten war, dass diese rigideren Strukturen auch zu disziplinierteren Gesprächsabläufen führen würden. Aufgrund der bereits getätigten Vorversuche [69] sind damit weniger kritische Qualitätsanforderungen verknüpft.

Versuch	Art	# TN	Task	Beschreibung
1	VC  VC	2	a) Sprichwortraten  b) Provokative Thesen	Die Kandidaten müssen vorgegebene Sprichwörter erraten. Dabei kennt Kandidat A das Sprichwort, Kandidat B soll dieses durch die Umschreibungen von Kandidat A herausfinden.  Die beiden Kandidaten sollen über von den Versuchsleitern aufgestellte, provokative Thesen eine möglichst hitzige Diskussion führen.
2	VC  VC	2	a) Flugplan  b) Legomodell	Kandidat A hat die Aufgabe, anhand eines Flugplanes Auskunft über Flugziele und Flugzeiten an Kandidat B zu geben, welcher seinerseits innerhalb von 10 Tagen möglichst viele Städte in den USA besuchen soll, ohne eine Stadt doppelt anzufliegen.  Die beiden Kandidaten müssen zwei Legomodelle zusammenbauen, wobei jeweils ein Kandidat die Anleitung zur Verfügung hat und der andere die Bausteine.
3	VC  VC + SW	3	a) WG-Gründung  b) Aktienhandel	Kandidat A sucht zwei Mitbewohner für seine Wohnung und möchte mit den Kandidaten B und C eine WG gründen.  Ein Anlageteam bestehend aus Kandidat A und B betreibt möglichst effektiv Aktienhandel von zwei Titeln, wobei es durch einen sich ändernden Kurs und einen Börsenmakler (Kandidat C) beeinflusst wird.
4	AC	3	1. WG-Gründung	Wie 3a) ohne Videokanal

*Tabelle 5.5: Die vier Versuchsphasen mit den verwendeten Tasks*

### 5.1.4 Variablen

Für die Durchführung eines kontrollierten Experimentes muss zuerst herausgefunden werden, welche Variablen das Ergebnis beeinflussen. Idealerweise sollten alle Variablen bis auf eine konstant gehalten werden, so dass ein wirklich aussagekräftiges Ergebnis erwartet werden kann. Diese Aufgabe ist jedoch sehr schwierig bis unmöglich, denn die Variablen beeinflussen sich gegenseitig und eher selten hängt der zu untersuchende Effekt nur von einem Parameter alleine ab. Es werden folgende Arten von Variablen unterschieden:

- Abhängige Variablen
- Unabhängige Variablen
- Kontrollvariablen

#### 5.1.4.1 Abhängige Variablen

Die abhängigen Variablen sind die während der Versuche bewusst zu messenden Grössen. Im Fall dieser Arbeit ist dies die Wahrnehmungsschwelle<sup>1</sup> bzw. Akzeptanzschwelle<sup>2</sup>, die Benutzer bei einer Videokonferenz bezüglich verschiedener Qualitätsparameter haben. Die Messung dieser Grössen ist jedoch problematisch, da Benutzer dazu neigen, sich mit einer bestimmten Dienstgüte (in Abhängigkeit von verschiedenen Einflüssen und Faktoren) zufrieden zu geben, obwohl diese eigentlich unterhalb der erforderlichen Schwelle liegt. Dies kommt daher, dass subjektive Bewertungen immer wahrnehmungsgesteuert und somit mit einer persönlichen Interpretation behaftet sind. Aus diesem Grund kann eine subjektive Bewertung niemals alleine als Messgrösse für objektive Betrachtungen über eine bestimmte QoS dienen. Nichts desto trotz ist es möglich, aus subjektiven Bewertungen wichtige Daten und Entscheidungsgrundlagen herauszuziehen. Zusätzlich müssen jedoch weitere Bewertungsgrundlagen herangezogen werden wie *task performance*<sup>3</sup> und *Wechselwirkungsaufwand*<sup>4</sup> [91]. Selbst diese sind aber nicht über alle Zweifel erhaben. *Task performance* ist zwar eine objektive Messgrösse, kann aber durch andere Einflüsse als durch die unabhängigen Variablen beeinflusst werden. Zur

---

<sup>1</sup> Als Wahrnehmungsschwelle wird diejenige Schwelle bezeichnet, ab der Benutzer eine Videokonferenz die Zweiwegverzögerung eines technischen Systems wahrnehmen.

<sup>2</sup> Als Akzeptanzschwelle wird diejenige Schwelle bezeichnet, ab der Benutzer eine Videokonferenz die Zweiwegverzögerung eines technischen Systems als störend empfinden.

<sup>3</sup> Die *Task Performance* ist ein Leistungskennwert, der gemessen wird, während die Probanden miteinander kommunizieren (beispielsweise die Zeit, die es für das Lösen einer Aufgabe braucht).

<sup>4</sup> siehe Seite 62

Messung des Wechselwirkungsaufwands schlägt Wilson psychometrische Methoden vor, die teilweise jedoch einen starken Eingriff in die Bewegungsfreiheit der Teilnehmer voraussetzen würden.

Aus der Reihe der von aussen messbaren, objektiven Parameter wurden die *task performance* gemessen, d.h. durch das Festhalten der Zeitdauer, die die Testteilnehmer zum Lösen der Aufgabe bzw. von Teilaufgaben benötigt haben, wurde die Leistungsfähigkeit des entsprechenden „Versuchsteams“ in Abhängigkeit der jeweils verwendeten Störparameter festgehalten.

Psychometrische Messungen, die eine „Verkabelung“ der Teilnehmer erforderlich gemacht hätten bzw. invasive Messungen bei den Probanden wurden keine gemacht, da der Autor der Ansicht ist, dass dies erstens zu „abschreckend“ gewesen wäre und zweitens einen hohen Mehraufwand bei fraglichem Nutzen bedeutet hätte. Statt dessen wurde in einem Versuchsdurchlauf gemessen, wie sich *Bestätigungsäusserungen* (backchannels), *Gegenseitige Unterbrechungen*<sup>1</sup> (interruptions) und *Sprechphasen* (turns) bei verschiedenen Dienstgütern verhalten. O'Malley [61], Boyle [17] und Anderson [1] haben in ihren Untersuchungen nicht übereinstimmende Ergebnisse erhalten. Dies hat die Versuchsleiter<sup>2</sup> dazu bewogen, eine eigene These H2.2 (siehe Abschnitt 5.3) zu formulieren und die gegenseitigen Unterbrechungen als weiteren Parameter hinzuzunehmen.

Folgende abhängige Variablen wurden gemessen:

- Subjektive Wahrnehmung

Während der Versuchsdurchführung wurden die Teilnehmer in regelmässigen Abständen darüber befragt, ob sie während der Durchführung des Tasks eine Verzögerung bemerkt hatten oder nicht. Technisch wurde dies durch ein Popup-Fenster auf dem Bildschirm gelöst, in dem entweder *ja* oder *nein* als Antwort per Maus angewählt werden konnte.

- Subjektive Zufriedenheit

Hatten die Benutzer eine Verzögerung bemerkt (Antwort *ja*), wurde Ihnen eine zweite Frage gestellt, ob sich diese störend ausgewirkt hatte oder nicht. Die

---

<sup>1</sup> Dies sind Situationen, in denen B A ins Wort fällt, ohne diesen eigentlich unterbrechen zu wollen. Aufgrund der zusätzlich eingefügten Verzögerung merkt er jedoch zu spät, dass A mit seinen Ausführungen noch nicht fertig war.

<sup>2</sup> Diese Untersuchungen wurden in einem Team durchgeführt [59], [64].

Antworten auf die Fragen wurden dazu benutzt, die beiden Schwellwerte in Iterationsschritten anzunähern. Um dies auf möglichst effiziente Art und Weise zu erreichen, wurde nach einem Verfahren gesucht, das diesen Ansprüchen genügt. Dabei stellte sich das Best-PEST (Parameter Estimation of Sequential Trials) Verfahren als besonders geeignet heraus, das eine Variante der Maximum-Likelihood Methoden darstellt. Dieses Verfahren maximiert die Information, die nach jedem Messintervall zur Verfügung steht, und stellt somit sicher, dass der gesuchte Schwellwert in der kleinstmöglichen Anzahl von Intervallen erreicht wird. Für eine detaillierte Beschreibung des Algorithmus wird auf [53] und [102] verwiesen.

- Subjektive Einschätzung und Meinung

Vor, während und nach den einzelnen Testdurchgängen wurden den Teilnehmenden Fragen zu ihren Erwartungen, ihrem technischen Vorwissen, ihrer Meinung und ihrer momentanen Verfassung gestellt. Einerseits dienten diese Fragen zur Ausmittlung diverser Kontrollvariablen (siehe Unterabschnitt 5.1.4.3), andererseits zur Verifizierung bzw. Falsifizierung der Hypothesen (Abschnitt 5.3) sowie weiteren qualitativen Analysen der Ergebnisse.

- *Wechselwirkungsaufwand*<sup>1</sup>

Wie bereits erwähnt, wurden folgende Parameter beobachtet:

#### **Backchannels**

Kurze Bestätigungsäusserungen des Zuhörers als Zeichen seiner Aufmerksamkeit wie „mmmhm“, „jawohl“, „ok“, „ja“, Es wird erwartet, dass diese bei grösserer Verzögerung abnehmen.

#### **Interruptions**

Gleichzeitiges Sprechen, ohne dass der Sprecher die Kommunikation abgibt oder seine Aussage beendet hätte. Es wird eine Zunahme solcher Unterbrechungen bei grösserer Verzögerung erwartet.

#### **Turn-taking**

Sprecherwechsel sollten bei zunehmender Verzögerung abnehmen und die Dauer von Sprechphasen somit zunehmen.

- Beobachtung von aussen, Notizen

---

<sup>1</sup> Dieser Begriff wird von Wilson [91] im Englischen mit *user cost* bezeichnet. Der Autor ist jedoch der Meinung, dass der von ihm gewählte Begriff *Wechselwirkungsaufwand* den Sachverhalt wesentlich verständlicher beschreibt. Nachfolgend wird deshalb der deutsche Begriff verwendet.

Zusätzlich zu den subjektiven Wahrnehmungen der Teilnehmer, der objektiv gemessenen Task-Performance und der Auswertung der Fragebögen wurden die Versuchsabläufe „von aussen“ beobachtet und auffällige Ereignisse bzw. Feststellungen notiert. Beispielsweise wurde beobachtet, dass Teilnehmer, die der deutschen Sprache nicht sehr gut mächtig waren, automatisch die Toleranzgrenze gegenüber Verzögerungen bei den anderen Versuchsteilnehmern erhöht haben, unabhängig von der Variation weiterer Parameter.

#### *5.1.4.2 Unabhängige Variablen*

Die unabhängigen Variablen sind die Grössen, die in den Experimenten bewusst verändert werden. Im Fall dieser Arbeit sind dies je nach Versuchsphase die folgenden:

##### *Versuch 1*

- Absoluter Verzögerungswert bei einer Videokonferenz
- Relativer Verzögerungswert bei einer Videokonferenz
- Bildwiederholrate bei einer Videokonferenz

##### *Versuch 2*

- Absoluter Verzögerungswert bei einer Videokonferenz
- Auszuführender Task
- Computer- und Netzwerkkennnisse

##### *Versuch 3*

- Absoluter Verzögerungswert bei einer Videokonferenz
- Auszuführender Task

##### *Versuch 4*

- Absoluter Verzögerungswert bei einer Audiokonferenz

#### *5.1.4.3 Kontrollvariablen*

Die Kontrollvariablen sind diejenigen, die bei den Versuchen kontrolliert werden sollten (d.h. sie werden konstant gehalten, so dass sie die einzelnen Versuchsergebnisse nicht mitbeeinflussen). Ist dies nicht möglich, muss ihr Einfluss auf die Ergebnisse durch statistische Verfahren ausgeschaltet werden,

z.B. in dem die verwendete Testgruppe genügend gross gewählt wird. Bei unseren Tests wurden folgende Kontrollvariablen berücksichtigt:

#### Variablen im Zusammenhang mit den Testpersonen

- Erfahrung mit Videokonferenzsystemen oder ähnlichen Anwendungen
- Muttersprache
- Erfahrung mit der gestellten Aufgabe
- Alter
- Geschlecht
- Einfluss von Drogen und Alkohol

Variablen wie Muttersprache, Vertrautheit mit den gestellten Aufgaben usw. wurden durch die Auswahl der Teilnehmer konstant gehalten. Alle anderen wurden entweder durch eine hohe Anzahl der Tests statistisch ausgemittelt oder aber es wurde anhand von Fragebögen versucht, deren Einfluss auf die Ergebnisse abzuschätzen und zu erklären. Die jeweiligen Statistiken und Ergebnisse finden sich in Kapitel 6 bei den einzelnen Versuchen.

#### Umgebungsvariablen

- Licht
- Lärm
- Umgebungstemperatur
- Luftqualität und -feuchtigkeit
- Bedienung der Workstation

Die Umgebungsvariablen wurden für alle Versuchsteilnehmer einer Phase konstant gehalten. Es wurde für eine gleichmässige, angenehme und nicht blendende Beleuchtung gesorgt, Umgebungslärm gab es keinen. Nach jedem Durchgang wurde für 5 Minuten gelüftet.

#### Variablen in Verbindung mit dem eigentlichen Experiment

- Aufmerksamkeitsschwankungen
- Müdigkeit (Tageszeit)
- Gewöhnungseffekt



- Stressfaktor
- Wissen über die beteiligten Teilnehmer
- Erwartungen an die Tests

Diese Variablen waren am schwierigsten in den Griff zu bekommen. Teilweise konnten ihre Einflüsse wieder durch das Auswerten von Fragebögen bzw. durch aktive Beobachtung des Versuchsablaufs von aussen eliminiert werden. Die jeweiligen Statistiken und Ergebnisse finden sich in Kapitel 6 bei den einzelnen Versuchen.

## 5.2 *Auswahl der Versuchsplattform*

Anhand der Ergebnisse der Literaturrecherche und der Erfahrungen aus der ersten Versuchsphase wird davon ausgegangen, dass die Versuchsplattform für weitere Versuche strengen Kriterien genügen muss. Insbesondere darf die Verzögerung des Messsystems selbst auf keinen Fall gleich gross oder grösser sein als die zu messende Wahrnehmungsschwelle der Testteilnehmer. Zusätzlich dürfen auch keine weiteren, wichtigen Qualitätsparameter den zu untersuchenden maskieren (siehe 6.1.4). Das bedeutet, dass eine Plattform gewählt werden muss, die insbesondere bei der Bildqualität (Bildwiederholrate, Bildgrösse), der Tonqualität (Auflösung und Abtastrate) sowie bei der Synchronität zwischen Bild und Ton (Lippensynchronität) die hohen Ansprüche der Testpersonen erfüllt.

Im Rahmen von [9] wurden verschiedene mögliche Plattformen für die subjektiven Versuche auf ihre Tauglichkeit hin geprüft. Insbesondere wurden Überlegungen angestellt, ob die Verwendung eines kommerziell erhältlichen Systems erstens die Reproduzierbarkeit solcher Untersuchungen erleichtern würde und zweitens Versuchsbedingungen entstehen würden, wie sie in realen Umgebungen (zu Hause oder am Arbeitsplatz) anzutreffen sind.

Es wurden verschiedene käufliche, auf IP-Technologie basierende Videokonferenzsysteme untereinander und mit der bereits vorhandenen Lösung aus den ersten Vorversuchen verglichen. Es wurde festgestellt, dass keine unter ihnen auch nur annähernd an die geforderte Leistung herankam. Weder die Auflösung noch die Wiederholrate des Bildes war zufriedenstellend, noch war die Lippensynchronität gewährleistet. Einzig die Audioqualität konnte überzeugen, allerdings waren die restlichen Parameter derart schlecht, dass keines der Systeme für die Untersuchungen in Frage kommen konnte. Ausserdem hat das

bereits vorhandene System die Vorteile, dass die für das Leistungsverhalten massgeblichen Mechanismen genau bekannt sind und mit dem selbstentwickelten ATM-Switch sowie der Box, die Verzögerungen in eine ATM-Verbindung einfügen kann, auch die Parameter des Netzwerks vollkommen unter Kontrolle sind. Es sollte sich im Verlauf der Versuche herausstellen, dass dies ein nicht zu unterschätzender Vorteil war und vor grösseren Problemen bewahren sollte. Der Nachteil, der durch das proprietäre System in Kauf genommen werden musste, nämlich dass die exakte Reproduzierbarkeit nur wieder mit einem vergleichbaren System möglich sein wird, war im Vergleich zu den vielen Vorteilen das kleinere Übel.

Somit wurde in allen vier Versuchsreihen ein System [67] verwendet das aus den am eigenen Institut entwickelten Geräten ETHMICS (Multimediaworkstation auf Basis eines Macintosh Computers) [28], [32], ARES (ATM Real-Time Emulation System) [50], einem 4-Port ATM-Switch sowie einem Steuercomputer bestand. Die verwendeten Systemkonfigurationen sind in [69] (Versuchsreihe 1), und [64] (Reihen 2 bis 4) genauer beschrieben.

### 5.3 *Arbeitshypothesen und Fragestellungen*

Die durchgeführten Versuche wurden in mehrere Versuchsreihen unterteilt, da man sich langsam an die eigentlich interessanten Fragestellungen heranarbeiten wollte. Die Versuchsreihe 1 (Abschnitt 6.1) diente als Vortest, in dem herausgefunden werden sollte, ob wirklich die absolute Verzögerung eine solch grosse Rolle spielte wie angenommen und in welchem zeitlichen Bereich nachfolgende Reihen diese genauer untersuchen sollten. Die Versuchsreihen 2 (Abschnitt 6.2) und 3 (Abschnitt 6.3) führten genau diese Untersuchungen über Schwellwerte bezüglich der absoluten Verzögerung durch, einmal mit zwei und einmal mit drei Teilnehmern einer Videokonferenz. Versuchsreihe 4 (Abschnitt 6.4) dient als Vergleichsbasis mit verwandten Arbeiten, in denen bereits Erkenntnisse über reine Audiokonferenzen gewonnen wurden.

#### 5.3.1 *Hypothesen und Fragen Versuchsphase 1*

*H1.1:* Die verwendete *Versuchsplattform* eignet sich, um damit auch die weiteren Versuchsreihen durchführen zu können.

*H1.2:* Bei zeitlicher Asynchronität zwischen Video- und Audiosignal wird die Situation *Ton vor Bild* zeitlich gesehen früher wahrgenommen bzw. als störend empfunden als die Situation *Bild vor Ton*.<sup>1</sup>

*F1.1:* Welches sind die *wichtigsten Qualitätsparameter* hinsichtlich Benutzerzufriedenheit bei einer Videokonferenz?

### 5.3.2 Hypothesen und Fragen Versuchsphase 2

*H2.1:* Die *Akzeptanzschwelle* der absoluten Verzögerung (die Schwelle, ab der man sich durch die Verzögerung gestört fühlt) bei einer Videokonferenz hängt vom auszuführenden Task ab.

*H2.2:* Die *Akzeptanzschwelle* der absoluten Verzögerung bei einer Videokonferenz hängt vom Vorwissen der Benutzer über Netzwerke und Computer ab.

*H2.3:* Die *Akzeptanzschwelle* der absoluten Verzögerung bei einer Videokonferenz liegt für diejenigen Benutzer höher, die über die zu Grunde liegende Netzwerktechnologie Bescheid wissen, als bei denjenigen, die davon keine Ahnung haben.

*H2.4:* Die *Wahrnehmungsschwelle* hängt nicht vom Vorwissen über Netzwerke und Computer ab.

*H2.5:* Die *Wahrnehmungsschwelle* der absoluten Verzögerung liegt tiefer als die Akzeptanzschwelle.

*H2.6:* Die *Wahrnehmungsschwelle* der absoluten Verzögerung bei einer Videokonferenz hängt nicht vom spezifischen Task ab.

*H2.7:* *Gegenseitige Unterbrechungen* nehmen ab, wenn die absolute Verzögerung bei einer Videokonferenz zunimmt.

*F2.1:* Wie hoch liegt die *absolute Verzögerung*, die bei den durchgeführten Tasks *wahrgenommen* wird?

*F2.2:* Wie hoch liegt die *absolute Verzögerung*, die bei den durchgeführten Tasks noch *toleriert* wird?

---

<sup>1</sup> Diese Hypothese entstand einerseits aus der Überlegung, dass die menschliche Erfahrung sich lange Zeit an die Situation *Bild vor Ton* gewöhnen konnte, da die Schallgeschwindigkeit kleiner als die Lichtgeschwindigkeit ist und somit in der Natur die umgekehrte Situation nicht vorkommt. Andererseits liefern die Untersuchungen in [79] keine klare Aussage, ob dem so ist, weshalb man sich entschlossen hat, diesen Aspekt genauer zu untersuchen.

### 5.3.3 Hypothesen Versuchsphase 3

*H3.1:* Die *Wahrnehmungsschwelle* der absoluten Verzögerung liegt tiefer als die *Akzeptanzschwelle*<sup>1</sup>.

*H3.2:* Die *Wahrnehmungsschwelle* der absoluten Verzögerung bei einer Videokonferenz hängt nicht vom *spezifischen Task* ab.

*H3.3:* Die *Wahrnehmungsschwellen* bei Versuchen mit *drei* Teilnehmern liegen tiefer als mit *zwei* Teilnehmern.

*H3.4:* Die *Akzeptanzschwelle* bei Versuchen mit *drei* Teilnehmern liegt tiefer als mit *zwei* Teilnehmern.

*H3.5:* Die *Wahrnehmungs-* und *Akzeptanzschwellen* sind abhängig von der *Tageszeit* und liegen früh morgens höher als am späteren Vormittag oder Nachmittag.

### 5.3.4 Hypothesen Versuchsphase 4

*H4.1:* Die sich ergebende *Akzeptanzschwelle* der Benutzer der Audiokonferenz wird tiefer sein als bei einer vergleichbaren Videokonferenz.

*H4.2:* Die sich ergebende *Wahrnehmungsschwelle* der Benutzer der Audiokonferenz wird tiefer sein als bei einer vergleichbaren Videokonferenz.

Die beiden Hypothesen stützen sich u.a. darauf, dass die Testpersonen bei einer reinen Audiokonferenz deshalb kritischer sein werden, weil sie in diesem Fall bereits eine Vergleichsmöglichkeit (Telefon) haben und dort zumindest bis vor einigen Jahren kaum mit Verzögerungen zu rechnen war (die meisten Telefongespräche werden innerhalb eines Landes oder mit Nachbarländern geführt). Dazu muss gesagt werden, dass das Telefonnetz ursprünglich darauf ausgelegt war, möglichst keine wahrnehmbaren Verzögerungen zu erzeugen. Lediglich durch die Einführung der interkontinentalen Gespräche via Fernmeldesatellit wurde eine wahrnehmbare und auch störende Verzögerung eingeführt.

---

<sup>1</sup> Obwohl diese auf den ersten Blick klar zu sein scheint, sind Situationen möglich, in denen subjektiv keine Störung empfunden wird, obwohl der Körper diese bereits wahrnimmt. Es fehlt jedoch die Sensitivität, dies zu erkennen und zuzuordnen.

# *Kapitel 6*

## *Durchführung der Versuche*

---

In diesem Kapitel werden die vier durchgeführten Versuchsreihen mit den jeweiligen Zielen, Methoden und Ergebnissen vorgestellt. Die erhaltenen Resultate werden nach jedem Abschnitt kurz diskutiert; eine ausführliche Betrachtung und Diskussion erfolgt in Kapitel 8.

### *6.1 Versuchsreihe 1*

#### *6.1.1 Ziele*

An die erste Versuchsreihe wurden insbesondere folgende Ansprüche gestellt:

- Es sollte eruiert werden, ob die gewählte Testumgebung (siehe 5.2) für die Verwendung im QED-Projekt geeignet ist.
- Im Sinne einer Exploration des Versuchsraums sollte ermittelt werden, welche Qualitätsparameter für eine Videokonferenzsituation besonders wichtig sind und auf welche man sich in weiteren Untersuchungen besonders konzentrieren sollte.
- Es sollte sich zeigen, welche Tasks bzw. Aufgaben für subjektive Tests geeignet sind und inwiefern die gewählte Aufgabe die Ergebnisse beeinflussen würden.

#### *6.1.2 Methode*

Die Erarbeitung der Experimente fand zum Teil in Zusammenarbeit mit dem Institut für Hygiene und Arbeitsphysiologie (IHA, ETH Zürich) statt. Als Resultat der Literaturrecherche und den Überlegungen aus Kapitel 2 und 4

wurde abgeleitet, dass eine erste Untersuchung folgende drei Qualitätsparameter umfassen sollte:

- Bildwiederholrate
- Absolute Verzögerung von Bild und Ton
- Relative Verzögerung von Bild zu Ton (Lippensynchronität)

Um den zeitlichen Rahmen von 60 Minuten für ein Experiment nicht zu sprengen, musste ein Kompromiss zwischen der Anzahl verschiedener Parameterkombinationen und der Zeitdauer, während der eine Kombination den Testteilnehmern präsentiert wird, eingegangen werden. Es wurde festgelegt, dass drei Werte pro Parameter (also insgesamt 27 mögliche Kombinationen) verwendet werden sollten und die Dauer von 2 Minuten pro Darbietung nicht überschritten werden sollte.

Für die Durchführung wurden 32 Studierende der Abteilung Pharmazie rekrutiert; zusätzlich haben vier Assistenten des Instituts TIK als 'Experten-Kontrollgruppe' bei den Experimenten mitgemacht.

#### *6.1.2.1 Auswahl des Tasks*

Zusätzlich zu den Anforderungen aus 5.1.3 wurden folgende Überlegungen bei der Taskauswahl mit einbezogen:

- Die Geschwindigkeit beim Lösen der Aufgaben sollte Rückschlüsse über die Qualität der Einstellung erlauben.
- Zwischen den beiden Personen sollte eine intensive Diskussion stattfinden, damit am Schluss auch wirklich beide die Qualität des Gesehenen und Gehörten bewerten können.
- Die Testpersonen sollten bei der Kommunikation ihre Hände nicht mit ins Spiel bringen, weshalb die Kamera auf 'Portrait-Ansicht' eingestellt wurde.

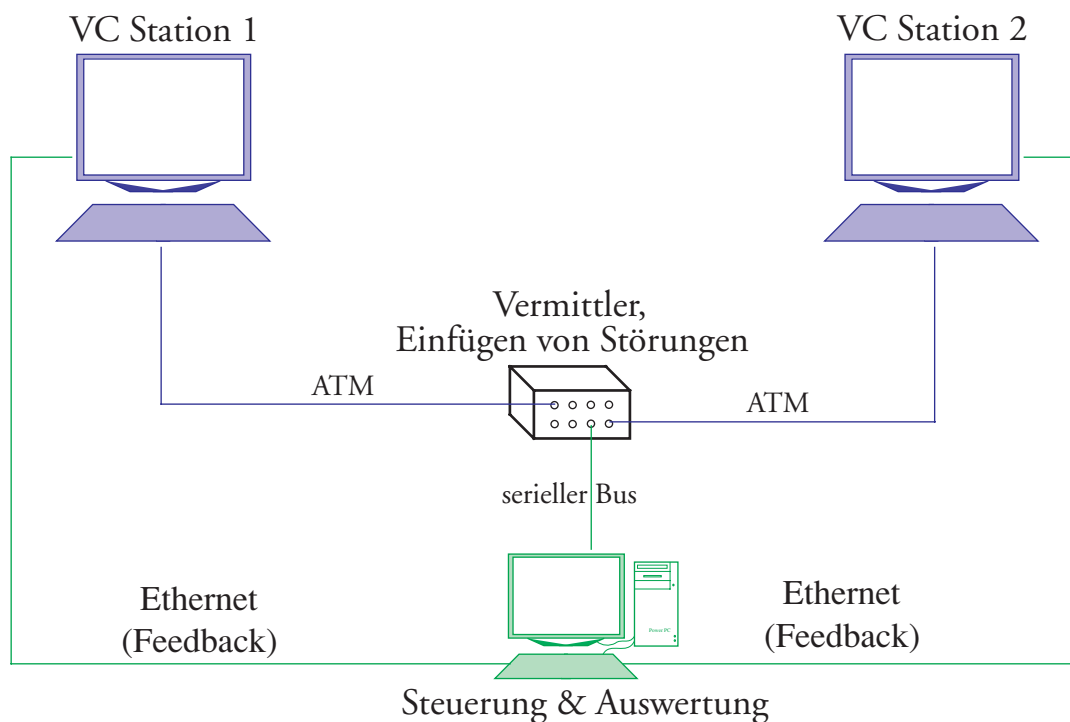
Für die Dauer der Versuche wurden insgesamt zwei unterschiedliche Aufgabenstellungen festgelegt:

- (a) **Sprichwortraten:** Die Kandidaten mussten alternativ von uns vorgegebene Sprichwörter erraten. Dabei kannte Kandidat A das Sprichwort, Kandidat B musste dieses durch seine Umschreibungen herausfinden, und umgekehrt.

(b) **Provokative Thesen:** Die beiden Kandidaten mussten über von uns aufgestellte, provokative Thesen eine möglichst hitzige Diskussion führen.

### 6.1.2.2 Versuchsaufbau

Die Abbildung 6 zeigt schematisch, wie die Komponenten zusammenhängen. Die Testpersonen sitzen an den Multimedia-Workstations, welche über ein ATM-Netzwerk miteinander verbunden sind. Für das Einspeisen der Übertragungsverzögerung und das richtige Verbinden der Datenströme ist der im Kontext einer früheren Promotionsarbeit entwickelte Emulator zuständig [50]. Das Feedback der Teilnehmer sowie die Ausgabe von zufälligen Kombinationen der drei Parameter an den Emulator wird von einem Steuercomputer bearbeitet.



*Abbildung 6.10: Vernetzung der Komponenten*

Um die beiden Testteilnehmer sowohl optisch als auch akustisch möglichst gut voneinander zu trennen, wurden die beiden Videokonferenzstationen in zwei separaten Räumen aufgebaut. Abbildung 6.11 zeigt den Grundriss des Versuchsaufbaus.

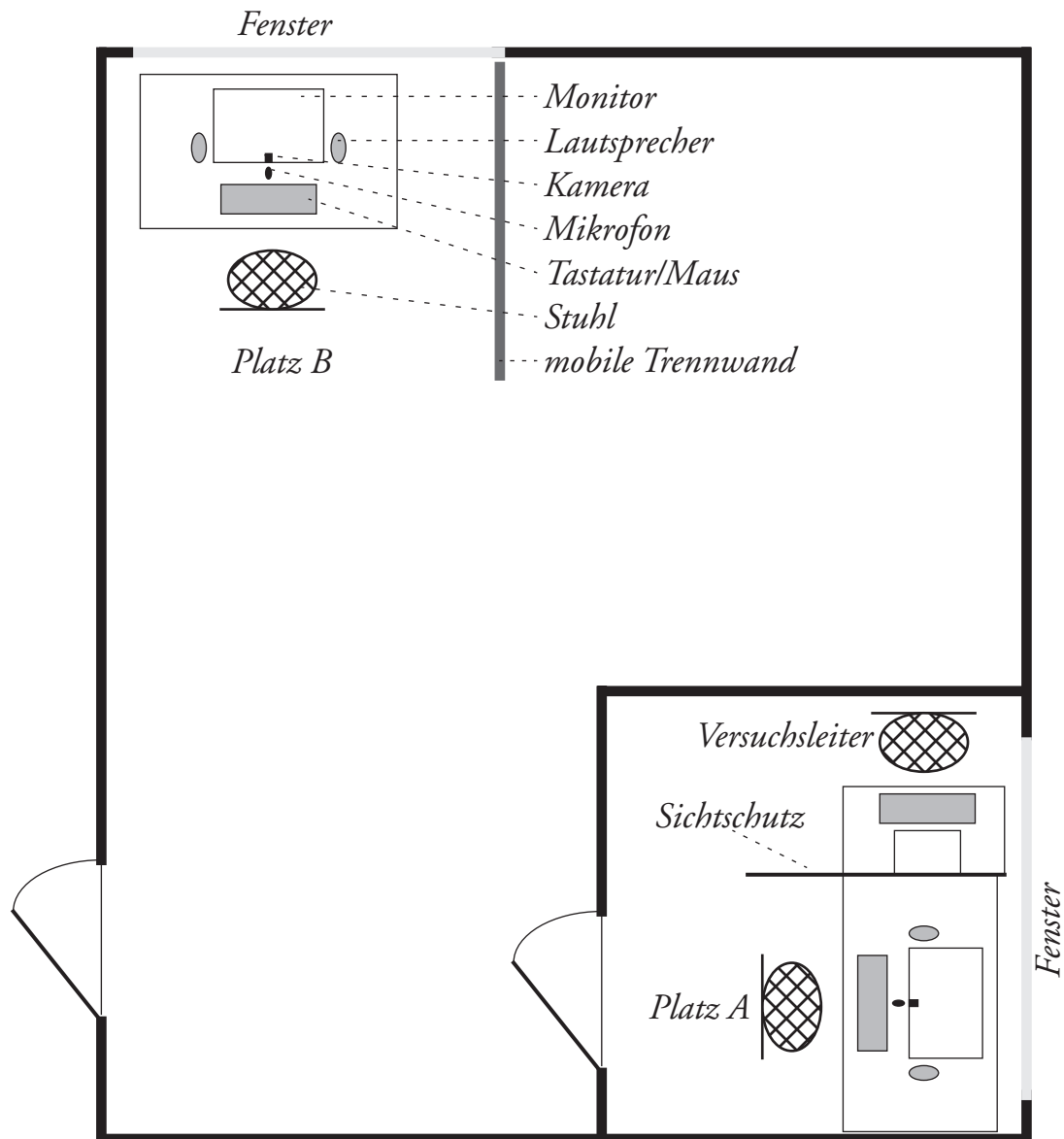
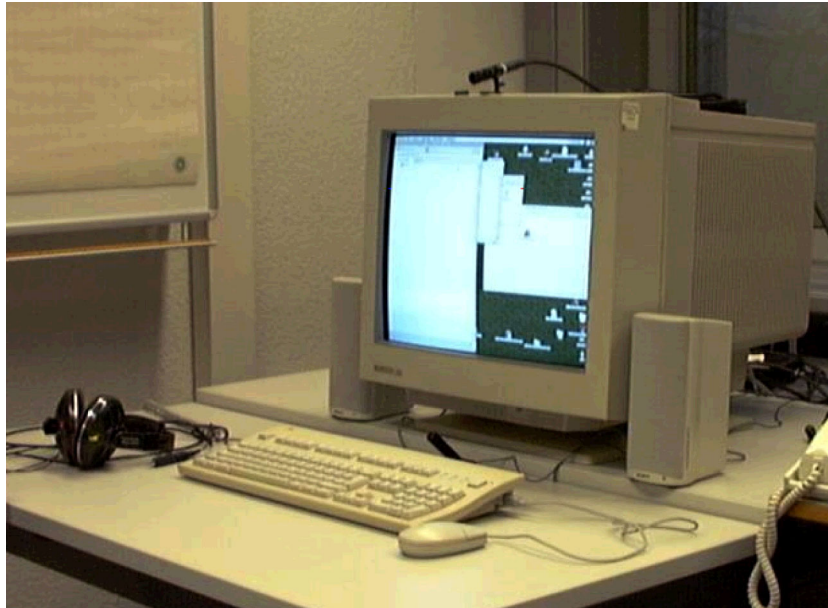


Abbildung 6.11: Räumliche Situation des Versuchsaufbaus

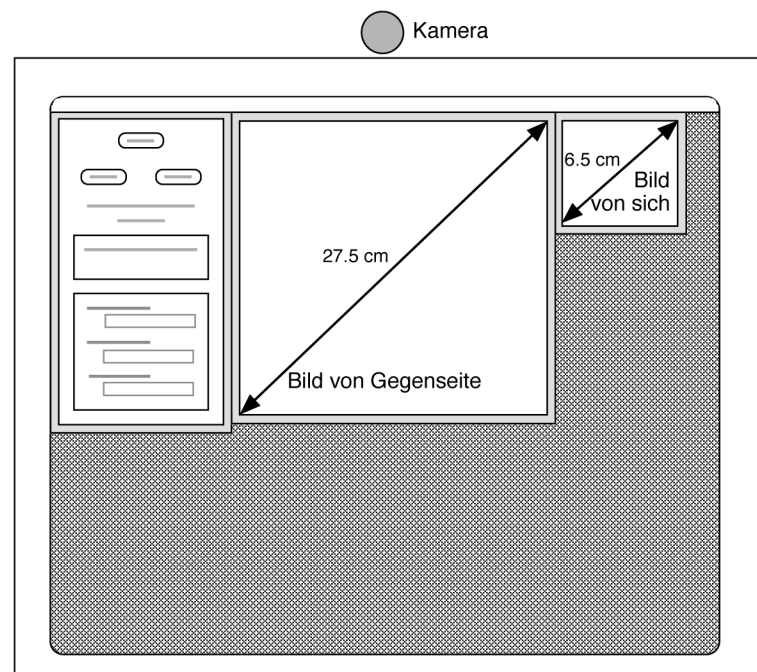
Um eine Vorstellung davon zu bekommen, wie ein Platz in der Wirklichkeit aussah, ist Abbildung 6.12 zu betrachten. Auf dem Tisch steht der Bildschirm, auf welchem oben in der Mitte die Videokamera befestigt ist. Links und rechts sind die Aktivlautsprecher angebracht. Auf dem Tisch, in der Mitte des Monitors, befindet sich das Richtmikrofon.





*Abbildung 6.12: Workstation (Platz B)*

Abbildung 6.13 zeigt Position und Dimension der im Experiment vorkommenden Fenster. Am linken Bildschirmrand wurde jeweils das 'Control Panel'-Window positioniert, in der Mitte das Bild der Gegenseite in voller Grösse (ca. 27.5 cm) und auf der rechten Seite das eigene Bild. Es wurde gerade so gross eingestellt (ca. 6.5 cm), dass die Testperson kontrollieren konnte, wie sie sich beim Gesprächspartner präsentiert, ohne jedoch vom Geschehen abgelenkt zu werden.



*Abbildung 6.13: Positionierung der Fenster auf dem Bildschirm*

### 6.1.2.3 Ablauf der Versuche

1. Den Testpersonen wurde das Vorgehen erklärt (Sinn des Experiments, Bedeutung der Bildschirmfenster und der Bewertungspunkte).
2. Zwei Testrunden mit je einem Betreuer pro Testperson wurden durchgeführt, um auf Fragen eingehen zu können und bei Problemen zu helfen.
3. Bevor eine Runde begann, waren beide Videobilder (das eigene und das der Gegenseite) abgedunkelt. Durch Drücken auf den 'Weiter'-Knopf wurde das eigene Bild aufgehellt, was das Zeichen für die andere Person war, dass man für die nächste Runde bereit war. Hatten beide auf 'Weiter' gedrückt, so konnte die Runde beginnen.
4. Aufgabe (a) Die Person, welche das zu erratende Sprichwort erklären musste, las die Umschreibung vor und die Unterhaltung begann (Phase 1)<sup>1</sup>. Aufgabe (b) Die Person, welche die zu diskutierende Behauptung vor sich hatte, las sie der anderen vor und die Diskussion begann (Phase 2-4)<sup>1</sup>.
5. War eine Minute verstrichen, wurden die Testpersonen vom Testleiter aufgefordert die Bewertung durchzuführen, was mit dem Drücken auf den Knopf 'Bewertung' eingeleitet wurde. Die Videobilder waren wieder abgedunkelt.
6. Während die Testpersonen die Bewertung durchführten, wurden die neuen Verzögerungen und Bildwiederholraten eingestellt, was dazu führte, dass kurzzeitig (ca. 10 s) die Tonverbindung ausfiel und das Videobild einfro. Anschliessend ging es mit Punkt zwei weiter, bis das Experiment nach zwei Proberunden und 27 'ernsten' Runden beendet war.
7. Im Anschluss an das Experiment wurden die Testkandidaten zu ihren Erfahrungen befragt und es wurden Anregungen und Verbesserungsvorschläge entgegengenommen.

## 6.1.3 Ergebnisse

### 6.1.3.1 Messungen

Im Verlauf der Versuchsreihe stellte es sich heraus, dass die Wahl der richtigen Parameter-Bereiche neben der Auswahl zweckmässiger Tasks die wohl schwierigste Aufgabe ist. Insgesamt wurden vier verschiedene Test-Setups verwendet, die im folgenden als Phase 1 bis 4 beschrieben werden. Die nachfolgend angegebenen Verzögerungswerte sind Zweiwegverzögerungen im Sinne von Abbil-

---

<sup>1</sup> Die erste Versuchsreihe wurde in vier Phasen unterteilt, in denen unterschiedliche Setups verwendet wurden. Auf diese wird nachfolgend genauer eingegangen.

dung 5.9. Diese beinhalten also sowohl die künstlich hinzugefügten Verzögerungswerte auf Hin- und Rückweg als auch die Zeiten, die das Versuchssystem zur Signalverarbeitung benötigt (40 ms pro Workstation).

### *Phase 1 (Experimente 1-3)*

In der ersten Phase wurden die Experimente mit dem Task (a) *Sprichwortraten* durchgeführt. Die Testparameter hatten folgende Grössen:

Framerate (Bilder/s):	16.66	12.5	10
Video Delay (ms):	160	200	240
Audio Delay (ms):	200	300	400

Diese neun Werte wurden für jedes Experiment zufällig zu 27 verschiedenen Kombinationen zusammengestellt. Als weiteres wichtiges Ergebnis dieser ersten Phase stellte sich heraus, dass der Zeitbedarf für das Lösen der ausgewählte Aufgabe sehr stark von den einzelnen Testpersonen abhing; offensichtlich war es gar nicht möglich, die Rategeschwindigkeit als Bewertungsgrundlage miteinzubeziehen.

### *Phase 2 (Experimente 4-8):*

Durch die Erfahrungen aus Phase 1 musste das Experiment umgestellt werden. Das Ratespiel wurde durch eine einminütige Diskussion ersetzt. Als Diskussionsgrundlage wurde jeweils eine provokative Behauptung oder These auf dem Bildschirm dargestellt, was eine spontane und ungezwungene Unterhaltung auslösen sollte. Mit diesem Taskwechsel wurde erreicht, dass die Teilnehmer wirklich eine angeregte Videokonferenz führten und auch von der Videoverbindung Gebrauch machten. Allerdings konnte so nicht mehr die Zeit als Messgrösse verwendet werden, die die Kandidaten zum Lösen der Aufgabe benötigten.

Bei den kurzen Besprechungen mit den Testpersonen im Anschluss an die Experimente von Phase 1 stellte sich heraus, dass die Unterschiede in den einzelnen Settings nicht oder nur schlecht erkannt wurden. Aus diesem Grunde wurden die Einstellungen so abgeändert, dass folgende zwei Problemkreise der Verzögerung betrachtet wurden: die zeitliche Differenz zwischen Bild und Ton

und die absolute Verzögerung. Die Bildwiederholraten aus Phase 1 wurden beibehalten. Dies führte zu folgenden Einstellungen:

Framerate (Bilder/s):	16.66	12.5	10
-----------------------	-------	------	----

Video Delay (ms):	160	220	280	0	0	0	480	680	880
Audio Delay (ms):	0	0	0	200	400	600	400	600	800

### *Phase 3 (Experimente 9-11):*

Nach rund der Hälfte aller Experimente wurde der bisherige Verlauf der Versuche erneut besprochen. Dabei wurde festgestellt, dass in Phase 2 der Aspekt der Verzögerungsdifferenz, unter Hinzunahme einer konstanten Verzögerung der beiden Medien, nicht zum Tragen kam. Da aus den ersten Experimenten bereits bekannt war, dass auf Audioverzögerungen weniger stark reagiert wird als erwartet, wurde diese Verzögerung im Verhältnis höher angesetzt. In den vorherigen Phasen wurde auf die Bildwiederholrate nur begrenzt reagiert. Aus diesem Grund wurde die Bildwiederholrate mit tieferen Werten angepasst. Die neun neuen Werte, welche in der folgenden Tabelle dargestellt sind, wurden wiederum zu 27 verschiedenen Einstellungen kombiniert.

Framerate (Bilder/s):	16.66	8.33	3.85
Video Delay (ms):	280	580	1280
Audio Delay (ms):	200	600	1400

### *Phase 4 (Experimente 12-18):*

Als Ergebnis aus Phase 3 wurde festgestellt, dass der tiefste Wert der Bildwiederholrate einen allgemein inakzeptablen Wert aufwies, so dass die Testpersonen in diesem Fall nicht genügend geduldig waren, die Dauer von einer Minute abzuwarten, da sie sich durch die schlechte Qualität zu stark gestört fühlten und sofort die Bewertung ausführen wollten. So wurde beschlossen, die Bildwiederholrate, unter Beibehaltung der Verzögerungen aus Phase 3, auf folgende Werte zu setzen:

Framerate (Bilder/s):	16.66	10	6.25
-----------------------	-------	----	------

Somit wurden in den letzten sieben Durchgängen die folgenden Werte verwendet:

Framerate (Bilder/s):	16.66	10	6.25
Video Delay (ms):	280	580	1280
Audio Delay (ms):	200	600	1400

### 6.1.3.2 Die Online-Bewertung durch die Teilnehmer

Die Bewertung der Sessions ist in zwei Teile gegliedert. Zum einen geht es um den Gesamteindruck, also die Qualität im allgemeinen, und zum anderen um das Verhältnis zwischen Bild und Ton.

#### Gesamteindruck:

Mit der Frage: *War eine natürliche Unterhaltung möglich?* sollten die Testkandidaten beurteilen, wie gut ihnen die Qualität der Session gefiel. Die fünfstufige Skala reichte in Anlehnung an die des MOS (Mean Opinion Square) von 'sehr gut' (++) bis 'sehr störend' (- -).

#### Verhältnis zwischen Bild und Ton:

Mit der zweiten Frage: *Haben Bild und Ton zusammengepasst?*, welche eine Spezialisierung der ersten Frage darstellt, sollte herausgefunden werden, ob den Testkandidaten die zeitliche Differenz zwischen Video und Audio aufgefallen ist. Wiederrum konnte in fünf Stufen angegeben werden, wie gut die Synchronität empfunden wurde. Als Erweiterung dazu kam die dritte Frage: *Wie standen Bild und Ton zueinander?* Mit Hilfe dieser sollte herausgefunden werden, ob die Kandidaten beurteilen können, welches Medium mehr verzögert wurde. Es standen vier Antwortmöglichkeiten zur Auswahl: *Bild vor Ton, gleichzeitig, Ton vor Bild, weiss nicht.*

### 6.1.3.3 Resultate des Abschlussgesprächs

Die folgende Aufstellung gibt wichtige Erkenntnisse wieder, welche in der kurzen Besprechung nach den Experimenten gewonnen wurden und somit auch statistisch gesehen eine gewisse Relevanz haben.

- Bei subjektiv schlechter Qualität hat die Kommunikation via Videokonferenz eher den Charakter eines Telefongesprächs, da das Bild des Gesprächspartners nicht beachtet wird und man sich nur auf den Ton konzentriert.
- Die Kommunikation per Videokonferenz ist einengender und anstrengender als eine natürliche Unterhaltung von Angesicht zu Angesicht, da man die ganze Zeit einen fest vorgegebenen Ausschnitt des Gesprächspartners beobachten muss.
- Es braucht eine gewisse Gewöhnungsphase an das System; besonders zu Anfang wird die Versuchsanordnung unnatürlich wahrgenommen. Die Testpersonen gewöhnten sich aber mit der Zeit daran und fühlten sich sogar von einer schlechten Qualität weniger gestört als zu Beginn.
- Für einen kurzen Einsatz muss die Qualität nicht unbedingt perfekt sein. Wenn man aber längere Zeit mit dem System arbeiten muss, werden höhere Ansprüche gestellt. Das gelegentliche Führen einer privaten Videokonferenz mit einem Bekannten in den USA stellt somit beispielsweise nicht die gleichen Qualitätsanforderungen wie eine häufige geschäftliche Nutzung.
- Die Frage hinsichtlich relativer Verzögerung war bei dieser Versuchsanlage sehr schwierig zu beantworten. Selbst wenn sich eine klare Asynchronität abzeichnete, musste man sich intensiv darauf konzentrieren, um sagen zu können, welches Medium mehr verzögert wurde.
- Es wurde als sehr störend empfunden, dass man sich nicht in die Augen schauen und die andere Person immer nur mit nach unten geneigtem Kopf sehen konnte.
- Eine Minute war oft sehr knapp, um die Qualität des Systems beurteilen bzw. um sie in Relation zu den anderen Qualitätseinstellungen setzen zu können.

Im Verlauf der Versuche haben sich einige Erfahrungen und Erkenntnisse rund um die Experimente ergeben, welche nachfolgend aufgelistet werden.

- Als eigentliche Schwierigkeit für das Erarbeiten der Experimente hat sich der Schritt von der Festlegung der zu untersuchenden Parameter zur Definition der Durchführung der Experimente erwiesen. Die erste Umsetzung mit Aufgabe (a) war grundsätzlich eine gute Sache. Die Rategeschwindigkeit hätte

als Qualitätsmerkmal in die Auswertung miteinbezogen werden können. Als sich dann aber herausstellte, dass die Geschwindigkeit beim Erraten der Sprichwörter im wesentlichen vom Vorwissen und nicht von der Verbindungsqualität abhing, musste festgestellt werden, dass die Überlegungen im Vorfeld nicht zutrafen. Bei zukünftigen Tests sollten somit vermehrt Vortests darüber Aufschluss geben, ob ein Task wirklich geeignet ist oder nicht.

- Der dreidimensionale Raum war wie angenommen viel zu gross; selbst eine grobe Abschätzung ist so nur schwierig möglich. Man sollte sich auf weniger Parameter beschränken, dafür aber die Auflösung feiner abstimmen.
- Die Erfahrungen aus der ersten Versuchsreihe haben ergeben, dass sich die Aufmerksamkeit der Testpersonen bei den Videokonferenzen nach ca. 45 Minuten spürbar verringert. Dies sollte bei weiteren Versuchen mit Probanden beachtet werden.
- Am Schluss der Experimente konnte fast immer festgestellt werden, dass die Kandidaten Spass an der Aufgabe hatten. Dies ist nach Ansicht des Versuchsleiters eine wichtige Grundvoraussetzung, um aussagekräftige Ergebnisse zu erhalten. Würden die Kandidaten nur widerwillig am Experiment teilnehmen, so würde dies die Resultate verfälschen, da nicht damit zu rechnen ist, dass Videokonferenzen im „normalen Leben“ widerwillig durchgeführt würden.
- Die Zusammensetzung der Testgruppen hat ebenfalls einen nicht zu vernachlässigenden Einfluss auf die Resultate. Deshalb sollte gleich in der Planungsphase mit geeigneten Massnahmen sichergestellt werden, dass eine derartige Beeinflussung der Ergebnisse vermieden werden kann.
- Die „Ehrfurcht“ der Teilnehmer vor den anfangs ungewohnten und unbekanntem Geräten und Funktionen hat dazu geführt, dass sehr diszipliniert ablaufende Unterhaltungen geführt wurden und infolge dessen der absoluten Verzögerung eine subjektiv geringere Relevanz beigemessen wird.

#### *6.1.4 Diskussion*

Wie bereits erwähnt, war es nicht das Ziel dieser ersten Versuche, Testreihen mit statistisch korrekten und abschliessenden Aussagen über die untersuchten Teilgebiete zu machen. Mit den 18 Testgruppen konnte von Phase 1 bis 4 ein genaueres Bild des zu untersuchenden Versuchsraumes erarbeitet werden. Die statistische Auswertung im Abschnitt 6.1.4.1 zeigt Tendenzen auf, welche in den ausgedehnteren Versuchsreihen 2 bis 4 aufgegriffen und verfeinert wurden. Es hat sich bei der Auswertung der Daten herausgestellt, dass die Bewertung der Experimente durch die 'Experten' gegenüber den anderen Teilnehmern

keine erkennbaren Unterschiede hervorbrachte, weshalb die beiden 'Experten-  
gruppen' zusammen mit den anderen ausgewertet wurden.

#### *6.1.4.1 Statistische Auswertung*

Zur statistischen Auswertung gelangten lediglich die Phasen 2 und 4. Aufgrund des veränderten Test-Setups musste die Auswertung von Phase 1 ignoriert werden. Phase 3 wurde wegen unbefriedigender Parameter und der zu geringen Anzahl Testgruppen ebenfalls nicht ausgewertet. Nicht zur Auswertung herangezogen wurde auch die zweite und dritte Frage der subjektiven Bewertung (Unterabschnitt 6.1.3.2), da diese keine aussagekräftigen Informationen lieferten. Speziell bei der dritten Frage wurde von den Probanden oft „geraten“. Die Auswertung der Ausgabedateien bestätigte dieses Verhalten. Die folgenden Graphiken stellen eine Auswahl aller Statistiken dar. Die kompletten Grafiken finden sich in [69]. In den Balkendiagrammen ist jeweils der Mittelwert angegeben, wobei '5' der Bewertung '++' entspricht und '1' der Bewertung '--'.

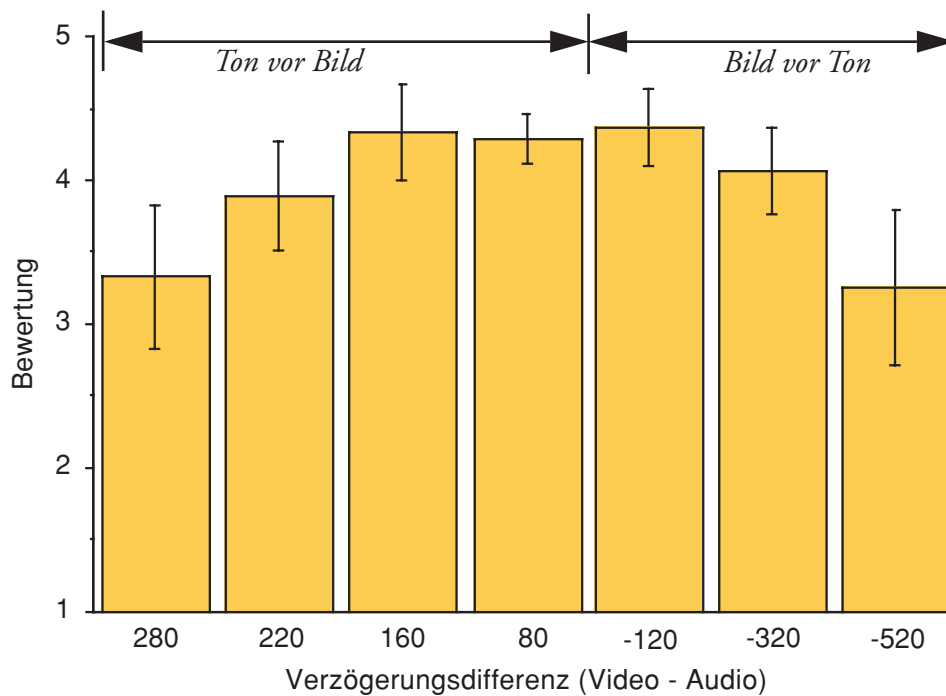
#### *Phase 2: Verzögerungsdifferenz*

In Abbildung 6.14 ist die Auswertung von Phase 2 mit Blick auf die Verzögerung zwischen Bild und Ton dargestellt<sup>1</sup>. Auf der Abszisse ist die Differenz von Video- minus Audioverzögerung angegeben, wobei die Angaben von 80 ms durch die Verzögerungspaare 480/400 ms, 680/600 ms und 880/800 ms erzeugt wurden. Es ist zu beachten, dass eine Verzögerungsdifferenz von 140 ms in beide Richtungen (+160 ms bzw. -120 ms) von den Testpersonen als etwa gleichwertig wie die kleinste Differenz bewertet wurde. Bei der Videoverzögerung scheint dieser Wert jedoch gerade ein Grenzwert zu sein, denn 60 ms mehr bedeuten eine merkliche Verschlechterung der Qualität. Anders sieht es im Audiobereich aus. Eine um 200 ms grössere Audioverzögerung bringt etwa eine ähnliche Verschlechterung mit sich wie zusätzliche 80 ms im Videobereich. Audioverzögerungen von über 200 ms liegen scheinbar noch in einem akzeptablen Bereich. Die Verzögerung des Bildes von 280 ms wird klar als schlechter empfunden. Für die Tonverzögerung lässt sich aus diesen Testreihen keine Grenze angeben, denn es ist anzunehmen, dass -520 ms einen zu grossen Wert darstellen und sich bereits bei kleineren Werten klare Qualitätseinbussen ergeben würden.

---

<sup>1</sup> Die dargestellten Balken geben jeweils die Mittelwerte der gemessenen Verzögerungswerte an. Zusätzlich ist das Intervall angegeben, in denen 95% der gegebenen Antworten lagen.





*Abbildung 6.14: Verzögerungsdifferenzen der Phase 2*

#### *Phase 4: Verzögerungsdifferenz*

Wie bereits in Phase 2 ist auch hier die Differenz der Verzögerungen besonders interessant. In Abbildung 6.15 ist jeweils die effektive Differenz oberhalb des Balkens angegeben. Die Folgerungen aus Phase 2 bestätigen sich grösstenteils. Es ist lediglich erstaunlich, wie gut die Audioverzögerung von -320 ms aufgenommen wurde (vierter Balken). Es kann sein, dass die Testpersonen durch die allgemein sehr grossen Verzögerungen (im Vergleich zu in der Literatur angegebenen Werten) an die Qualität „gewöhnt“ wurden und sich deshalb auch durch 320 ms Differenz nicht gestört fühlten. Das erstaunlichste Resultat wird im letzten Balken sichtbar. Selbst eine nahezu synchrone Verzögerung von über 1200 ms von Audio und Video wurde von den Testpersonen nicht als Verschlechterung der Qualität beurteilt. Dieses Resultat lässt sich einerseits damit erklären, dass sich die Testpersonen sehr diszipliniert miteinander unterhielten und sich nicht gegenseitig ins Wort gefallen sind, sondern den Kommunikationspartner immer aussprechen liessen und ihm eine gewisse Bedenkzeit einräumten. Oder aber die multimodale Videokonferenz beanspruchte derart viel Aufmerksamkeit, dass die Toleranzgrenze automatisch anstieg. Bei intensiver Diskussion über die eingestellten Werte machten sich bei den Testleitern in eigenen Vortests bereits Zweiwegverzögerungen von 600 ms als störend bemerkbar, da sie sich gegenseitig ins Wort gefallen sind. Diese Beobachtung bestätigt letztere Vermutung.

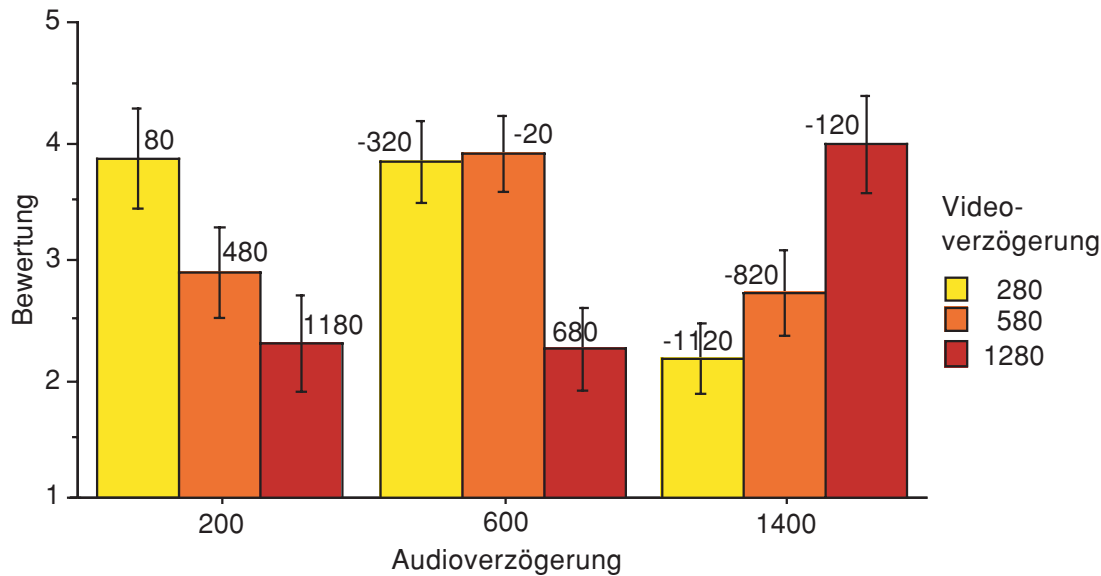


Abbildung 6.15: Verzögerungsdifferenzen der Phase 4

### Betrachtung aller drei Dimensionen

Abbildung 6.16 zeigt die Zusammenhänge aller drei Testparameter in Phase 4. Die neun Balken der vorherigen Abbildung sind nun in je drei weitere aufgeteilt, so dass der Einfluss der Bildwiederholrate sichtbar wird. Bei den Dreierblöcken 1, 5 und 9 handelt es sich jeweils um kleine Verzögerungsdifferenzen zwischen Bild und Ton. In diesen Fällen zeigt sich die schlechtere Qualität durch die niedrigere Bildwiederholrate am deutlichsten. Die Werte fallen von der höchsten zur niedrigsten Bildwiederholrate um rund 1.5 Punkte in der Bewertungsskala ab. Allgemein ist festzustellen, dass die schlechteste Bildwiederholrate jeweils die schlechteste Bewertung erhielt. Eine Ausnahme ist im vierten Balkentrio festzustellen. Es ist anzunehmen, dass dieser Effekt durch die geringe Anzahl von Testpersonen aufgetreten ist. Eine bemerkenswerte Feststellung kann man bei der Betrachtung der Blöcke 3, 6, 7 und 8 machen. Bei einer sehr grossen Differenz zwischen Bild und Ton ist jeweils der mittlere Balken des Trios im Durchschnitt am besten bewertet worden. Daraus lässt sich die Vermutung ableiten, dass bei deutlichen Qualitätseinbussen durch grosse Verzögerungswerte eine Bildwiederholrate von 16.66 Bildern pro Sekunde keinen besseren subjektive Qualitätseindruck mehr hinterlässt als bei 8.33 Bildern/s. Aus diesen Erkenntnissen wurde abgeleitet, dass in einem nächsten Schritt die absolute Verzögerung bei einer Videokonferenz zu untersuchen ist bei gleichzeitiger zeitlicher Synchronität der beiden Medienströme sowie einer ausreichend guten Bildwiederholrate von 25 Bildern pro Sekunde. Es wurde auch klar, dass das verwendete Versuchssystem sich eignete, um weitere Versuchsreihen durchzuführen (Frage F1.1).

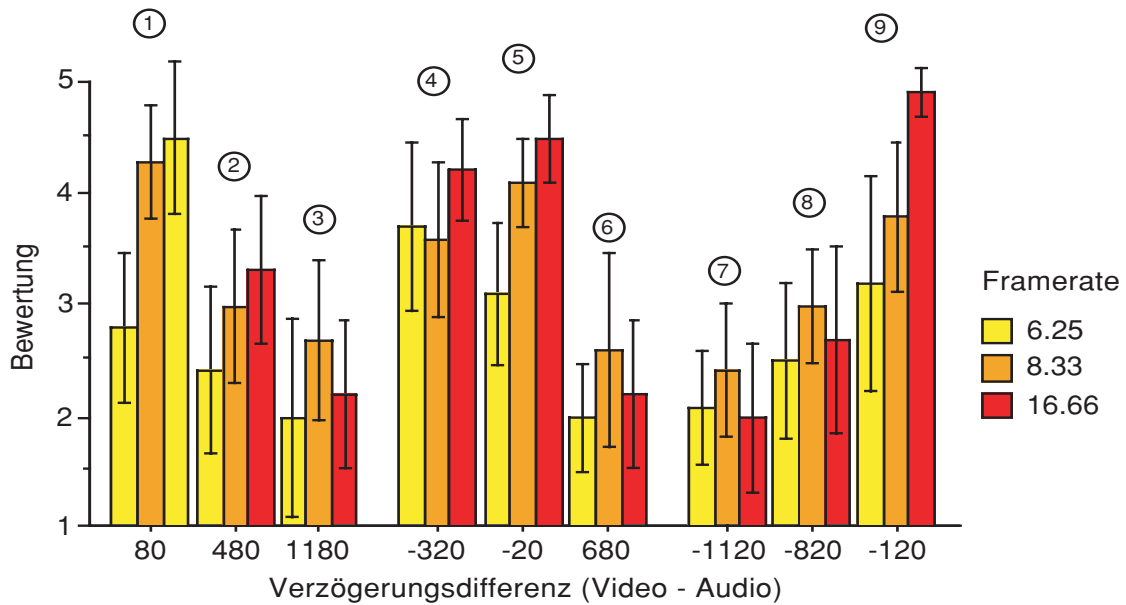


Abbildung 6.16: Phase 4: Alle Dimensionen

#### 6.1.4.2 Ergebnisse

- Es sollte möglichst keine Asynchronität zwischen Audio- und Videosignal bestehen, wobei Differenzen von weniger als 160 ms als nicht störend wahrgenommen werden.
- Der Mensch neigt dazu, Asynchronitäten, bei denen das Bild länger verzögert wird als der Ton, schneller als unangenehm zu empfinden als umgekehrt. Die aufgestellte Hypothese H2.1 konnte somit bestätigt werden.
- Es hat sich anhand der durchgeführten Experimente gezeigt, dass bei für gut befundener Verzögerungsqualität eine Bildwiederholrate von 10 Bildern/s nur schwach negativ auffiel. Dies kann aber am statischen Bildinhalt liegen, bei dem sich praktisch nur die Lippen bewegen.
- Bei einer synchronen Verzögerung von Bild und Ton (Differenz < 160 ms) haben die Testpersonen selbst bei Werten von 1400 ms keine Verschlechterung der Qualität festgestellt, wenn die Versuchssituation zu disziplinierten Gesprächen führt.
- Es gibt einen Maskierungseffekt zwischen den verschiedenen qualitätsbeeinträchtigenden Mechanismen. Das bedeutet, dass eine qualitative Verbesserung von Parameter A nicht unbedingt zu einer subjektiv besseren Qualität führt, falls Parameter B für den momentanen Eindruck dominiert. Erst wenn dieser auch verbessert wird, steigt die wahrgenommene Qualität an.

## 6.2 *Versuchsreihe 2*

### 6.2.1 *Ziele*

An den zweiten Versuchsdurchgang wurden insbesondere folgende Ansprüche gestellt:

- Untersuchen des Einflusses von absoluter Verzögerung bei Videokonferenzen auf die Qualitätswahrnehmung der Benutzer und auf deren Leistungsverhalten beim Lösen bestimmter Aufgaben
- Verifikation bzw. Falsifikation der Hypothesen H2.1 bis H2.7
- Finden von Antworten auf die Fragen F2.1 und F2.2

### 6.2.2 *Methode*

Auch im zweiten Versuchsdurchgang wurde die Methode mit einem Maximum an Effizienz gewählt, ohne jedoch die statistische Gültigkeit zu beeinträchtigen. Da sich drei der Hypothesen mit Unterschieden bei den Tasks auseinandersetzen, wurde besonderes Augenmerk auf deren Auswahl gelegt. Sie sollten so realistisch wie möglich und trotzdem experimentell aussagekräftig sein. Die Auswahl der Teilnehmer erfolgte unter der Vorgabe, dass sich männliche und weibliche Teilnehmer sowie solche mit guten und schlechten Netzwerkkennnissen die Waage halten sollten. Die Dauer eines Versuchsdurchgangs war begrenzt, weshalb zur schnellen Annäherung ein effizienter Algorithmus gewählt wurde (Best-PEST, siehe 5.1.4.1). Nicht berücksichtigt werden konnte in der Folge der Wunsch nach der Messmethode *2AFC* (*2-alternative-forced-choice*). Diese hätte bedingt, dass die Tests jeweils die doppelte Zeitdauer in Anspruch genommen hätten, was den Probanden nicht zugemutet hätte werden können und ausserdem die Ergebnisse durch Ermüdung der Teilnehmer verfälscht hätte.

Netzwerk- und Computerkenntnisse	Task 1	Task 2
hoch	15	15
niedrig	15	15

*Tabelle 6.6: Geplante Verteilung von Versuchsteilnehmern über die beiden Tasks*

Die Versuchsteilnehmer wurden wiederum im universitären Umfeld rekrutiert. Via einer speziellen Webseite wurden Anmeldungen entgegengenommen, bei denen folgende Daten erhoben wurden:

- Alter
- Grad der Extrovertiertheit entsprechend dem *Meyers-Briggs-Type-Indicator* [8]

Der Grad der Computer- und Netzwerkkenntnisse wurde vor den eigentlichen Versuchen anhand von speziellen Fragebogen ermittelt. In Tabelle 6.6 ist die geplante, ideale Verteilung der Versuchsteilnehmer über die beiden Tasks dargestellt.

### 6.2.2.1 Auswahl des Tasks

Für diesen Versuch wurden zwei unterschiedliche Tasks ausgewählt. Einer sollte einer Zusammenarbeit, der andere einer Verhandlungssituation entsprechen. Ersterer wurde zum *Legospiel*, letzterer zum *Reisespiel*.

Das Legospiel ist für die Versuchsteilnehmer sehr einfach zu erklären und läuft folgendermassen ab: Beide Versuchspersonen haben je ein Legomodell sowie die Anleitung zum Modell, dessen Bauteile der Partner besitzt. Das Ziel ist es, mit Hilfe der Videokonferenz beide Modelle möglichst schnell korrekt zusammenzubauen. Der Zeitrahmen für beide Modelle beträgt 15 Minuten. Der Vorteil dieses Tasks gegenüber anderen (wie z.B. dem Kartenspiel) ist das aktive Benutzen des Videokanals, indem das Modell nach jedem Bauschritt gezeigt werden kann und der Partner mit der Anleitung kontrollieren kann, ob die Teile richtig zusammengefügt wurden. Da während des Spiels die Rollen des Konstruierens und Anweisungen-Gebens getauscht werden, ist zudem ein sehr symmetrischer Kommunikationsablauf gewährleistet. Die Vorteile dieses Tasks sind folgende: Das Spiel macht Spass, die Bekanntheit von Lego ist derart gross, dass alle Personen auf Anhieb verstehen sollten, wie die Klötzchen zusammenzustecken sind und die Anleitung zu lesen ist. Ausserdem hat man ein grosses Interesse am richtigen Ergebnis, da es kein theoretisches Gedanken-spiel, sondern ein physisches Modell zum Anfassen ergibt. Als Leistungskennwert wurde gemessen, wie weit die Modelle in der vorgegebenen Zeit fertiggestellt werden konnten.

Das Reisespiel ist bis auf wenige Änderungen mit dem „Travel Game“ von Anderson [1] identisch. Bei diesem Spiel gibt es einen Reisenden sowie einen Agenten eines Reisebüros. Die Aufgabe des Reisenden ist es, innerhalb von 15 Minuten so viele Städte wie möglich zu besuchen, ohne eine doppelt anzureisen. Als Hilfsmittel besitzt er eine Landkarte mit eingezeichneten Städten. Der Agent ist im Besitz einer Liste mit möglichen Städteverbindungen. Dieser Task wurde ausgewählt, da er eher auf einer Verhandlungsstrategie basiert und somit

hilfreich ist für die Stützung der These H2.1. Zudem wurde er schon mehrfach bei anderen Untersuchungen angewandt, was einen Vergleich der Ergebnisse mit denen anderer Untersuchungen erlaubt.

### 6.2.2.2 Fragebogen

Die Teilnehmer mussten sowohl vor als auch nach den Tests einen Fragebogen ausfüllen. Der erste Bogen diente insbesondere der Erfassung der Computererfahrung als auch der aktuellen Verfassung der Versuchspersonen. Der Fragebogen nach dem Versuch zielte darauf ab, Erfahrungen und Hinweise der Versuchspersonen festzuhalten und später auswerten zu können. Der vollständige Fragebogen findet sich in [59].

### 6.2.2.3 Versuchsaufbau

Der Versuchsaufbau geschah analog dem aus Versuchsreihe 1 (siehe Abbildung 6.17). Dies sollte sicherstellen, dass die Ergebnisse möglichst gut untereinander vergleichbar sein sollten. Zusätzlich wurden die beiden Audio- und Videosignale der Workstations für eine Offline-Auswertung aufgezeichnet und archiviert.

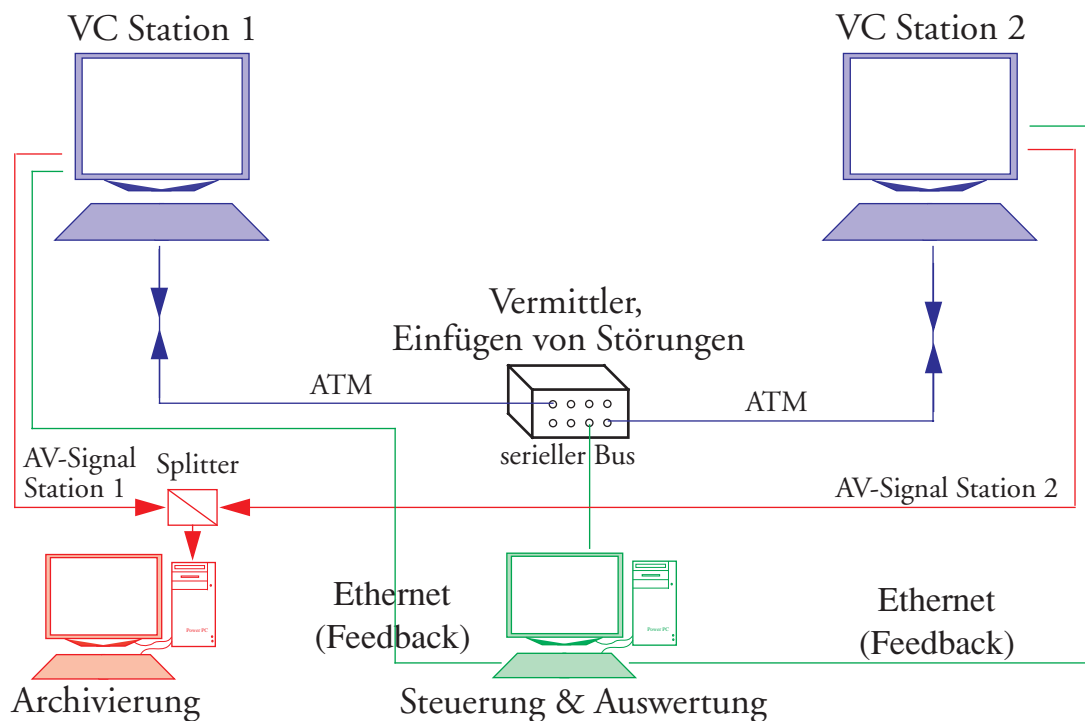


Abbildung 6.17: Vernetzung der Komponenten

Einzig der Platz für den Versuchsleiter und den Beobachter wurde neu gewählt, damit keiner der beiden Versuchsteilnehmer sich gestört fühlen konnte. Der Grundriss in Abbildung 6.18 zeigt die Anordnung der Versuchsplätze und der Kontrollstation. Zusätzlich wurde ein Konferenztisch eingesetzt, an dem die Nachbesprechungen mit allen Beteiligten stattfanden.

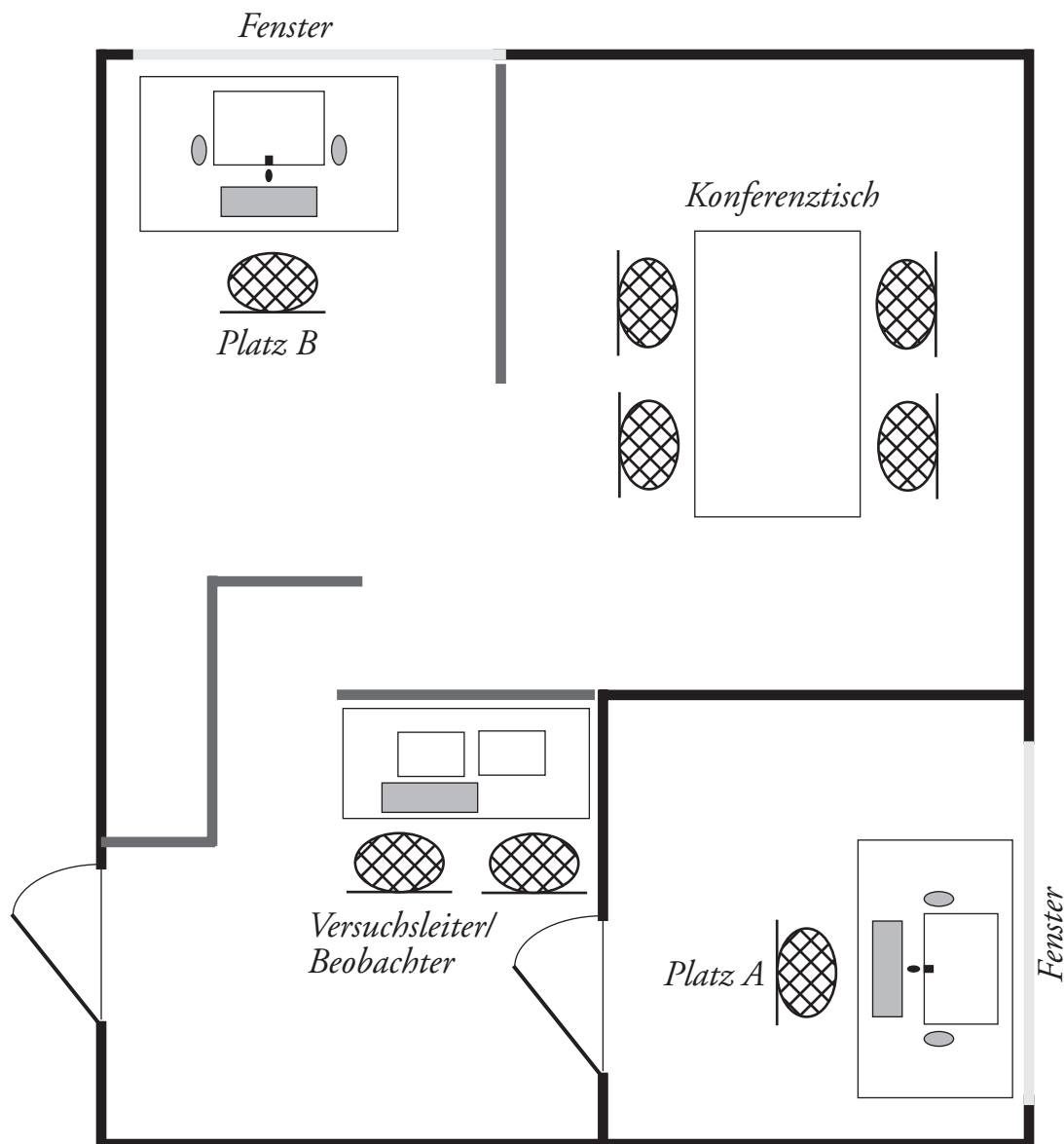


Abbildung 6.18: Räumliche Situation des Versuchsaufbaus

#### 6.2.2.4 Ablauf der Versuche

Die Versuchsprozedur bestand aus folgenden Teilen:

- Einleitung (10 Minuten): Die beiden Versuchspersonen wurden am Konferenztisch begrüßt und in den Ablauf eingewiesen. Um eine ungezwungene Atmosphäre zu schaffen, wurde zuerst das Experiment als Teil einer grösseren Arbeit vorgestellt und es wurde Mineralwasser angeboten. Die Tasks wurden im Detail besprochen und die Versuchsumgebung erklärt. Danach wurden die beiden Personen an ihre Plätze geschickt und es wurden zwei Probedurchgänge gestartet, so dass sie sich an die Videokonferenz gewöhnen konnten. Vor dem Beginn des ersten Tasks musste der erste Fragebogen ausgefüllt werden.
- Task 1 (30 Minuten): Die Reihenfolge von Lego- und Reisespiel erfolgte zufällig. Zu Beginn wurde den Teilnehmern eine Situation ohne Verzögerung als Referenz vorgeführt. Anschliessend wurde die maximale Verzögerung von 1.4 Sekunden (einweg) demonstriert. Der anschliessende Versuch dauerte 14 Minuten, wobei je 7 Werte für die Akzeptanz- und 7 für die Wahrnehmungsschwelle verwendet wurden. Nach jeder Minute wurden die Teilnehmer gebeten, auf die Frage zu antworten, ob Sie eine Verzögerung wahrgenommen hatten oder nicht. Wenn dies der Fall gewesen sein sollte, musste die zweite Frage beantwortet werden, ob die Verzögerung einen störenden Einfluss hatte oder nicht. Aus den Antworten wurde mit dem Best-PEST Algorithmus der nächste Verzögerungswert berechnet, um während des Tasks die entsprechende Schwelle zu ermitteln. Welcher der beiden Probanden für die Steuerung des Algorithmus zuständig war, wurde zu Beginn des Tasks zufällig festgelegt. Nach Beendigung des Tasks musste ein kleiner Fragebogen ausgefüllt werden.
- Task 2 (15 Minuten): Der Ablauf geschah analog zu Task 1
- Fragebogen und Besprechung am Ende (20 Minuten): Nach der Beendigung des zweiten Tasks galt es, einen weiteren Fragebogen auszufüllen (ca. 10 Minuten). Abschliessend diskutierten sowohl Teilnehmer als auch Versuchsleiter über den Ablauf. Hier wurden insbesondere Erfahrungen und Meinungen der Teilnehmer auf emotionaler Ebene registriert.

#### 6.2.3 Ergebnisse

Das auffälligste und wichtigste Ergebnis dieser zweiten Versuchsphase stützt gleichzeitig und in besonderem Masse die Hypothese H2.1: Die gemessene Akzeptanzschwelle bei Verzögerungen lag unerwartet hoch. Obwohl mit dem



Versuchsaufbau eine Einwegverzögerung von 1.4 Sekunden hinzugefügt werden konnte, reichte dies bei der Hälfte der Versuche nicht aus, um die Akzeptanzschwelle mit dem Best-PEST Algorithmus exakt anzunähern. Die Wahrnehmungsschwelle lag mit 800 bis 1000 ms ebenfalls deutlich über den Erwartungen.

### *6.2.3.1 Qualitative Ergebnisse: Reisespiel*

Bei diesem Task hat sich herausgestellt, dass der Nutzen des Videokanals äusserst gering bis nicht vorhanden war. Die meisten Teilnehmer waren der Meinung, dass der Task genausogut per Telefon hätte durchgeführt werden können. Es hat sich gezeigt, dass die eingefügten Verzögerungen die Leistung der beiden Teilnehmer nicht auffällig beeinträchtigte. Dies lag einerseits sicherlich daran, dass der Task keine grosse Interaktivität verlangte (die Teilnehmer studierten die meiste Zeit die Karte mit den Reisezielen), andererseits wurden die Verzögerungen zum grössten Teil nicht bemerkt, weil in dieser Zeit überlegt werden musste und somit gar keine Antwort erwartet wurde. Wenn allerdings interaktive Diskussionen aufkamen, machten sich grosse Verzögerungswerte sofort störend bemerkbar, indem sich die Teilnehmer wiederholen mussten oder sich gegenseitig ins Wort fielen. Der wichtigste Parameter, der die Leistungsfähigkeit bei diesem Task einschränkte, war somit nicht die zusätzliche Verzögerung, sondern vielmehr die Vorkenntnisse über die nordamerikanische Geographie sowie die Begabung, solch eine Reise effizient zu organisieren und mit Flugplänen umgehen zu können.

### *6.2.3.2 Qualitative Ergebnisse: Legospiel*

Im Gegensatz zum Reisespiel wurde der Videokanal bei diesem Task wirklich benötigt, um die Modelle innerhalb kurzer Zeit zusammenzubauen. Davon abgesehen war der Informationsfluss sehr einseitig: Derjenige Teilnehmer mit der Anleitung gab Anweisungen und kontrollierte über den Videokanal, ob sein Partner das Modell richtig zusammensetzte. Somit war ersterer auch wesentlich sensitiver auf Verzögerungen, da er auf Rückmeldungen angewiesen war. Auch bei diesem Task waren die Akzeptanzwerte sehr hoch, was teilweise daran lag, dass beim Suchen der richtigen Legosteine Zeit verstrich und diese fälschlicherweise nicht dem System sondern der „langen Leitung“ des Konstruierenden zugeschrieben wurde. Zusätzlich zu einer Verbindung mit niedrigen Verzögerungswerten war es bei diesem Task leistungssteigernd, wenn man sich mit dem Bau von Legomodellen bereits gut auskannte oder aber Konstruktionsanweisungen effizient abgeben konnte.

### 6.2.3.3 Messergebnisse

Wie bereits erwähnt, traten mit der Best-PEST Methode in einigen Fällen Probleme auf, da die Teilnehmer besonders zu Beginn auch hohe Verzögerungswerte als nicht störend betrachteten und somit der Algorithmus eine hohe Toleranzschwelle annahm, die teilweise oberhalb des technisch Möglichen lag und in den sieben zur Verfügung stehenden Iterationsschritten nur schwierig wieder nach unten zu korrigieren war. Um dies zu kompensieren, wurden die Experimente zusätzlich nach einer holistischen (ganzheitlichen) Methode ausgewertet, indem alle Messwerte wie folgt bewertet wurden:

- Offensichtliche Ausreisser wurden eliminiert (beispielsweise wenn ein Teilnehmer eine Verzögerung von 2800 ms nicht erkannte, obwohl im Verlauf mehrere Werte um 1400 ms sinnvoll erkannt und teilweise sogar als störend empfunden wurden).
- Es mussten mindestens drei Werte existieren, bei denen die Verzögerung durch die Teilnehmer erkannt wurde.
- Für den Fall, dass Verzögerungswerte höher als die vom Teilnehmer bemerkten nicht erkannt wurden, war Vorsicht geboten. Waren diese Werte mengenmässig mehr als doppelt so gross wie die Anzahl der festgestellten, wurde der Datensatz verworfen.
- Bei einem gültigen Datensatz wurde der Schwellwert aus dem Durchschnitt dicht zusammenliegender, gerade noch erkannter Verzögerungswerte und denjenigen gerade nicht mehr erkannter Werte errechnet.

Tabelle 6.7 zeigt die Wahrnehmungsschwellen für beide Tasks, einmal mit der PEST-Methode und einmal holistisch ausgewertet.

Methoden	Wahrnehmung Reise	Wahrnehmung Lego
Best-PEST	1998 ms	1994 ms
holistisch	1800 ms	1904 ms

*Tabelle 6.7: Durchschnittswerte der Wahrnehmungsgrenzwerte*

## 6.2.4 Diskussion

### 6.2.4.1 Hypothesen H2.1, H2.2 und H2.3: Akzeptanzschwelle

Aufgrund der Tatsachen, dass die Akzeptanzschwelle bei beiden Tasks höher zu sein scheint als die 2800 ms, die der Versuchsaufbau erlaubte, und dass die Best-PEST Messmethode in 7 Schritten nicht ausreichte, um nach einer Eingewöhnungsphase einen niedrigeren Wert zuzulassen, kann die Akzeptanzschwelle mit den vorhandenen Messwerten nicht bestimmt werden. Insofern kann über die Hypothesen H2.1, H2.2 und H2.3 keine Aussage gemacht werden.

### 6.2.4.2 Hypothese H2.4: Die Wahrnehmungsschwelle hängt nicht vom Vorwissen über Netzwerke und Computer ab

Die Auswertung der Daten über die Wahrnehmungsschwelle von Teilnehmern mit guten gegenüber denen mit schlechten Computer- und Netzwerkkenntnissen zeigte signifikante Unterschiede bei den Durchschnittswerten sowohl mit der PEST-Methode als auch mit der ganzheitlichen. Teilnehmer mit viel Vorwissen über die zu Grunde liegende Technik waren allgemein weniger tolerant gegenüber absoluter Verzögerung als solche mit wenig Vorwissen (siehe Tabelle 6.8)

Task	Schwelle viel Wissen (ms)	Schwelle wenig Wissen (ms)	ø-Diff. (ms)	Sig.	n <sup>a</sup> (viel/wenig)
Reise PEST	1794	2148	354	0,146	19/26
Reise hol.	1522	2050	528	0,102	9/10
Lego PEST	1846	2098	252	0,23	19/27
Lego hol.	1876	2026	150	0,837	10/12

**Tabelle 6.8:** Durchschnittswerte im Vergleich (Teilnehmer mit viel und wenig Computer- und Netzwerkwissen)

a. n bezeichnet die Anzahl Teilnehmer. Die erste Zahl gibt die Anzahl mit viel, die zweite die mit wenig Computerkenntnissen an.

Diese Ergebnisse widersprechen denen von Sasse et. al (Höheres Netzwerkwissen führe zu höherer Toleranz gegenüber Verzögerungen). Die festgestellten Ergebnisse widersprechen auch der These, weswegen diese ebenfalls fallen gelassen werden muss. Für die gemessenen Werte sind mehrere Interpretatio-

nen denkbar. Erstens ist es möglich, dass Personen, die im Umgang mit Computern geübt sind, sich mehr auf die Verzögerungen konzentrieren können als auf das Bedienen des Systems und somit empfindlicher reagieren. Zweitens könnte es sein, dass solche Personen andere Strategien für die Qualitätsbewertung besitzen als Personen mit wenig Netzwerkkenntnissen und auch eher wissen, auf was es ihnen bei QoS ankommt.

Ein weiteres interessantes Resultat zeigt sich, indem die tolerierte Verzögerung bei Leuten mit viel Computerwissen im Legospiel grösser war als im Reisespiel. Bei den Personen mit weniger Kenntnissen war es genau umgekehrt.

#### 6.2.4.3 Hypothese H2.5: Die Wahrnehmungsschwelle der absoluten Verzögerung liegt tiefer als die Akzeptanzschwelle

Bei beiden Tasks unterscheiden sich Wahrnehmungs- und Akzeptanzschwelle signifikant. Die Hypothese H2.5 wird somit gestützt.

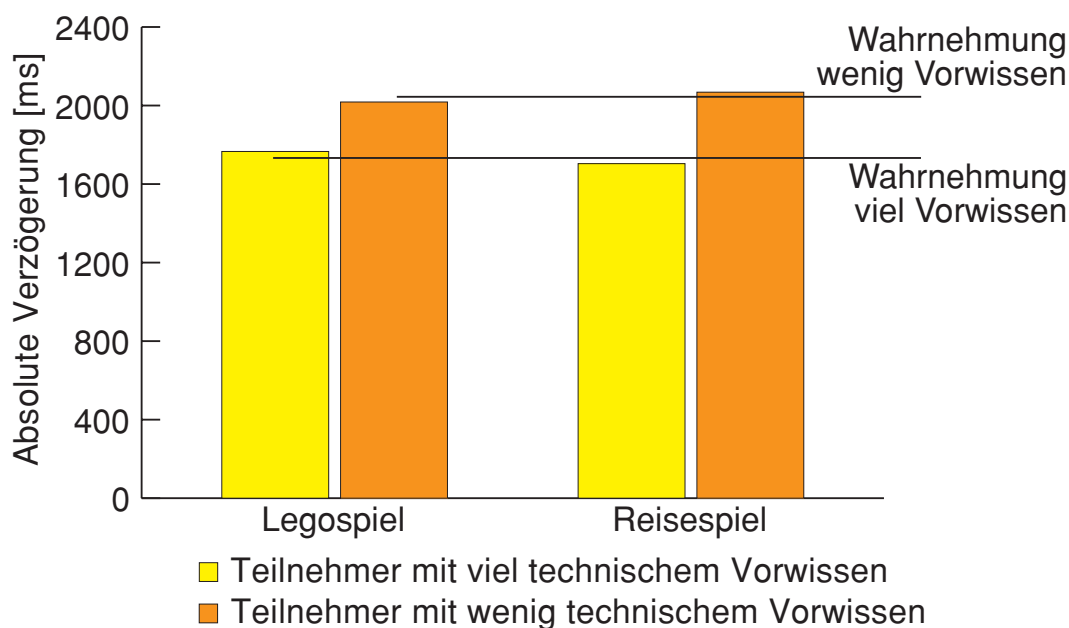


Abbildung 6.19: Ergebnisse der Versuche betreffend Wahrnehmung

#### 6.2.4.4 Hypothese H2.6: Die Wahrnehmungsschwelle der absoluten Verzögerung bei einer Videokonferenz hängt nicht vom spezifischen Task ab.

Bei beiden Tasks liegen die Wahrnehmungsschwellen in einem Bereich, der statistisch gesehen als gleichwertig gewertet werden muss (zwischen 1800 ms und 1998 ms). H2.6 wird somit erwartungsgemäss gestützt.

#### *6.2.4.5 Hypothese H2.7: Gegenseitige Unterbrechungen nehmen ab, wenn die absolute Verzögerung bei einer Videokonferenz zunimmt*

Die Auswertung der Daten zeigte, dass Unterbrechungen in diesem Fall keine gute Messgrösse für den Wechselwirkungsaufwand sind. Es waren maximal acht Unterbrechungen pro Minute und die zeitliche Verteilung über diese Minute ist nicht normal, was eine statistische Analyse verunmöglicht. Alternativ wurden Bestätigungsäusserungen und Sprecherwechsel registriert, wobei nur letztere normalverteilt waren. Es bietet sich also an, die Hypothese neu wie folgt zu schreiben:

#### *H2.7a: Sprecherwechsel nehmen ab, wenn die absolute Verzögerung bei einer Videokonferenz zunimmt.*

Diese Hypothese wurde schon in mehreren Studien untersucht und auch bestätigt [60]. Allerdings ergeben sich aus unseren Werten keine Hinweise auf eine Stützung dieser These. Somit kann H2.7a nicht bestätigt werden und muss verworfen werden.

### 6.3 *Versuchsreihe 3*

An den dritten Versuchsdurchgang wurden unter Ausdehnung der Dialogsituation von zwei auf drei Personen insbesondere folgende Ansprüche gestellt:

- Untersuchen des Einflusses von absoluter Verzögerung bei Videokonferenzen auf die Qualitätswahrnehmung der Benutzer und auf deren Leistungsverhalten beim Lösen bestimmter Aufgaben
- Verifikation bzw. Falsifikation der Hypothesen H3.1 bis H3.5

#### 6.3.1 *Methode*

Ausgehend von den Erfahrungen und Erkenntnissen der vorangehenden Versuchsphasen wurde auch in diesem Versuch Wert auf eine realistische Versuchsumgebung gelegt. Wiederum wurde für eine schnelle Schwellenannäherung der Best-PEST-Algorithmus gewählt. Für weniger interessant hielten wir Fragestellungen, bei denen das Vorwissen der Testkandidaten bezüglich Computern und Netzwerken eine Rolle spielt. Stattdessen wurde noch mehr Augenmerk auf die Auswahl der Tasks gelegt als in Versuchsreihe 2.

Durch die erhöhte Anzahl erforderlicher Teilnehmer (geplant wurden 20 Experimente mit je drei Teilnehmern) musste das Rekrutierungsverfahren auf mehrere Departemente der ETH ausgeweitet werden (Chemie, Biologie, Informationstechnologie und Elektrotechnik, Umweltwissenschaften, Agrar- und Lebensmittelwissenschaften). Es wurde ein ausgewogenes Verhältnis zwischen weiblichen und männlichen Teilnehmern angestrebt. Ausserdem sollten nach Möglichkeit mindestens zwei der drei Teilnehmer eines Experiments eher extrovertiert als introvertiert sein, um Monologsituationen oder solche, in denen überhaupt keine Gespräche stattfinden, möglichst zu umgehen.

##### 6.3.1.1 *Auswahl des Tasks*

Aus Gründen der Vergleichbarkeit wurden wiederum zwei unterschiedliche Tasks ausgewählt. Einer sollte einer Zusammenarbeit, der andere einer Verhandlungssituation entsprechen. Letzterer wurde zum *WG-Spiel*<sup>1</sup>, ersterer wurde zum *Börsenspiel*.

---

<sup>1</sup> WG steht für *Wohngemeinschaft*

Das *WG-Spiel* hatte dabei folgende Ausgangslage: Einer der Teilnehmer ist der Mieter einer 4-Zimmer-Wohnung und sucht zwei Mitbewohner, um darin eine Wohngemeinschaft (WG) zu gründen. Die anderen beiden Teilnehmer sind Interessenten, die gerne ein Zimmer mieten würden. Alle drei Teilnehmer führen die Diskussion über die Gründung der WG per Videokonferenz und müssen sich dabei Fragestellungen annehmen wie das Zuweisen der unterschiedlich grossen Zimmer oder das gerechte Aufteilen der Miete. Oder aber sie diskutieren, wer was zu putzen hat, ob Haustiere und Rauchen in der Wohnung erlaubt sind, ob gemeinsam Lebensmittel eingekauft werden, etc. Als Hilfsmittel stand den Teilnehmern der Grundriss der Wohnung zur Verfügung, wobei ergänzende Hinweise zur Umgebung der Wohnung jeweils nur auf einem der drei Pläne eingezeichnet waren. Ausserdem stand noch eine Auflistung möglicher Fragestellungen als Diskussionsgrundlage zur Verfügung, an die sich die Teilnehmer jedoch bei genügend Eigenkreativität nicht halten mussten.

Beim *Börsenspiel* bestand folgende Situation: Zwei der Teilnehmer bilden ein Anlageteam, das zwei verschiedene Aktien kaufen und verkaufen kann. Als Entscheidungsgrundlagen standen einerseits die sich stetig ändernden Kursverläufe dieser beiden Aktien in Echtzeit zur Verfügung, andererseits die Hinweise eines Börsenmaklers (dritter Teilnehmer), der aufgrund des Kursverlaufs Kauf- bzw. Verkaufsempfehlungen aussprach und Aktientransaktionen ausführte. Das Ziel für das Anlageteam war es, möglichst bei tiefen Kursen Aktien zu beziehen und bei hohen Kursen diese zu veräussern. Als Feedback wurde der Kauf bzw. Verkauf jeweils farblich im entsprechenden Kursverlauf markiert. Als zusätzliche Bedingungen musste sich das Anlageteam zuerst einigen, bevor eine Transaktion durch den Börsenmakler ausgeführt werden konnte.

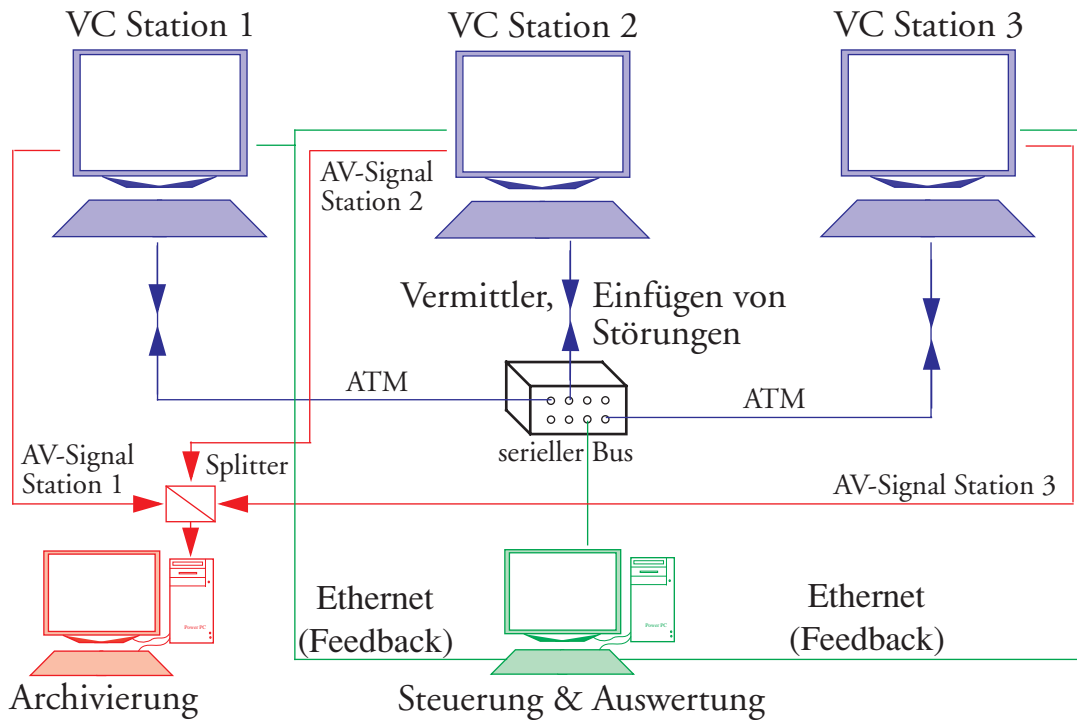
### *6.3.1.2 Fragebogen*

Die Teilnehmer mussten sowohl vor als auch nach den Tests einen Fragebogen ausfüllen. Der erste Bogen diente insbesondere der Erfassung der aktuellen Verfassung der Versuchspersonen sowie einiger statistischer Angaben. Der Fragebogen nach den Versuchen zielte darauf ab, Erfahrungen und Hinweise der Versuchspersonen festzuhalten und später auswerten zu können. Der vollständige Fragebogen ist in [44] abgedruckt.

### *6.3.1.3 Versuchsaufbau*

Für die Versuchsphase mit drei Teilnehmern wurde ein weiterer Versuchsplatz notwendig. Aus diesem Grund wurden drei identische Versuchskabinen einge-

richtet, in denen die Testteilnehmer sowohl visuell als auch akustisch voneinander abgeschirmt waren. Der technische Aufbau erfolgte analog zu den beiden bisherigen (siehe Abbildung 6.20). Dies sollte sicherstellen, dass die Ergebnisse möglichst gut untereinander vergleichbar sein sollten. Zusätzlich wurden wie schon bei Versuch 2 die drei Audio- und Videosignale der Workstations für eine Offline-Auswertung aufgezeichnet und archiviert.



*Abbildung 6.20: Vernetzung der Komponenten*

Abbildung 6.21 zeigt die Anordnung der Versuchskabinen, der Plätze für Versuchsleiter und Beobachter sowie des Konferenztisches. Es wurde darauf geachtet, dass in den Kabinen ein möglichst angenehmes Umfeld herrschte. Vor jedem Experiment wurde fünf Minuten gelüftet, damit wieder genügend Sauerstoff und frische Luft für die Versuche zur Verfügung stand. Zur Schaffung einer ansprechenden Umgebung wurden pro Kabine zwei Pflanzen platziert. Die Lichtverhältnisse waren angenehm und nicht blendend. Um die Videobilder den wechselnden Lichteinflüssen von aussen anzupassen, wurde vor jedem Experiment ein Weissabgleich durchgeführt. Glücklicherweise herrschte aber während der ganzen Versuchsphase schönsten Sommerwetter und mit der Ausrichtung der Fenster gegen Südwesten blieben die Wärmeverhältnisse einigermaßen konstant.



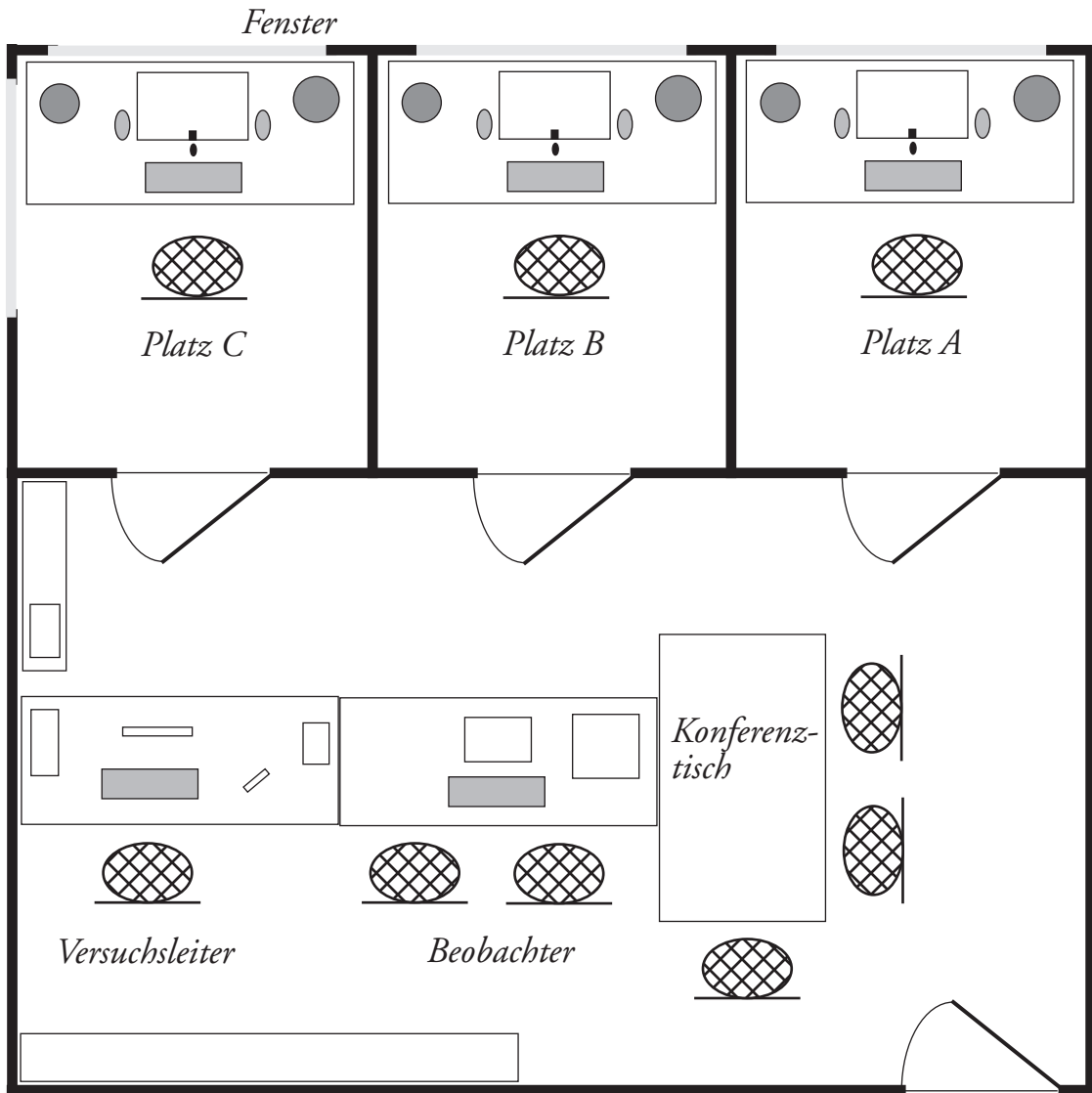


Abbildung 6.21: Räumliche Situation des Versuchsaufbaus

#### 6.3.1.4 Ablauf der Versuche

Die Versuchsprozedur bestand aus folgenden Teilen:

- Einführung (ca. 10 Minuten): Die drei Teilnehmer setzten sich an den Konfereztisch; ihnen wurden Getränke offeriert. Der Versuchsleiter erklärte kurz den Hintergrund der Experimente, anschliessend wurde den Probanden der Ablauf des Experiments erklärt und ihnen die Versuchskabinen gezeigt.
- Task 1 (ca. 30 Minuten): Die Teilnehmer nahmen in den Kabinen Platz, füllten den Fragebogen 1 aus und lasen anschliessend die Anweisungen zu Versuch 1 durch (WG-Spiel). Die Türen wurden geschlossen und die wei-

tere Kommunikation erfolgte via Konferenzsystem. In einer ersten Phase wurden erst Verzögerungen von 0 ms (Minimum) und 2800 ms (Maximum) eingestellt, um den Teilnehmern eine Referenz zu ermöglichen. Danach wurde die echte Versuchsphase mit der Schwellwertbestimmung nach PEST gestartet. Die Teilnehmer diskutierten über die WG-Gründung und mussten im Minutentakt die bereits in Versuch 2 beschriebenen Fragen beantworten (wurde der Delay bemerkt -ja/nein - falls ja, hat er gestört oder nicht). Insgesamt wurden 20 Abfragen durchgeführt, je zehn zur Annäherung der Wahrnehmungsschwelle und zehn zur Akzeptanzschwelle. Im Anschluss an Task 1 füllten die Teilnehmer Fragebogen 2 aus.

- Task 2 (ca. 15 Minuten): Der Ablauf geschah analog zu Task 1. Die Teilnehmer lasen zuerst die Anweisungen durch, danach wurde direkt mit dem Börsenspiel begonnen. Im zweiten Durchgang wurde nur noch nach der Wahrnehmungsschwelle gesucht (10 Durchgänge zu je einer Minute), um das Experiment nicht zu arg in die Länge zu ziehen.
- Fragebogen und Schlussbesprechung (ca. 15 Minuten): Anschliessend verliessen die Teilnehmer die Kabinen und füllten den dritten Fragebogen am Besprechungstisch aus. Zum Abschluss gab es eine offene Diskussion, in denen den Teilnehmern von den Versuchsleitern je nach Ablauf des Experimentes Fragen gestellt wurden bzw. in der die Probanden Anmerkungen zu den Versuchen abgeben konnten.

### 6.3.2 Ergebnisse

Es konnten 26 Versuche in Dreiergruppen durchgeführt werden (jüngster Teilnehmer 20 Jahre, ältester 45 Jahre), Leicht mehr als ein Drittel waren weiblich. Bis auf einen Durchgang in Hochdeutsch wurden alle Konferenzen in Schweizerdeutsch abgehalten.

#### 6.3.2.1 Qualitative Ergebnisse: WG-Gründung

Die Aufgabenstellung einer WG-Gründung stellte sich schnell als sinnvoll heraus. Alle Teilnehmer kamen auf Anhieb mit den Bedingungen zurecht und hatten auch grossen Spass im Verlauf der Experimente, was meist zu lebhaften Sitzungen führte. Vielfach übernahm ein Teilnehmer die Rolle des Sitzungsleiters, indem er Überleitungen zu neuen Themen initiierte oder die Gesprächsinitiative allgemein an sich riss. Die Versuchsdauer von 20 Minuten erwies sich ebenfalls als gut gewählt. Gruppen, die nur harzige Diskussionen zustande brachten, blieben klar die Ausnahme (3 von 26).

Die hinzugefügten Verzögerungen wirkten aus Sicht der Versuchsleitung nicht störend auf die Ergebnisse der Diskussion, beeinflussten den Gesprächsverlauf mit zunehmenden Werten jedoch deutlich, indem sich Teilnehmer wiederholen mussten oder indem Missverständnisse infolge verzögerter Antworten bereinigt werden mussten. Diese konnten meist sofort geklärt werden („Aha, das war jetzt wohl eine sehr grosse Verzögerung“).

Die Bemerkungen der Teilnehmer auf den Fragebogen lassen sich wie folgt zusammenfassen:

- Die überwiegende Mehrheit beurteilte die Videokonferenz als sinnvolles Kommunikationsmittel, um grosse Distanzen zu überwinden und dabei eine Aufgabe wie die gestellte zu lösen.
- Viele bemängelten die fehlende Gestik oder den fehlenden Augenkontakt. Oft wurde auch der Wunsch geäussert, nicht nur ein Portrait der Person, sondern die ganze Körperhaltung per Video zu Gesicht zu bekommen.
- Der Mehrheit der Teilnehmer ist aufgefallen, dass eine grosse Gesprächsdisziplin einzuhalten ist (z.B. das Ansprechen per Namen).
- Des öfteren wurde bemerkt, dass eine Videokonferenz im privaten Rahmen deutliche Vorteile gegenüber einer Telefonkonferenz habe, auch wenn ein Gespräch am runden Tisch immer noch die erste Wahl sei.
- Einige Teilnehmer erwähnten, dass ihrer Ansicht nach der Umgang mit einer Videokonferenz Gewöhnungssache sei, gerade auch im Hinblick auf auftretende Verzögerungen und im Umgang mit diesen.

#### *6.3.2.2 Qualitative Ergebnisse: Börsenspiel*

Auch das Börsenspiel machte den Teilnehmenden generell Spass, Aktien wurden rege ge- und verkauft. Als sinnvoll hat es sich erwiesen, wie geplant nur zehn und nicht wie beim WG-Spiel 20 Durchgänge zu machen, da das Spiel viel schneller monoton und einseitig wurde. Trotzdem liessen die meisten Gruppen auch gegen Ende der zehnminütigen Session noch grossen Enthusiasmus erkennen. Im Vergleich zur WG-Gründung äusserten sich die Teilnehmer in den Fragebögen vermehrt, dass in diesem Fall Verzögerungen fatale Auswirkungen hätten (gerade natürlich auch, wenn echtes Geld im Spiel wäre). Mehrmals wurde auch bemerkt, dass für diesen Task der Videokanal nicht benötigt würde, da man sich rein auf den Kursverlauf konzentriert habe und eine Audiokonferenz für die kurzen Absprachen genügen sollte.

### 6.3.2.3 Quantitative Ergebnisse

Wie schon im zweiten Durchlauf wurden auch hier die Ergebnisse nach zwei verschiedenen Arten ausgewertet: Einmal nach der Methode Best-PEST und einmal durch Mittelwertbildung nach holistischem Ansatz. Zusätzlich wurde noch eine psychometrische Funktion eingesetzt und damit eine dritte Berechnungsmethode für die Schwellwerte hinzugenommen.

#### *WG-Spiel*

Obwohl in Phase 2 festgestellt werden musste, dass die systembedingte Obergrenze des zusätzlich einspeisbaren Verzögerungswertes von 2800 ms nicht ausgereicht hat, wurde für Phase 3 der Wert aus mehreren Gründen nicht erhöht:

- Mit einer Erhöhung auf 5600 ms hätte sich die Granularität erhöht und die Messwerte wären u. U. nicht mehr mit denen aus Phase 2 vergleichbar gewesen.
- Die höhere Teilnehmerzahl liess einen niedrigeren Wert für die Akzeptanzschwelle erwarten (Hypothese H3.5).
- Es wurde erwartet, dass das Testdesign mit den zwei Tasks *WG-Spiel* und *Börsenspiel* zu einer höheren Interaktivität führen sollten, was ebenfalls die Vermutung nahelegte, dass die Akzeptanzwerte niedriger sein sollten.

Während der Versuchsdurchführung haben sich die oben gemachten Annahmen und Erwartungen bestätigt, da sich nur in einzelnen Fällen (2 von 26) die technische Grenze von 2800 ms als zu niedrig erwies, um eine saubere Annäherung zu gewährleisten. In weniger als 8% der Messreihen wären Verzögerungswerte von mehr als 2800 ms nötig gewesen und selbst dann wurde dieser Wert ca. zu 50% als störend betrachtet, was den Schluss zulässt, dass die Akzeptanzschwelle in diesen Fällen nur unwesentlich höher sein könnte. Die Auswertungen wurden deshalb nochmals aufgeteilt: In einem ersten Schritt wurden diejenigen Messreihen, bei denen die Grenze erreicht wurde, weggelassen, in einem zweiten Schritt wurden alle Werte über 2800 ms mit 3000 ms angenommen. Dieses Verfahren ergibt die in Abb. 6.22 dargestellten Werte. Mit der entsprechenden Gewichtung (8%, 92%) sind ebenfalls die beiden Durchschnittswerte angegeben. Diese dürften den tatsächlichen Werten relativ nahe kommen. In Abbildung 6.23 ist die Auswertung mit Hilfe einer psychometrischen Funktion dargestellt. Tabelle 6.9 fasst die verschiedenen Auswertungsergebnisse zusammen.

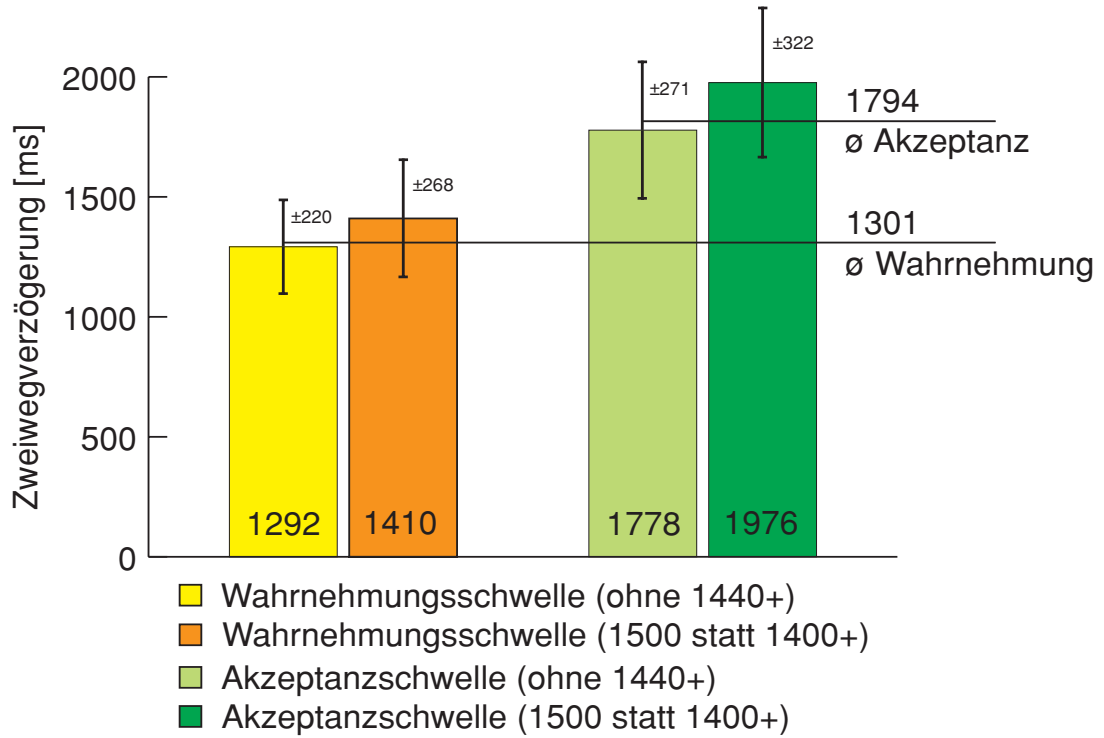


Abbildung 6.22: Ergebnisse des WG-Spiels (holistische Durchschnittswerte)

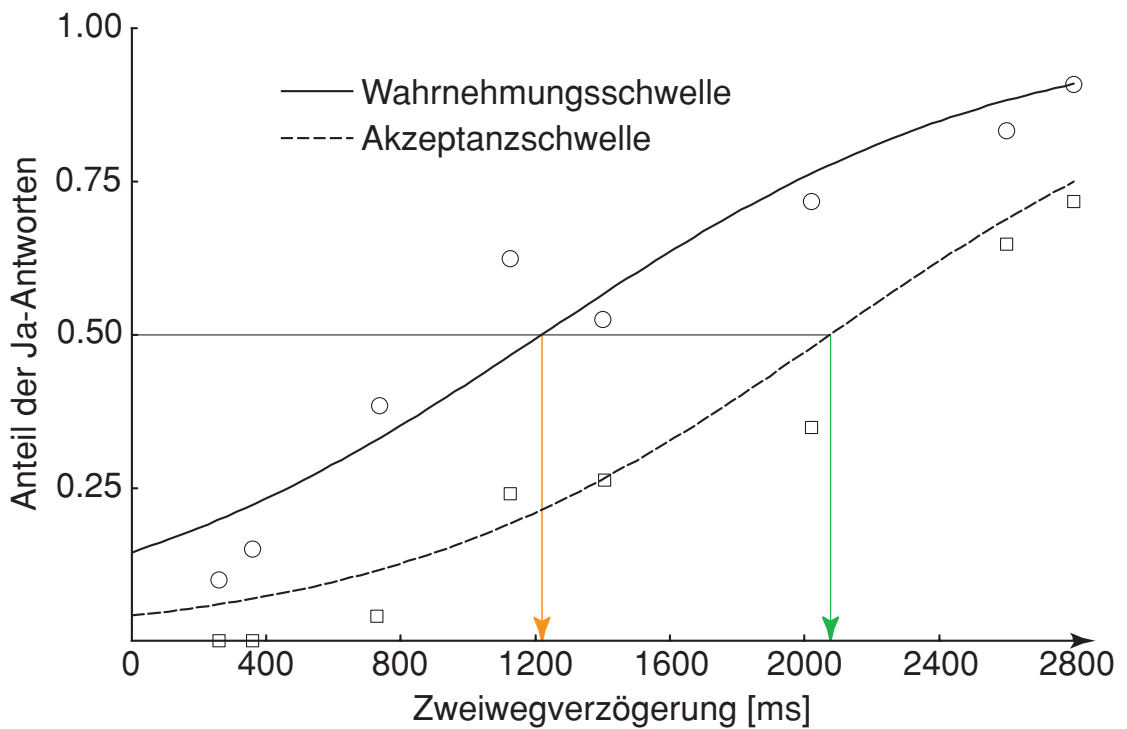


Abbildung 6.23: Ergebnisse des WG-Spiels (Psychometrische Funktion)  
 Bemerkung: Die angegebenen Zeiten in dieser Grafik beinhalten noch nicht die 80 ms Systemverzögerung

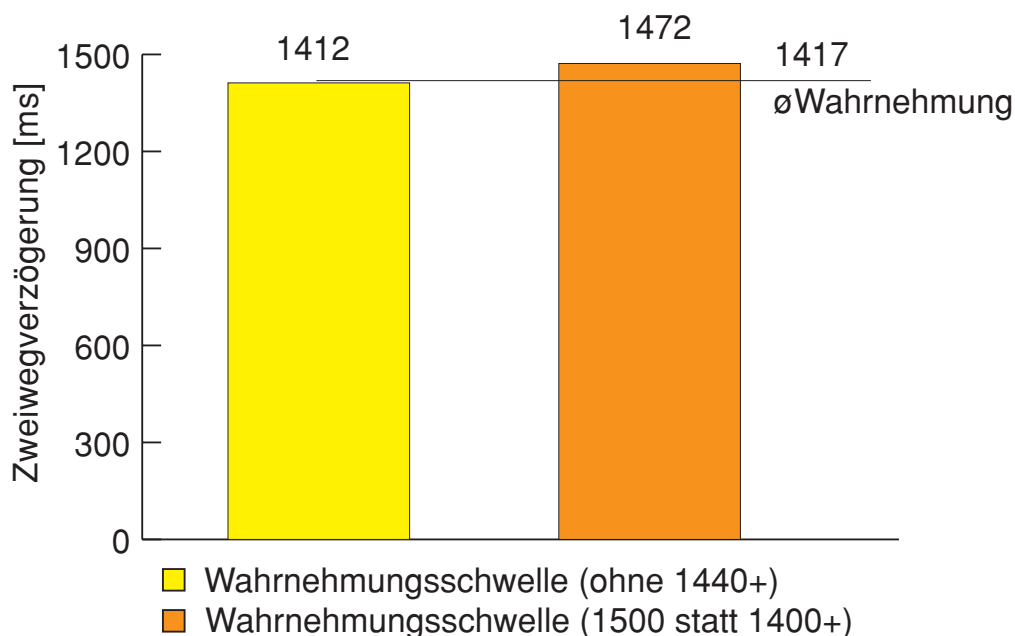
Methode	Wahrnehmung WG (ms)	Akzeptanz WG (ms)
Best-PEST	1326	1804
∅ holistisch	1301	1794
psych. Fkt.	1300	2160 <sup>a</sup>

**Tabelle 6.9:** Übersicht der Ergebnisse des WG-Spiels

- a. Die grosse Abweichung dieses Wertes im Vergleich zu den anderen beiden liegt an der Berechnungsmethode. Da bei der psychometrischen Funktion alle Messwerte berücksichtigt wurden und nicht nur die letzten drei (die bei Best-PEST und dem holistischen Verfahren die Schwellwerte direkt angeben), wird dieser Wert in Richtung 2800 ms verschoben. Dies liegt darin begründet, dass der Wert 2800 ms als Initialwert des Algorithmus einerseits überproportional oft vorkam und andererseits in der Tatsache, dass die Probanden besonders in den ersten drei Durchgängen unkritischer auf grössere Verzögerungswerte reagierten.

### Börsenspiel

Bei der zweiten zu lösenden Aufgabe wurde, wie bereits erwähnt, nur die Wahrnehmungsschwelle gemessen. Die beiden Berechnungsmethoden ergeben die Ergebnisse in Abbildung 6.24. Tabelle 6.10 fasst die verschiedenen Auswertungsergebnisse zusammen.



**Abbildung 6.24:** Ergebnisse des Börsenspiels (holistische Durchschnittswerte)

Method	Wahrnehmung Börsenspiel (ms)
Best-PEST	1414
ø holistisch	1417
psych. Fkt.	n.n.

*Tabelle 6.10: Übersicht der Ergebnisse des WG-Spiels*

### 6.3.3 Diskussion

*Hypothese H3.1: Die Wahrnehmungsschwelle der absoluten Verzögerung liegt tiefer als die Akzeptanzschwelle*

Diese Hypothese, die sich auch in Versuchsreihe 2 bereits bestätigt hat und die eigentlich auch offensichtlich ist, sollte der Vollständigkeit halber überprüft werden. Die Akzeptanzschwelle liegt ca. 40% höher als die Wahrnehmungsschwelle (WG-Spiel). Die Hypothese wird einmal mehr gestützt.

*Hypothese H3.2: Die Wahrnehmungsschwelle der absoluten Verzögerung bei einer Videokonferenz hängt nicht vom spezifischen Task ab*

Geht man beim WG-Spiel von einer Wahrnehmungsschwelle von 1300 bis 1326 ms aus, beim Börsenspiel von 1414 bis 1417 ms (abhängig von der Berechnungsmethode), dann wird auch diese Hypothese von den Versuchsergebnissen gestützt. Die Differenz von maximal 117 ms liegt fast noch innerhalb der Messungenauigkeit der Annäherungsschritte (80 ms) und auf jeden Fall innerhalb der Standardabweichung von jeweils ca. 360 ms.

*Hypothese H3.3: Die Wahrnehmungsschwellen bei Versuchen mit drei Teilnehmern liegen tiefer als mit zwei Teilnehmern*

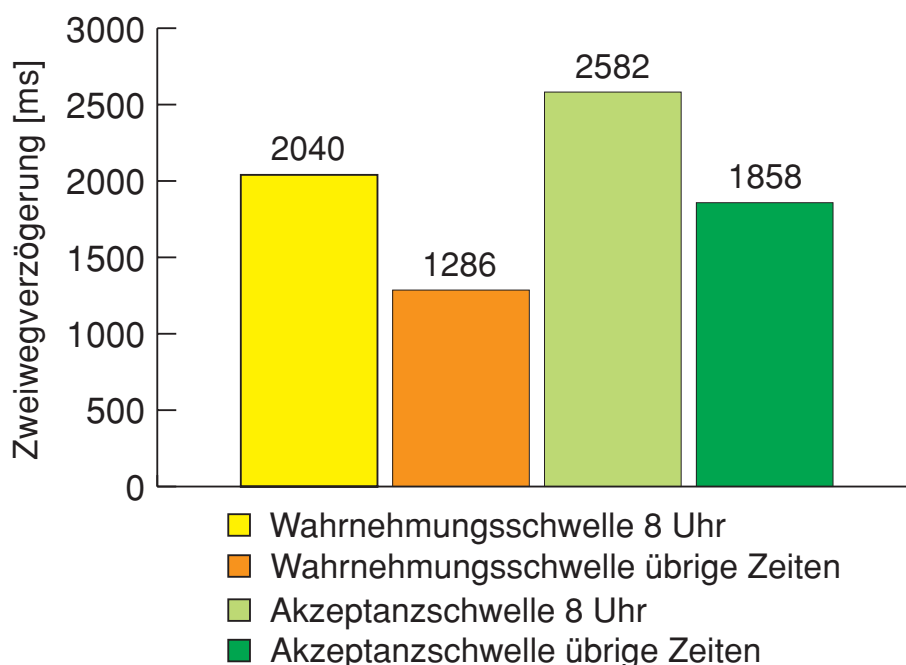
Bei einer durchschnittlichen Wahrnehmungsschwelle von 1358 ms (WG- und Börsenspiel) und einer Schwelle aus den Versuchsdurchgängen 2 mit je nur zwei Teilnehmern von durchschnittlich 1924 ms ist klar ersichtlich, dass Hypothese H3.3 bestätigt werden konnte. Dies liegt vermutlich daran, dass bei drei Teilnehmern in der Regel der jeweils nicht angesprochene wie ein Beobachter von aussen wirkt und somit in dieser Situation gegenüber Verzögerungen wesentlich kritischer reagiert und den anderen Teilnehmern eher ins Wort fällt.

*Hypothese H3.4: Die Akzeptanzschwellen bei Versuchen mit drei Teilnehmern liegen tiefer als mit zwei Teilnehmern*

Die Akzeptanzschwelle im WG-Spiel mit Werten zwischen 1794 bis 2160 (abhängig von der Berechnungsmethode) liegt deutlich unterhalb der Resultate, die in Versuchsreihe 2 gefunden wurden. Zwar konnte dort kein exakter Wert angegeben werden, aber es wurde postuliert, dass der Wert oberhalb von 2800 ms liegen muss. Insofern bestätigt diese grosse Differenz selbst bei einer Annahme von größeren „Messfehlern“ die Hypothese.

*Hypothese H3.5: Die Wahrnehmungs- und Akzeptanzschwellen hängen von der Tageszeit ab und liegen früh morgens höher als am späteren Vormittag oder Nachmittag*

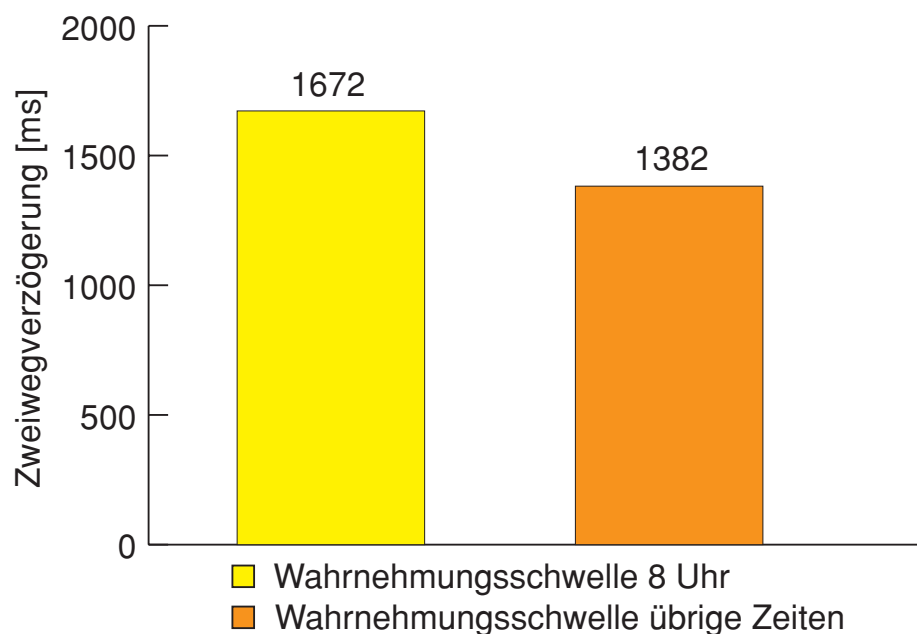
Zu dieser These gibt es bisher keine Untersuchungen und auch in der Literatur keine Aussagen. Die These stützt sich erstens auf Beobachtungen aus den Versuchsphasen 1 und 2 sowie auf die Tatsache, dass die Probanden ausschliesslich Studierende waren. Auch bei unterschiedlichen Biorhythmen lag der Verdacht nahe, dass diese morgens doch eher unaufmerksamer sein würden als am restlichen Tag, da Studierende oft spät ins Bett gehen und morgens entsprechend unausgeschlafen und unaufmerksam sind.



*Abbildung 6.25: Schwellwerte WG-Spiel nach Zeit aufgeteilt*



In Abbildung 6.25 sind die Schwellwerte für Wahrnehmung und Akzeptanz beim WG-Spiel aufgeteilt in Messwerte aus Experimenten um 8 Uhr sowie allen anderen (10, 13, 15 und 17 Uhr). Es ist deutlich zu erkennen, dass die Schwellen morgens signifikant höher sind als am restlichen Tag.



*Abbildung 6.26: Schwellwerte Börsen-Spiel nach Zeit aufgeteilt*

Nicht ganz so deutlich sind die Messwerte für das Börsenspiel. Dies kann daran liegen, dass für diesen Task allgemein eine höhere Aufmerksamkeit notwendig war und die Teilnehmer wie beobachtet auch viel enthusiastischer bei der Sache waren. Auch wenn der Unterschied in diesem Fall statistisch nicht relevant ist, ist der Autor der Meinung, dass zusammen mit den Ergebnissen aus dem WG-Spiel die These H3.5 verifiziert werden konnte.

### *Hypothese H3.6: Die Wahrnehmungsschwellen von Börsenmakler und Anlageteam unterscheiden sich nicht*

Diese Hypothese ist im Prinzip identisch mit der in H3.1. Allerdings wird hier davon ausgegangen, dass die Schwelle nicht nur unabhängig des ausgeführten Tasks ist, sondern dass auch unterschiedliche Bedingungen innerhalb einer Videokonferenz nicht zu unterschiedlichen Werten führen. Die Werte in Abbildung 6.27 lassen eindeutig erkennen, dass die Mittelwerte dieser beiden Gruppen statistisch betrachtet identisch sind. Die Hypothese wird somit bestätigt.

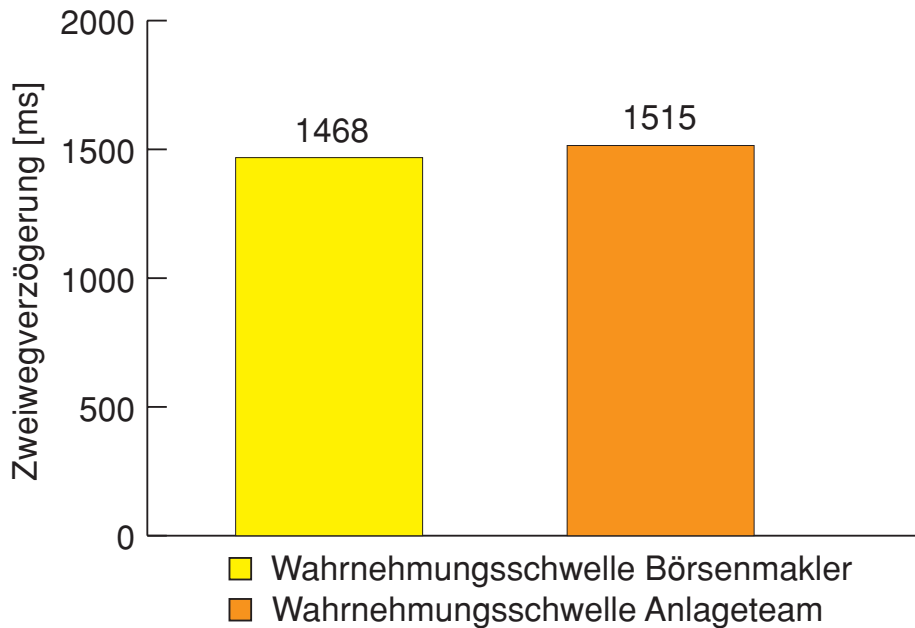


Abbildung 6.27: Wahrnehmungsschwellen von Makler und Anlageteam

## 6.4 Versuchsreihe 4

In der Testphase 4 wurde eine reine Audiokonferenz mit drei Konferenzteilnehmern durchgeführt. Reines Audio, und nicht wie in den vorhergehenden Phasen auch Video, wurde insbesondere aus den folgenden beiden Gründen verwendet:

- Eine reine Audiokonferenz bietet eine hervorragende Vergleichsmöglichkeit mit anderen subjektiven Tests (Tests mit Videokonferenzen in der Art, wie sie innerhalb dieser Arbeit durchgeführt wurden, gibt es bisher nicht).
- Der Autor war daran interessiert, in welcher Weise sich die Wahrnehmungs- und Akzeptanzschwellen bei einem Wechsel von *multimedialer* zu *monomedialer* Kommunikation ändern würde.

### 6.4.1 Methode

Ausgehend von den Erfahrungen und Erkenntnissen der vorangehenden Versuchsreihe wurde auch in dieser vierten Reihe Wert auf eine realistische Versuchsumgebung gelegt. Wiederum wurde für eine schnelle Schwellenannäherung der Best-PEST-Algorithmus gewählt. Für weniger interessant wurden die Fragestellungen gehalten, bei denen das Vorwissen der Testkandidaten bezüglich Computern und Netzwerken eine Rolle spielt.

Da diese Tests insbesondere als Vergleichsbasis zu anderen im Audiobereich durchgeführten Untersuchungen dienen sollten, wurde nur noch eine statistisch genügende, minimale Anzahl an Versuchen geplant. Insgesamt wurden 10 Experimente zu je drei Personen durchgeführt. Angestrebt war wiederum, zwischen weiblichen und männlichen Teilnehmern ein ausgewogenes Verhältnis zu erreichen. Ausserdem sollten nach Möglichkeit mindestens zwei der drei Teilnehmer eines Experiments eher extrovertiert als introvertiert sein, um Monologsituationen oder solche, in denen überhaupt keine Gespräche stattfinden, möglichst zu vermeiden.

#### *6.4.1.1 Auswahl des Tasks*

Aus Gründen der direkten Vergleichbarkeit zu Versuchsreihe 3 wurde diesmal lediglich das *WG-Spiel* als Task ausgesucht. Bei diesem waren bereits die Ergebnisse der Videokonferenz bekannt, wobei sowohl die Wahrnehmungs- als auch die Akzeptanzschwelle gemessen wurde. Wie schon bei der Auswertung der Fragebögen aus Reihe 3 festgestellt wurde, waren einige Teilnehmer der Meinung, dass es für diesen Task keinen Videokanal brauche. Insofern stellte er sich als ideal heraus, um ihn zu Vergleichen heranzuziehen.

#### *6.4.1.2 Fragebogen*

Die Teilnehmer mussten vor und nach der Durchführung des Tasks einen Fragebogen ausfüllen. Der erste Bogen diente insbesondere der Feststellung der aktuellen Verfassung der Versuchspersonen sowie einiger statistischer Angaben. Der Fragebogen nach dem Versuch zielte darauf ab, Erfahrungen und Hinweise der Versuchspersonen festzuhalten und später auswerten zu können, siehe Kapitel 7.

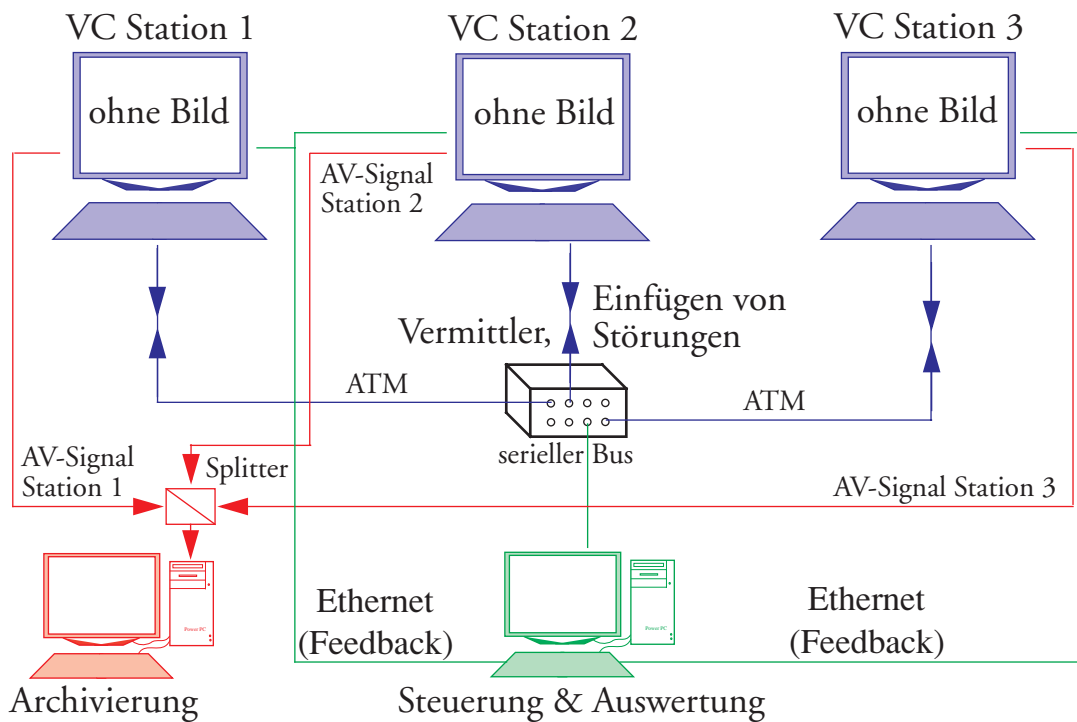
#### *6.4.1.3 Versuchsaufbau*

Der Versuchsaufbau war bis auf folgende Ausnahmen identisch mit demjenigen von Testreihe 3:

- Die Teilnehmer hatten keine Videobilder zur Verfügung.
- Die Kamera wurde zu Überwachungszwecken unterhalb des Monitors platziert (es wurde in den vorigen Tests festgestellt, dass die Teilnehmer meistens auf die Unterlagen auf dem Tisch blicken und deshalb wurden von der neuen Position bessere Kontrollbilder erwartet).

Die Börsenapplikation wurde nicht mehr gestartet, da nur noch das *WG-Spiel* durch die Teilnehmer zu absolvieren war. Somit wurde der Monitor lediglich

dazu benutzt, Feedback über die Qualität des jeweiligen Versuchsabschnitts zu geben.



*Abbildung 6.28: Vernetzung der Komponenten in Versuchsphase 4 (Audiokonferenz)*

Abbildung 6.28 stellt die technische Versuchsanordnung dar. Es ist zu beachten, dass zwar weiterhin Audio- und Videosignal für spätere Auswertungszwecke aufgezeichnet wurde, die Teilnehmer jedoch kein Bild ihrer Gesprächspartner mehr sehen konnten. Die Monitore dienten nur der Rückmeldungen über die wahrgenommene Qualität.

Für die Darstellung der räumlichen Gegebenheiten wird auf Abbildung 6.21 weiter oben verwiesen, da sich im Unterschied zu Versuchsreihe 3 nichts geändert hat.

#### *6.4.1.4 Ablauf der Versuche*

Die Versuchsprozedur bestand aus folgenden Teilen:

- Einführung (ca. 10 Minuten): Die drei Teilnehmer nahmen am Besprechungstisch Platz und der Versuchsleiter erklärte kurz den Hintergrund der Experimente. Anschliessend wurde den Probanden der Ablauf des Experiments erklärt und ihnen die Versuchskabinen gezeigt.
- Task (ca. 30 Minuten): Die Teilnehmer nahmen in den Kabinen Platz, füllten den Fragebogen 1 aus und lasen anschliessend die Anweisungen zum WG-Spiel durch. Die Türen wurden geschlossen und die weitere Kommunikation erfolgte via Konferenzsystem. In einer ersten Phase wurden erst Verzögerungen von 0 ms (Minimum) und 2800 ms (Maximum) eingestellt, um den Teilnehmern eine Referenz zu ermöglichen. Danach wurde die echte Versuchsphase mit der Schwellwertbestimmung nach PEST gestartet. Die Teilnehmer diskutierten über die WG-Gründung und mussten im Minutentakt die bereits in Versuch 2 beschriebenen Fragen beantworten (wurde der Delay bemerkt -ja/nein - falls ja, hat er gestört oder nicht). Insgesamt wurden 20 Abfragen durchgeführt, je zehn zur Annäherung der Wahrnehmungsschwelle und zehn zur Akzeptanzschwelle.
- Fragebogen und Schlussbesprechung (ca. 15 Minuten): Anschliessend verliessen die Teilnehmer die Kabinen und füllten den zweiten Fragebogen am Konferenztisch aus. Abschliessend gab es eine offene Diskussion, in denen wir den Teilnehmern je nach Ablauf des Experimentes Fragen stellten bzw. in der sie Anmerkungen zu den Versuchen abgeben konnten.

#### *6.4.2 Ergebnisse*

Insgesamt wurden wie geplant zehn Versuche in Dreiergruppen durchgeführt (30 Teilnehmer, jüngster Teilnehmer 19 Jahre, ältester 32 Jahre, Ein Viertel davon waren weiblich). Alle Konferenzen wurden in Schweizerdeutsch abgehalten.

##### *6.4.2.1 Qualitative Ergebnisse*

Wie zu erwarten, gab es mit der Aufgabe, die sich bereits in der vorhergehenden Reihe per Videokonferenz bewährt hatte, keine Probleme. Die Teilnehmer kamen gut mit der Aufgabenstellung und dem System zurecht. Einziger Unterschied war, dass sie sich zu Beginn ein Bild machen mussten, wer in welcher Kabine sass und welche Stimme zu welcher Person gehörte. Im Unterschied zu einer 'echten' spontanen Audiokonferenz hatten die Teilnehmer den Vorteil,

sich vor Beginn am Besprechungstisch schon kurz kennengelernt zu haben, falls sie sich nicht schon bereits kannten, was in ca. der Hälfte der Experimente der Fall war.

Die Bemerkungen der Teilnehmer auf den Fragebogen lassen sich wie folgt zusammenfassen:

- Praktisch alle Teilnehmer bemängelten die bei der Audiokonferenz fehlende Möglichkeit, Mimik und Gestik der Gesprächspartner nicht zu kennen. Besonders für eine Aufgabe wie diese sei es wichtig, seine Partner auch optisch einschätzen zu können.
- Fast alle Probanden würden bei kleinen geografischen Distanzen ein Treffen am runden Tisch bevorzugen. Nur wenn dies unter keinen Umständen möglich wäre, wäre ihrer Meinung nach eine Audiokonferenz in Betracht zu ziehen.
- Der Mehrheit der Teilnehmer ist aufgefallen, dass eine grosse Gesprächsdisziplin einzuhalten ist (z.B. das Ansprechen per Namen, Gesprächspausen abwarten, Koordination der Sprecher untereinander regeln).
- Bemerkte wurde auch, dass besonders die sich ändernden Verzögerungswerte grosse Aufmerksamkeit erforderten. Kaum habe man sich darauf eingestellt, praktisch normal miteinander kommunizieren zu können, würde eine grosse Verzögerung die eigene Strategie zunichte machen.

#### *6.4.2.2 Quantitative Ergebnisse*

Die Messergebnisse wurden auf drei verschiedenen Arten ausgewertet: Nach der Methode Best-PEST, durch Mittelwertbildung nach holistischem Ansatz (Definition siehe Versuchsreihe 2) sowie mit einer psychometrischen Funktion.

Wie erwartet (Hypothese H4.1) lag der Schwellwert für die Wahrnehmungs bei der Audiokonferenz tiefer als bei der Videokonferenz. In Abbildung 6.29 ist die Auswertung mit Hilfe der ermittelten Durchschnittswerte dargestellt. Tabelle 6.11 fasst die verschiedenen Auswertungsergebnisse auf einen Blick zusammen.

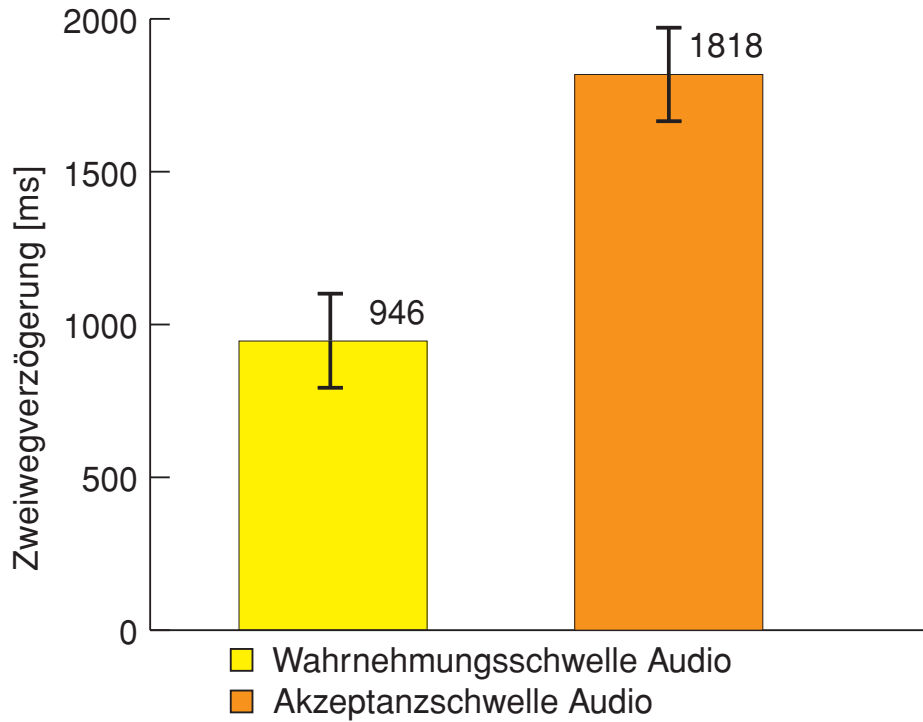


Abbildung 6.29: Ergebnisse (holistische Durchschnittswerte)

Abbildung 6.30 zeigt die Ergebnisse mit Annäherung durch eine psychometrische Funktion.

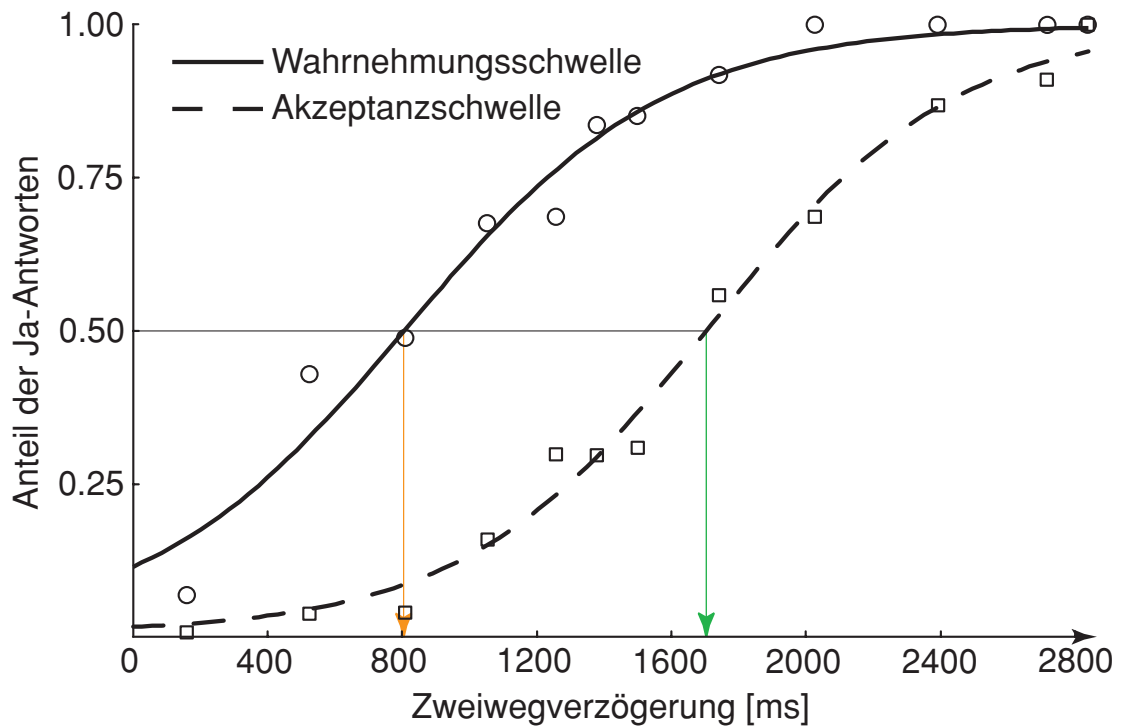


Abbildung 6.30: Ergebnisse des WG-Spiels (Psychometrische Funktion)

Bemerkung: Die angegebenen Zeiten in dieser Grafik beinhalten noch nicht die 80 ms Systemverzögerung

Methoden	Wahrnehmung WG (ms)	Akzeptanz WG (ms)
Best-PEST	960	1838
ø holistisch	946	1818
psych. Fkt.	882	1766

*Tabelle 6.11: Übersicht der Ergebnisse des WG-Spiels*

### 6.4.3 Diskussion

*Hypothese H4.1: Die sich ergebende Wahrnehmungsschwelle der Benutzer der Audiokonferenz wird tiefer sein als bei einer vergleichbaren Videokonferenz*

Im Vergleich zum gemessenen Schwellwert der Videokonferenzen aus Versuchsreihe 3 lag die Wahrnehmungsschwelle um ca. 25% tiefer. Die Hypothese wurde durch die Versuche somit bestätigt wie auch die Vermutung, dass ein zusätzliches Videobild beim WG-Spiel die Teilnehmer entweder mehr ablenkt oder aber mehr fordert, was beides zu einer höheren Wahrnehmungsschwelle führen kann.

*Hypothese H4.2: Die sich ergebende Akzeptanzschwelle der Benutzer der Audiokonferenz wird tiefer sein als bei einer vergleichbaren Videokonferenz*

Die Akzeptanzschwelle bei der Audiokonferenz lag genau im selben Bereich wie die vergleichbare Messgrösse aus Versuchsreihe 3. Somit muss diese Hypothese verworfen werden und es zeigt sich, dass das Gegenteil zutrifft. Es wäre in weiteren Versuchen zu untersuchen, ob die Akzeptanzschwelle zwar vom auszuführenden Task abhängt, nicht jedoch davon, ob es sich dabei um eine Audio- oder um eine Videokonferenz handelt.

## 6.5 Schlussfolgerungen

Es muss festgestellt werden, dass die Durchführung der Versuchsreihen trotz kleinerer Rückschläge äusserst erfolgreich verlaufen ist. Als wichtigstes Ergebnis wird die Feststellung betrachtet, dass die Akzeptanzschwelle betreffend absoluter Verzögerung bei einer Videokonferenz (und auch bei einer reinen Audiokonferenz) zu einem grossen Teil vom ausgeführten Task abhängt. Wei-



tere Einflüsse liefern das Szenario (beispielsweise die Anzahl Teilnehmer, die Dauer der Nutzung, die Tageszeit etc.)

Die zweitwichtigste Aussage betrifft die Wahrnehmungsschwelle. Die Versuche haben ergeben, dass diese wohl abhängig vom Szenario ist, nicht jedoch vom ausgeführten Task und auch nicht von der Anzahl der verwendeten Modi (Video und Audio oder nur Audio).

Leider haben sich, bedingt durch die Messmethode, einige Unsicherheiten eingeschlichen. Besser wäre ein 2-AFC-Verfahren gewesen, was aber zeitlich in den beschriebenen Fällen nicht möglich war. In weiteren Versuchsreihen könnte dieses Verfahren jedoch erfolgreich eingesetzt werden. Denn erstens sind jetzt die Bereiche genau definiert, in denen die Schwellen erwartungsgemäss zu liegen kommen, für eine Annäherung wären also weniger Testintervalle notwendig. Zum anderen müsste pro Experiment nur ein Task (und nicht wie bisher zwei) untersucht werden, da die Taskunabhängigkeit bzw. Taskabhängigkeit der beiden Schwellen in den durchgeführten Versuchen bewiesen wurde. Somit wäre es also möglich, ohne die Probanden länger als 60 Minuten zu beanspruchen, die beiden Schwellen bei einem entsprechenden Szenario genauer einzugrenzen. Sogar eine Verlängerung des Messintervalls von einer auf zwei Minuten wäre möglich, was einen weiteren Genauigkeitszuwachs bringen würde. Zu erwarten wären Ergebnisse mit kleinerer Varianz.

Weitere Folgerungen sind untenstehend aufgeführt:

- Sowohl die Akzeptanz- als auch die Wahrnehmungsschwelle waren in allen Versuchen deutlich höher als die Werte, die in der Literatur gefunden wurden. Dies liegt nach Meinung des Verfassers daran, dass multimodale Kommunikation mehr Aufmerksamkeit von den Benutzern verlangt und zu disziplinierteren Gesprächen führt. Dies bedeutet, dass unter Einbezug mehrerer Modi (z.B. drei statt zwei) unter sonst gleichbleibenden Bedingungen die Schwellwerte noch weiter ansteigen.
- Der Ton ist bei einer Videokonferenz das wichtigste Standbein. Je nach Task ist der Videokanal ein nützlicher (WG-Spiel) oder sogar notwendiger (Lego-Spiel) Zusatz, in anderen Fällen ist er aber nutzlos oder gar störend (Börsenspiel, Reisespiel). Dieses Ergebnis wird von anderen Studien bestätigt [38].
- Wenn bei einer Videokonferenz das Bild eine zu schlechte Qualität aufweist, wird es automatisch nicht mehr beachtet und man konzentriert sich alleine auf den Ton. Auch dieses Resultat wird in anderen Untersuchungen bestätigt [83].

- Es gab in allen Fällen einen Gewöhnungseffekt sowohl an das System als auch an die Kommunikationsbedingungen. Es hat sich herausgestellt, dass während einer ca. fünfminütigen Eingewöhnungsphase höhere Verzögerungen nicht wahrgenommen werden, die von den gleichen Personen unter gleichen Umständen im weiteren Verlauf plötzlich als störend beurteilt werden (Lernkurve). Dies bedeutet, dass die Benutzeransprüche auch davon abhängen, wie lange ein System zur Kommunikation benutzt wird. Sollte sich beispielsweise herausstellen, dass die durchschnittliche Dauer eines Videokonferenzgesprächs im Rahmen eines durchschnittlichen Telefongesprächs liegt (3 Minuten), dann werden die Benutzer eine schlechtere Qualität akzeptieren als wenn es sich um eine 30minütige Diskussion handelt.
- Die in Versuchsreihe 1 festgestellten Maskierungseffekte zeigen ansatzweise, wie kompliziert Modelle aussehen, wenn mehr als nur eine unabhängige Variable einbezogen wird. Es ist zu vermuten, dass viele Tests für ein aussagekräftiges Modell notwendig sind und dieses dann trotzdem noch sehr viele Vereinfachungen beinhalten wird.

# *Kapitel 7*

## *Messungen der Sprechaktivitäten*

---

Dieses Kapitel beschreibt die Ergebnisse, die durch die nachträgliche Auswertung der bei den Versuchen aufgezeichneten Video- und Audiodaten gewonnen wurden. Wie bereits im vorhergehenden Kapitel erwähnt, wurden die Audio- und Videosignale aus den Versuchsphasen 3 und 4 aufgezeichnet. Die Erwartungen an die Offline-Auswertungen werden nachfolgend beschrieben.

Entsprechend den Ausführungen von Abschnitt 4.3 wurde erwartet, dass sich (je nach betrachtetem Abschnitt) die Sprach- und Aktivitätsmuster der Diskussionsteilnehmer ändern, so dass aus diesen Mustern „objektive“ Parameter extrahiert werden können, aus denen Rückschlüsse auf die Zufriedenheit der Benutzer gezogen werden können. Das Ziel einer solchen Analyse ist es natürlich, den Vorgang letztlich umkehren zu können und daraus eine Mess- und Bewertungsmethode für zukünftige Multimedia-Kommunikations-Systeme abzuleiten. Gelingt dies, können bei einer Multimediakommunikation während einer bestimmten Zeit Messungen der Aktivitätsmuster durchgeführt und die charakteristischen Parameter extrahiert werden. Anschliessend können Aussagen über die Zufriedenheit der Benutzer mit diesem System gemacht werden.

Beispielsweise ist es denkbar, dass die Häufigkeit und die Zeitdauer zunimmt, während der zwei Personen gleichzeitig sprechen (sich ins Wort fallen), wenn die absolute Verzögerung bei einer Videokonferenz zunimmt. Oder es ist möglich, dass die Pausen zwischen einzelnen Sprechphasen (turns) bei zunehmender Verzögerung länger werden.

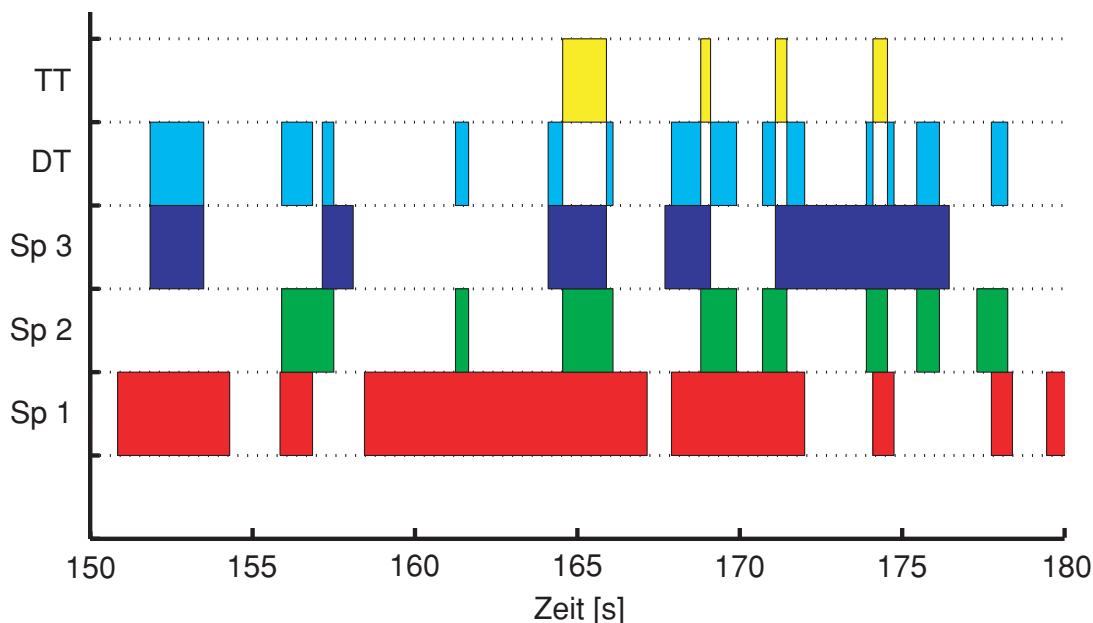
Nachforschungen und Literaturrecherchen haben ergeben, dass bisher keine derartigen Zusammenhänge untersucht worden sind. Für erste Untersuchungen im Zusammenhang mit den Versuchen aus dem vorigen Kapitel wurden folgende Hypothesen aufgestellt:

*H5.1:* Die Zeitdauer während einer Videokonferenz, in denen zwei Personen gleichzeitig sprechen (*Doubletalk*) nimmt zu, wenn die absolute Verzögerung grösser wird.

*H5.2:* Die Zeitdauer während einer Videokonferenz, in denen drei Personen gleichzeitig sprechen (*Tripletalk*) nimmt zu, wenn die absolute Verzögerung grösser wird.

## 7.1 Methode

Für die Analyse wurden die zehn durchgeführten Experimente aus Testphase 4 verwendet. In einem ersten Schritt wurden mit einem Algorithmus<sup>1</sup> für die drei Tonkanäle (im folgenden mit Sprecher 1 bis 3 bezeichnet) die Zeiten bestimmt, in denen der jeweilige Sprecher aktiv war (Sprechzeiten).



Sp1: Sprecher 1, Sp2: Sprecher 2, Sp3: Sprecher 3

DT: Doubletalk, TT: Tripletalk

(im gezeigten Zeitfenster beträgt die Zweiwegverzögerung 2880 ms)

*Abbildung 7.31: Darstellung der Aktivitätsmuster*

<sup>1</sup> Die Detektion der Sprechphasen wurde anhand des Kurzzeitmittelwertes der Signalleistung durchgeführt. Die lokale Unterscheidung zwischen Sprache und Pause geschieht anhand eines einstellbaren Schwellwertes. Sprechphasen müssen eine minimale Dauer von 'minspe' (hier 0.25 s) haben und dürfen Pausen enthalten, die maximal 'maxsil' (hier 0.6 s) lang sind. (*standard speech detector with fixed intensity threshold*)

Anschliessend wurden die drei erhaltenen Aktivitätsmuster zeitlich synchronisiert übereinandergelegt, um daraus Phasen mit Doubletalk und Tripletalk zu berechnen (siehe Abb. 7.31). In einer sechsten Spur wurde die jeweils verwendete Verzögerung zwischen den Kommunikationspartnern eingetragen.

In einem zweiten Schritt wurden drei Werte berechnet: Gesamtsprechzeit, Dauer mit Doubletalk und Dauer mit Tripletalk, jeweils in Abhängigkeit der momentanen Zweiwegverzögerung. Da während der Experimente die Verzögerungsintervalle nicht alle gleich häufig vorkamen, wurden die errechneten Werte in einem dritten Schritt normiert, so dass sie miteinander vergleichbar wurden.

## 7.2 Ergebnisse

Abbildung 7.32 zeigt die gemessenen Werte aus Versuchsphase 4. Um statistisch aussagekräftige Werte zu erhalten, wurden alle zehn Experimente ausgewertet und die Anteile addiert.

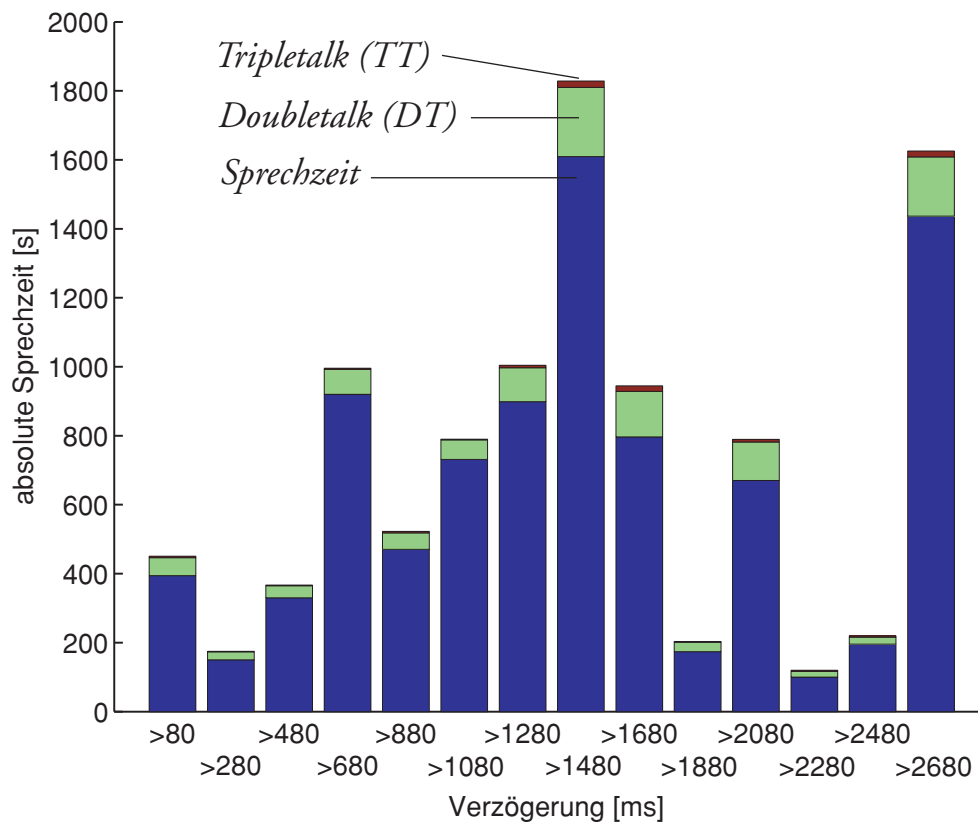


Abbildung 7.32: Darstellung der absoluten Sprechzeit mit Double- und Tripletalk<sup>1</sup>

Es ist gut zu erkennen, dass die Intervalle 4 und 7 bei den Experimenten besonders häufig vertreten waren. Dies liegt in der Tatsache begründet, dass darin die beiden Initialwerte für den Best-PEST Algorithmus enthalten sind, die bei jedem Experiment mindestens einmal vorkamen. Die Initialwerte lagen bei 1480 ms (Wahrnehmungsschwelle) sowie 2880 ms (Akzeptanzschwelle).

Die normierte Darstellung ist in Abbildung 7.33 gezeigt. Eine Übersicht über die prozentuale Zusammensetzung gibt Tabelle 7.12.

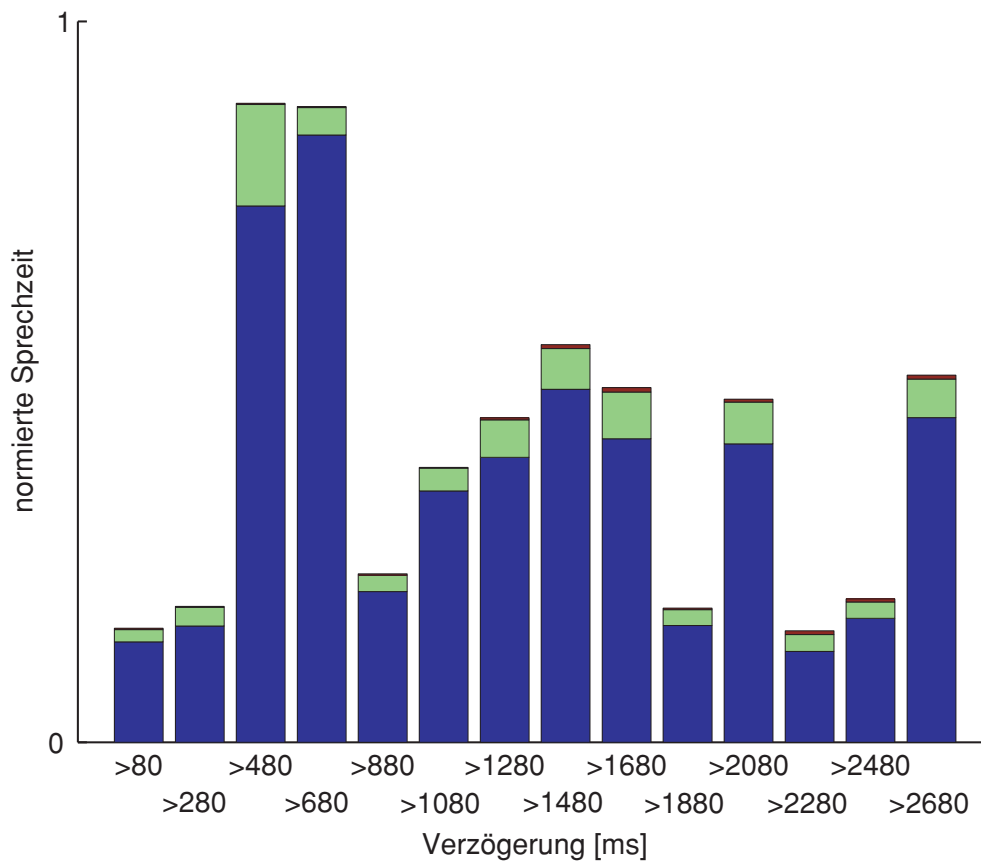


Abbildung 7.33: Darstellung der normierten Sprechzeit mit Double- und Tripletalk

Verzög. [ms]	80-280	281-480	481-680	681-880	881-1080	1081-1280	1281-1480
DT [%]	12.1	16.0	18.9	4.5	10.8	9.1	13.2
TT [%]	1.5	0.7	0.2	0.2	1.0	0.3	0.7

<sup>1</sup> Die Beschriftung der Abszisse ist wie folgt zu lesen: Das Intervall >80 beinhaltet alle Experimentphasen, in denen Zweiwegverzögerungen zwischen 80 und 480 ms verwendet wurden. Analog dazu beinhaltet das Intervall >480 alle zwischen 481 und 880 ms usw.

Verzög. [ms]	1480- 1680	1681- 1880	1881- 2080	2081- 2280	2281- 2480	2481- 2680	2681- 2880
DT [%]	11.6	15.4	13.6	14.0	18.5	13.2	11.9
TT [%]	1.1	1.4	1.4	1.0	4.1	2.5	1.2

*Tabelle 7.12: Prozentuale Anteile von Double- und Tripletalk bzgl. der Gesamtsprechzeit*

### 7.3 Diskussion

Die Werte des ersten Intervalls (80 - 480 ms) sind mit Vorsicht zu geniessen, da in den 10 Experimenten nur bei einem diese Werte überhaupt erreicht wurden. Insofern sind diese Messwerte statistisch gesehen nicht relevant und werden für die folgende Diskussion verworfen.

In Abbildung 7.33 (normierte Sprechzeit) ist sehr gut zu erkennen, dass bei Werten zwischen ca. 400 ms und 900 ms im Vergleich zu anderen Werten wesentlich mehr Sprechzeit stattgefunden hat. Dies bedeutet, dass die Probanden bevorzugt in dem Bereich mit den genannten Verzögerungszeiten miteinander kommuniziert haben. Gut erkennbar ist somit die Wahrnehmungsschwelle (Best-Pest bei 960 ms, psychometrische Funktion bei 882 ms), ab welcher die Sprechzeit deutlich gesunken ist.

Weniger klar erkennbar ist die gemessene Akzeptanzschwelle (Best-Pest bei 1838 ms, psychometrische Funktion bei 1766 ms). Bei einer Darstellung wie in Abbildung 7.34 wird diese Schwelle besser sichtbar. Hier sind klar die drei interessanten Bereiche erkennbar: Bereich 1 liegt noch unterhalb der Wahrnehmungsschwelle, die Probanden diskutieren ohne Beeinträchtigung. Bereich 2 liegt oberhalb der Wahrnehmungs- aber unterhalb der Akzeptanzschwelle. Bereich 3 liegt oberhalb der Akzeptanzschwelle. Allerdings muss gesagt werden, dass diese Erkenntnisse mit weiteren subjektiven Versuchen genauer verifiziert werden müssen.

Die Tatsache, dass die zweite Schwelle weniger ausgeprägt ist als die erste, kann darauf hinweisen, dass diese den Charakter einer „weicheren“ Schwelle hat. Die Benutzer sind der Meinung, dass bis zu einem gewissen Wert die Verzögerung tolerierbar ist. Tatsächlich werden die Benutzer von ihr aber schon ab der Wahrnehmungsgrenze beeinflusst, dies jedoch nicht bewusst sondern im Unterbewusstsein. Dies würde die Ergebnisse von Wilson/Sasse untermauern,

die vergleichbare Ergebnisse erhalten haben, als sie den Einfluss der Bildwiederholrate auf die Benutzerzufriedenheit gemessen haben [91]. Es scheint so zu sein, dass die Probanden subjektiv eine andere Meinung äussern als der Körper reaktiv feststellt. Hinsichtlich dieses Phänomens müssten Untersuchungen durchgeführt werden, die das Verhalten der Benutzer bei kürzerem Gebrauch bzw. bei längerem Gebrauch einer solchen Videokonferenz erforschen. Die ersten Ergebnisse deuten darauf hin, dass die Benutzer bei längerem Gebrauch kritischer werden und sich dann der Einfluss der Wahrnehmungsschwelle gegenüber der Akzeptanzschwelle weiter verstärken wird.

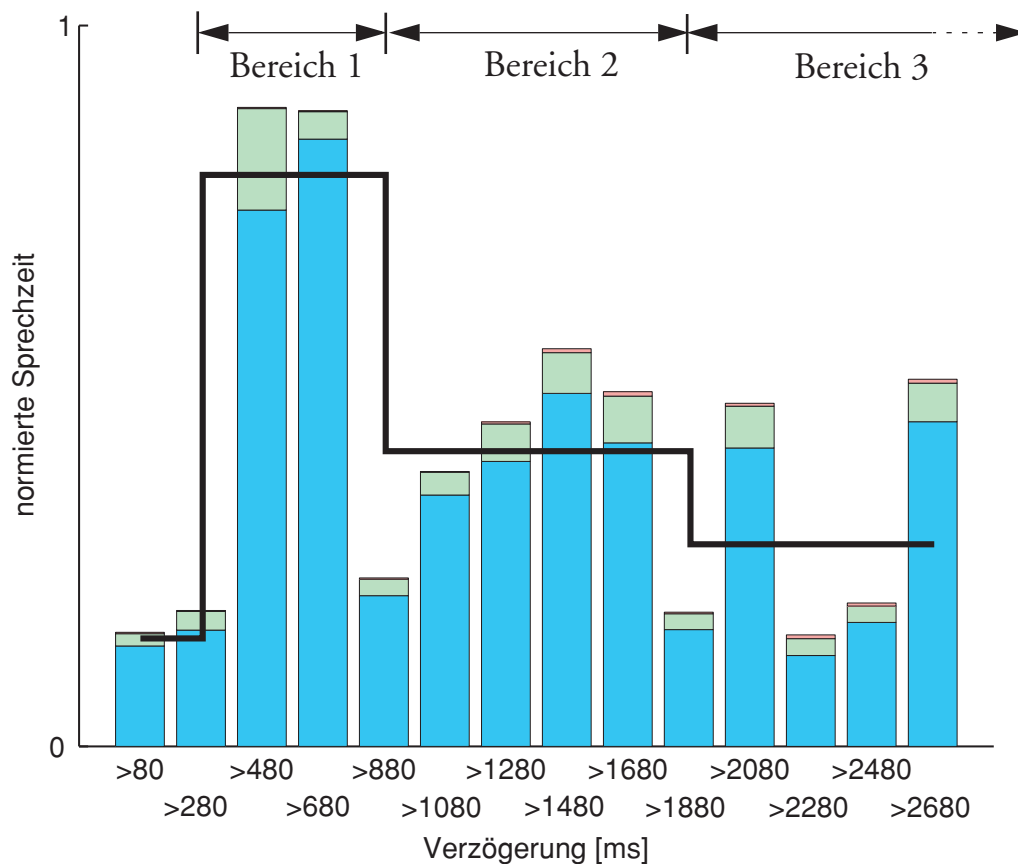


Abbildung 7.34: Bereiche innerhalb der normierten Sprechzeit

Möglicherweise zeigt sich dann, dass sogar noch weitere Abstufungen unterscheidbar sind:

Stufe 3: Es wird bewusst eine Beeinträchtigung empfunden (wahrgenommen) und die Ursache (der Störfaktor) ist bekannt.



Stufe 2: Es wird bewusst eine Beeinträchtigung empfunden (wahrgenommen) und die Ursache (der Störfaktor) ist nicht bekannt („es stört mich etwas, aber ich kann nicht sagen, was es ist...“).

Stufe 1: Es wird bewusst keine Beeinträchtigung empfunden, trotzdem reagiert der Körper mit Stress auf die unterbewusst registrierte Qualitätsverschlechterung.

Werden die prozentualen Anteile von Double- bzw. Tripletalk betrachtet, fällt auf, dass auch dort zwischen drei Bereichen unterschieden werden kann. Der Anteil an Doubletalk bewegt sich leicht ansteigend bis zu Verzögerungswerten von 700 ms im Bereich von 15.7 %, danach fällt er bis zu Verzögerungswerten von 1700 ms auf 9.8 % ab und steigt anschliessend wieder auf Werte um 15 %. Die bisherigen Erkenntnisse lassen den Schluss zu, dass Doubletalk im ersten Bereich besonders aus Bestätigungslauten sowie aus überlappenden Aussagen besteht, wie sie vorkommen, wenn eine Unterhaltung flüssig geführt wird (A weiss schon einige ms vor dem Satzende von B, was er sagen will und beginnt bereits mit seinen Ausführungen). Im Bereich zwei sind die eingefügten Verzögerungswerte gross genug, um die beiden eben angeführten Doubletalkquellen zu eliminieren, es ist aber nach wie vor eine praktisch störungsfreie Unterhaltung möglich. In Bereich 3 hingegen steigt Doubletalk wieder an, diesmal aber weil man sich gegenseitig ins Wort fällt, nachhaken und nachfragen muss sowie mehr Aufwand zur Koordination des Gesprächs treiben muss.

Bei Tripletalk ist eine ähnliche Interpretation möglich, allerdings müssten hier statistisch aussagekräftigere Daten verwendet werden, für die weitere subjektive Versuche mit mehr als 30 Probanden notwendig wären. Es ist zu erwarten, dass eine grössere Datenbasis einerseits zu genaueren Schwellwerten und andererseits zu korrekten Werten von Double- und Tripletalk führen wird.



# *Kapitel 8*

## *Schlussbetrachtungen und Ausblick*

---

Im Rahmen der vorliegenden Arbeit konnten drei wichtige Erkenntnisse gewonnen werden:

1. Die Benutzer von Videokonferenzen akzeptieren zum Teil erstaunlich grosse Werte bei der Zweiwegverzögerung im Vergleich zu bisher durchgeführten Studien.
2. Je multimodaler eine Kommunikation wird, desto disziplinierter wird die Unterhaltung und gleichzeitig unkritischer die Zweiwegverzögerung.
3. Es ist möglich, anhand der normierten Sprechzeit und den Anteilen von Double- und Tripletalk Rückschlüsse über die Benutzerzufriedenheit bei Videokonferenzen zu ziehen.

Die Arbeit konnte den Zusammenhang zwischen der absoluten Verzögerung bei Multimediakommunikation und der Benutzerzufriedenheit aufzeigen, insbesondere wurde der Weg zu einem neuen Forschungsgebiet bereitet und aufgezeigt, in welche Richtung zukünftige Untersuchungen gehen müssen sowie in welchem Bereich Ergebnisse zu erwarten sind.

### *8.1 Modell der Benutzerzufriedenheit*

Basierend auf den Erkenntnissen der subjektiven Tests hat sich gezeigt, dass die Akzeptanzschwelle der Benutzer einer Videokonferenz von folgenden Einflüssen abhängt:

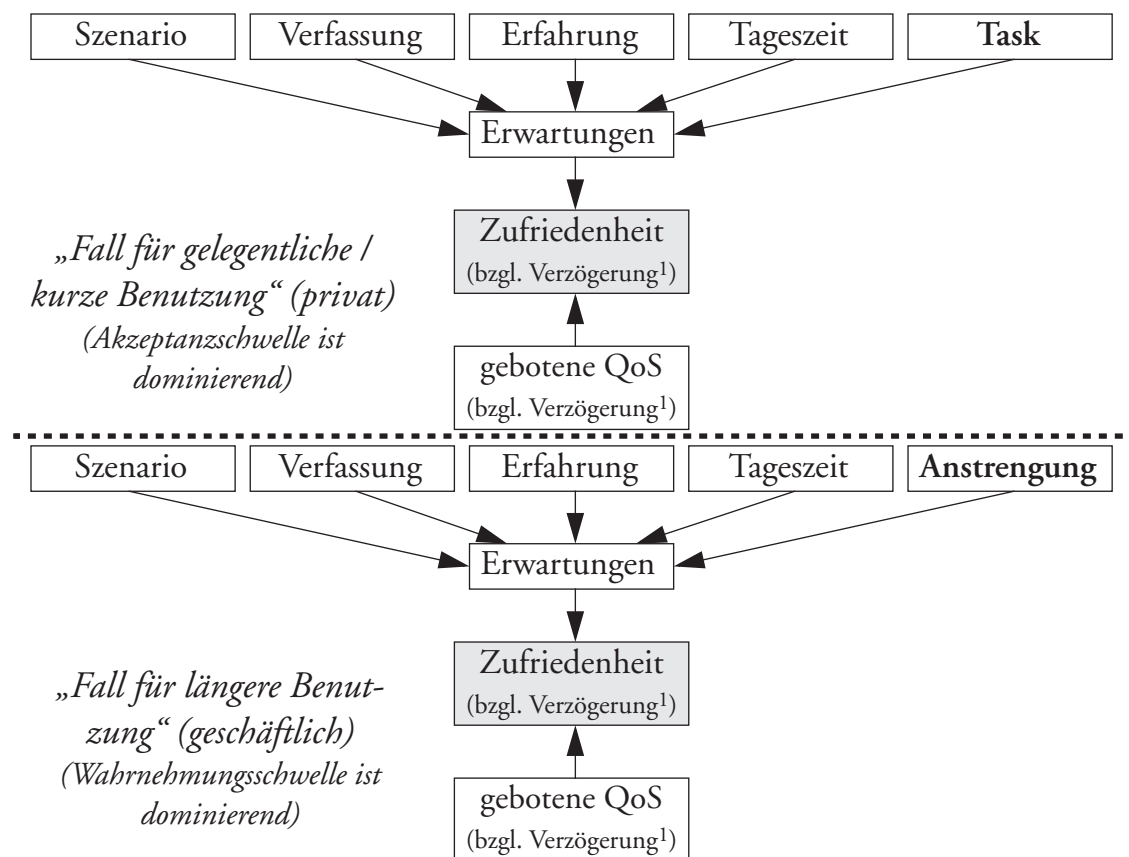
- Task
- Szenario (Anzahl Teilnehmer, Zweck der Kommunikation, etc.)
- Tageszeit

- Erfahrung der Benutzer
- Verfassung der Benutzer

Im Unterschied dazu hängt die Wahrnehmungsschwelle nicht vom Task ab, dafür aber von der Anzahl Modi, die bei der Kommunikation verwendet werden (z. B. *multimodal* mit Video und Audio oder *monomodal* mit Audio).

Bei der Auswertung der aufgezeichneten Sprecheraktivitäten hat sich herausgestellt, dass „objektiv“ gesehen besonders die Wahrnehmungsschwelle eine Rolle spielt. Sobald der Benutzer eine Verschlechterung der Qualität feststellt, ändert er sein Verhalten. Man kann somit von einer „harten“ Schwelle sprechen. Die zweite Schwelle, die hier als „weich“ bezeichnet wird, ist eher davon abhängig, wie tolerant sich der Benutzer im Moment gibt. Er behauptet zwar, dass ihn eine gewisse weitere Verschlechterung der Qualität nicht stört, sein Körper ist aber trotzdem grösserem Stress ausgesetzt und muss diesen kompensieren. Dies mag eine gewisse Dauer gut gehen, ab dann stellt aber auch dieser Benutzer eine schnellere Ermüdung oder eine abfallende Produktivität fest.

Diese Erkenntnisse führen zu einem ersten Modell mit zwei Fällen, das in Abbildung 8.35 grafisch dargestellt ist.



1} In dieser Arbeit wurde QoS nur anhand der absoluten Verzögerung gemessen, der Autor ist aber der Ansicht, dass dieses Modell auch auf weitere QoS-Parameter anwendbar und erweiterbar ist.

Abbildung 8.35: Modell der Benutzerzufriedenheit

## 8.2 *Empfehlungen für Netzwerkplaner und -gestalter*

Obwohl innerhalb dieser Arbeit nur ein kleiner Teil des angestrebten Gesamtkatalogs von denkbaren Multimediaanwendungen untersucht werden konnte, gibt es schon jetzt wichtige Erkenntnisse. Bei der Gestaltung von neuen Netzen und Diensten gibt es je nach Anwendung gewisse Verzögerungswerte, die nicht überschritten werden dürfen, ohne dass der Benutzer unzufrieden wird. Der Netzwerkplaner kann nun anhand dieses Wertes und den einzelnen Quellen, die zur Gesamtverzögerung beitragen, entscheiden, an welchen Stellen er wieviel Verzögerung einplanen will. Wie anhand von Tabelle 2.2 illustriert wurde, kann er beispielsweise mehr Verzögerung durch eine effizientere Kodierung zulassen, muss diese aber z.B. beim Traffic Shaping eventuell wieder einsparen. Er hat also viele Freiheiten bei der Gestaltung der Netze und kann für jede Anforderung eine optimale Zusammensetzung wählen.

Betreffend der Synchronität von Video- und Audiodaten sowie der Bildwiederholrate können einige Erkenntnisse aus der ersten Versuchsreihe übernommen werden:

- Wenn es nicht möglich ist, mit beiden Datenströmen unterhalb der Wahrnehmungsschwelle zu bleiben, muss je nach Anwendung abgewogen werden, ob die Synchronität zugunsten einer grösseren Gesamtverzögerung beibehalten oder aber der Ton ohne wahrnehmbare Verzögerung verarbeitet wird, währenddessen das Bild zu spät ankommt.
- Schwankungen der absoluten Verzögerungswerte sollten unter allen Umständen vermieden werden.
- Die Bildwiederholrate sollte der restlichen Qualität angepasst werden. Bei sehr guter Audio- und Videoqualität wird auch eine flüssige Wiederholrate erwartet. Werte bis hinab zu 8.33 Bilder/s sind für Videokonferenzen noch ausreichend. Bei niedrigeren Raten wird die Nützlichkeit des Videokanals in Frage gestellt.

## 8.3 *Offene Fragen und Ausblick*

Obwohl die Zusammenhänge zwischen der absoluten Verzögerung bei Multi-Mediakommunikation und der Benutzerzufriedenheit zu einem Teil erforscht werden konnten, sind weitere Untersuchungen auf diesem Gebiet nötig. Einerseits müssen die Ergebnisse dieser Arbeit durch weitere subjektive Versuche statistisch sauber abgesichert werden, unter anderem durch weitere Testreihen, in

denen die Methode 2AFC (2-alternative-forced-choice) eingesetzt wird und in denen die Initialwerte des Best-PEST Algorithmus dem Task entsprechend besser gewählt werden. Andererseits sind die Untersuchungen auf weitere QoS-Parameter auszudehnen, damit schliesslich ein komplettes Modell erstellt werden kann. Dabei ist zu beachten, dass sich die QoS-Parameter wie erwähnt gegenseitig maskieren können.

Parallel dazu muss im Bereich der Offline-Auswertung nach zuverlässigen Parametern gesucht werden, die es erlauben, Rückschlüsse auf die Benutzerzufriedenheit zu ziehen. Sobald statistisch gesehen aussagekräftigere Daten gesammelt worden sind, können die drei in dieser Arbeit untersuchten Merkmale (normierte Gesprächszeit, Double- und Tripletalk) und die betreffenden Schlussfolgerungen präzisiert und verifiziert werden.

Die vorliegende Arbeit bewegt sich in einem Gebiet, das zuerst noch komplett definiert werden muss. Es werden noch viele vergleichbare Arbeiten nötig werden bis zum ersehnten Ziel. Der Grundstein wurde nun gelegt.

# *Abkürzungen*

---

2AFC	2-Alternative-Forced-Choice
ARES	ATM Real-Time Emulation System
ATM	Asynchronous Transfer Mode
CCIF	Comité Consultatif International des Télécommunications à Grandes Distances
CCITT	Comité Consultatif International Télégraphique et Téléphonique
CSCW	Computer Supported Cooperative Work
CSG	Communication Studies Group
DiffServ	Differentiated Services
DSI	Digital Speech Interpolation
EHF	Extreme High Frequency
ETH	Eidgenössische Technische Hochschule
ETHMICS	ETH Multimedia Interactive Communication System
FDM	Frequency Division Multiplex
GUI	Graphical User Interface
IHA	Institut für Hygiene und Arbeitsphysiologie
IntServ	Integrated Services
IP	Internet Protocol
ISDN	Integrated Services Digital Network
IVR	Interactive Voice Response
JND	Just Notable Differences
MOS	Mean Opinion Score
MTFA	Modulation Transfer Function Area
OSI	Open System Interconnection
PCM	Pulse Code Modulation

---

PDA	Personal Digital Assistant
PEST	Parameter Estimation of Sequential Trials
PESQ	Perceptual Evaluation of Speech Quality
PSQM	Perceptual Speech Quality Measure
QED	Qualitätsmerkmale multimedialer Echtzeit-Dialogkommunikation
QoS	Quality of Service
RMS	Root Mean Square
RSVP	Resource Reservation Protocol
SQRI	Square Root Integral
SYMLOG	Systematic Multiple Level Observation of Groups
TCO	Total Cost Of Ownership
TCP	Transmission Control Protocol
TIK	Institut für technische Informatik und Kommunikationsnetze
UCL	University College London
UMTS	Universal Mobile Telecommunications System
VHE	Virtual Home Environment
VoIP	Voice over Internet Protocol



# Referenzen

---

- [1] Anderson, A., Newlands, A., Mullin, J., Fleming, A., Doherty-Sneddon, G. & der Velden, J. V. Impact of videomediated communication on simulated service encounters. *Interacting with Computers*, 8(2), S. 193-206, 1996
- [2] Ansoff, H. I. *Corporate Strategy. An Analytical Approach to Business Policy for Growth and Expansion*. New York: McGraw-Hill, 1965
- [3] Apteker, R., Fisher, J., Kisimov, V. & Neishlos, H. *Video Acceptability and Frame Rate*. *IEEE Multimedia*, 2, Herbst 1995
- [4] Baird, J. C., & Noma, E. *Fundamentals of scaling and psychophysics*. New York, Wiley, 1978
- [5] Bales, R. F. *How people interact in conferences*. *Scientific American*, 1955
- [6] Bales, R. F. *Social interaction systems: Theory and measurement*. New Brunswick: Transaction Publishers, 1999
- [7] Bauer, B. & Patrick, A. *A Human Factors Extension to the Seven-Layer OSI Reference Model*. Nortel Networks & Institute for Information Technology National Research Council of Canada. Draft copy., 15. April 2002
- [8] Bents, R. & Blank, R. *Der MBTI*. Claudius Verlag, 1998
- [9] Borner, N. & Braun, A. *Qualitätsmerkmale der Multimedia-Echtzeit-Dialogkommunikation*. Semesterarbeit am Institut TIK, ETH Zürich, 2001

- 
- [10] Bouch, A. & Sasse, M. *It ain't what you charge*. In Proc. IFIP/IEEE International Symposium on Integrated Network Management (IM'99), Madrid, September 1999
- [11] Bouch, A. & Sasse, M. *Network QoS: What do users need*. In IDC'99, Madrid, September 1999
- [12] Bouch, A. & Sasse, M. *Network quality of service - an integrated perspective*. In Proc. RTAs'99, Vancouver, Juni 1999
- [13] Bouch, A. & Sasse, M. *The case for predictable media quality in networked multimedia applications*. In Proc. MMCN'2000. San Jose, Kalifornien, Januar 2000
- [14] Bouch, A. *A User-centered Approach to Network Quality of Service and Charging*. Dissertation, University of London, 2001
- [15] Bouch, A., Sasse, M. & DeMeer, H. *Of packets and people: A user-centered approach to Quality of Service*. In Proceedings of IWQoS 2001
- [16] Bouch, A., Watson, A. & Sasse M. *QUASS - a tool for measuring the subjective quality of real-time multimedia audio and video*. in HCI'98 Conference Companion, 1998
- [17] Boyle, E., Anderson, A., & Newlands, A. *The effects of eye contact on dialogue and performance in a co-operative problem solving task*. Language and Speech, 37, 1, 1994
- [18] Braun, A. M. *Qualitätskriterien bei digitalem Fernsehen und Video*. Diplomarbeit am Institut für Kommunikationstechnik. ETH Zürich, 1998
- [19] Bray, J. *The Communication Miracle. The Telecommunication Pioneers from Morse to the Information Superhighway*. Plenum Press, New York, 1995
- [20] Chapanis, A. *Interactive human communication*. Scientific American, 1975

- 
- [21] Chapanis, A., Ochsmann, R., Parrish, R., & Weeks, G. *Studies in interactive communication: The effects of four communication modes on the behavior of teams during cooperative problem solving*. Human Factors, 14, 1972
- [22] Chen, T. M., Walrand, J. & Messerschmitt, D. G. Dynamic Priority Protocols for Packet Voice. IEEE Selected Areas in Communications, Juni 1989
- [23] Choi, C. A. *Techniques to Improve Upon a User's WWW Experience*. Reading Course Paper of the University of Western Ontario, 2000
- [24] Cohen, K. *Speaker interaction: video teleconferences versus face-to-face meetings*. In Proceedings of teleconferencing and electronic communications, Madison: University of Wisconsin Press, 1982
- [25] Crystal, D. *The Cambridge Encyclopedia of Language*. 2<sup>nd</sup> ed., Cambridge University Press, 1997
- [26] Dixon, N. F. & Spitz, L. *The detection of auditory visual desynchrony*. Perception, 9. 1989
- [27] Egado, C. *Videoconferencing as a Technology to Support Group Work: A Review of its failures*. Proceedings ACM Conference on Computer-Supported Cooperative Work, CSCW'88, 1988.
- [28] Estermann, P. *Verbindungs- und Ressourcenmanagement in neuen Netzwerken*. Dissertation ETH Zürich Nr. 14562, 2002
- [29] Fish, R., Kraut, R., Root R. & Rice, R. *Evaluating video as a technology for informal communication*. In Proceedings of CHI'92 Human Factors in Computing Systems, New York: ACM Press, 1992
- [30] Fish, R., Kraut, R., Root R. & Rice, R. *Video as a technology for informal communication*. Communications of the ACM. 36, 1993
- [31] Fukuda, K. Wakamiya, N., Murata, M. & Miyahara, H. QoS Mapping between User's Preference and Bandwith Control for Video Transport. In Proceedings of the International Workshop on QoS May 97, S. 291-302, 1997

- 
- [32] Frey, S. *Ein Komponentenmodell für die Verbindungssteuerung in Multimedia-Netzen*. Dissertation ETH Zürich Nr. 12763, 1997
- [33] Gescheider, G. A. *Psychophysics: The Fundamentals*. Lawrence Erlbaum Associates, Mahwah, NJ, 1997
- [34] Gieske, R. *Deutschland 2010 Szenarien Was wird der 11. September mit Deutschland machen?* ICEBERG Consulting, 2001.  
[http://www.iceberg-consulting.de/download\\_deutschland\\_2010\\_szenarios.pdf](http://www.iceberg-consulting.de/download_deutschland_2010_szenarios.pdf)
- [35] Gonsalves & Tobagi Comparative Performance of Voice/Data Local Area Networks. IEEE Selected Areas in Communications, Juni 1989
- [36] Habermas, J. *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*. suhrkamp taschenbuch wissenschaft 1176, 1995
- [37] Helder G. K. *Customer Evaluation of Telephone Circuits with Delay*. Bell System Technical Journal 38, 1966
- [38] Isaacs, E., & Tang, J. *What video can and can't do for collaboration: a case study*. In Proceedings of the ACM Multimedia 93 Conference. Anaheim., CA, 1993
- [39] ITU-T Recommendation G.107. *The E-Model, a Computational Model for Use in Transmission Planning*. International Telecommunication Union, Genf, 1998
- [40] ITU-T Recommendation G.114. *General Characteristics of International Telephone Connections and International Telephone Circuits: One-Way Transmission Time*. International Telecommunication Union, Genf, 1996.
- [41] Ives, H. *Two Way Television, Bell Labs Record*, 8. 1930
- [42] Kahneman, D. *Attention and Effort*. New Jersey: Prentice Hall, 1973

- 
- [43] Keshav, S. *An Engineering Approach to Computer Networks*. Addison Ewsley, 1997.
- [44] Kienholz, M., Braun, A. M. & Estermann, P. *QoS-Parameter bei Multimedia Kommunikation. Der Einfluss von Verzögerungen bei Videokonferenzen auf die Benutzerzufriedenheit*. Semesterarbeit am Institut TIK, ETH Zürich, 2002
- [45] Kleinke, C. *Gaze and eye contact: a research review*. Psychological Bulletin, 100, 1986
- [46] Kouvelas, I., Hardman, V. & Watson, A. Lip Synchronisation for use over the Internet: Analysis and Implementation. In Proceedings of IEEE Globecom '96, London, UK, 1996
- [47] Kraut, R., Fish, R., Root, B. & Chalfonte, B. *Informal communication in organizations*. In R. Baecker Ed., Goupware and Computer Supported Cooperative Work. San Mateo, Kalifornien: Morgan Kaufman, 1993
- [48] Kündig, A., Zuberbühler, H.-J. & Braun, A. QED-Report 2 – State of the Art. Key Parameters – their Relevance and their Determination. Institut TIK, ETH Zürich, 2000
- [49] Kündig, A. *A Basis for IT Assessment*. Zentrum für Technologiefolgen-Abschätzung, Bern, September 2002
- [50] Kurmann, H. M. *On the Emulation of Impairments in ATM-Networks*, Dissertation ETH Zürich Nr. 12362, 1997
- [51] Lamparter, O. *Ein Softwaresystem für die Verkehrsdatenanalyse in Kommunikationsnetzwerken*. Dissertation ETH Zürich Nr. 15070, 2003
- [52] Lewkowicz, D. J. *Perception of auditory-visual temporal synchrony in human infants*. Journal of Experimental Psychology: Human Perception and Performance, 22. 1996
- [53] Liebermann, H. & Pentland, A. Microcomputer-based estimation of psychophysical thresholds: The Best PEST. Behavior Research Methods and Instrumentation, 14(1): S. 21-25, 1982

- [54] Masoodian, M. & Apperley, M. *User Perceptions of Human-to-Human Communication Modes in CSCW Environments*. Aus ED-MEDIA 95 Conference Proceedings, 1995
- [55] Matarazzo, G. The effects of video quality and task on remote collaboration. A laboratory experiment. In Proceedings of 17th International Symposium on Human Factors in Telecommunications, Copenhagen May 1999, S. 143-148
- [56] McGrath M. & Summerfield, Q. *Intermodal timing relations and audio-visual speech recognition by normal-hearing adults*. J. Acoust. Soc. Am., 77. 1985
- [57] Minoli, D. *Satellite on-board processing of packetized voice*. Proc. ICC'79, paper 58.4, 1979
- [58] Möller, S. *Assessment and Prediction of Speech Quality in Telecommunications*. Kluwer Academic Publishers, Boston, 2000
- [59] Oberholzer, G., Braun, A. M. & Estermann, P. *How Long Is Too Long? The Influence of Absolute Delay on User Perception in Video-Conferencing*. Diplomarbeit am Institut TIK, ETH Zürich, 2002
- [60] O'Conaill, B., Whittaker, S., & Wilbur, S. *Conversations over videoconferences: an evaluation of the spoken aspects of video mediated interaction*. Human Computer Interaction, 8, 1993
- [61] O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G. & Bruce, V. *Comparison of face-to-face and video-mediated interaction*. Interacting with Computers, Vol. 8, no 2, S. 177-192, 1996
- [62] Pandey, P. C., Kunov, H. & Abel, S. *Disruptive effects of auditory signal delay on speech perception with lipreading*. The Journal of Auditory Research, 26. 1986
- [63] REID, A. *Comparing the telephone with face-to-face interaction* In I. Pool Ed., *The Social Impact Of The Telephone*. Cambridge, MA, MIT Press, 1977

- 
- [64] Renggli, P., Braun, A. M. & Estermann, P. *Qualitätsmerkmale der Multimedia-Echtzeit-Kommunikation*. Diplomarbeit am Institut TIK, ETH Zürich, 2002
- [65] Braden, R., Clark, D. & Shenker, S. *Integrated Services in the Internet Architecture: An Overview*. RFC 1633, Juli 1994.
- [66] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. & Weiss, W. *An Architecture for Differentiated Services*. RFC 2475, Dezember 1998
- [67] Röthlisberger, U. *The Architecture of an Interactive Multimedia Communication System*. Dissertation ETH Zürich Nr. 12796, 1998
- [68] Ruesch, J. & Bateson, G. *Communication: The Social Matrix of Psychiatry*. W. W. Norton & Co., New York, 1951
- [69] Scherer, P., Braun, A. M. & Estermann, P. *Multimediasoftware für Qualitätsuntersuchungen im ATM-Netz*. Diplomarbeit DA-99.13 am Institut für Technische Informatik und Kommunikationsnetze, ETH Zürich, 1999
- [70] Schilder, H.-J. *Wohin der Trend geht*. Partner-Meeting von Natural Microsystems, 2000.  
[http://www.telekom3.de/de-p/pres/2-pr/2001/03-m/010322-innovationsfuehrerschaft-ar,templateId=\\_2Fdt\\_2Fweb\\_2Fmod\\_2FModul-PM3-print-load.jsp.html](http://www.telekom3.de/de-p/pres/2-pr/2001/03-m/010322-innovationsfuehrerschaft-ar,templateId=_2Fdt_2Fweb_2Fmod_2FModul-PM3-print-load.jsp.html)
- [71] Sellen, A. *Remote Conversations: The Effects of Mediating Talk With Technology*. Human Computer Interaction, Vol. 10, No. 4, S. 401-444, 1995
- [72] Shannon, C. E. *A Mathematical Theory of Communication*. The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, 1948.
- [73] Shannon, C. E., Weaver, W. *Mathematische Grundlagen der Informationstheorie*. München, 1976

- [74] Short, J, Williams, E. & Christie, B. *The social psychology of telecommunications*. London: Wiley, 1976
- [75] Simon, H. A. *The Shape of Automation for Men and Management*. New York: Harper and Row, 1960
- [76] Smith, R. L., Richetto, G. M. & Zima, J. P. *Organizational behaviour: an approach to human communication*. In Approaches to Human Communication, R. W. Budd and B. D. Ruben, Eds. New York: Spartan Books, 1972
- [77] Snyder, F. Travel Patterns: Implication for new Communication Facilities, 1971. Bell Laboratories Memorandum.
- [78] Sommer, C. M. ALLES TREND ODER WAS? Über Moden, Trends und Trendforschung. Sommer Research Stuttgart, 1995. [http://www.sommer-research.com/Texte/Trend\\_D.PDF](http://www.sommer-research.com/Texte/Trend_D.PDF)
- [79] Steinmetz, R. Human Perception of Jitter and Media Synchronization. IEEE Journal on Selected Areas in Communications 14. 1996
- [80] Stevens, W. R. *TCP/IP Illustrated, Volume 1*. Addison-Wesley, 2000
- [81] Suchman, L. *Constituting shared workspaces*. In D. Middleton & Y. Engstrom Eds., Advances in information processing in organisations. Greenwich, Connecticut: JAI Press, 1992.
- [82] Talley, T. M. *A Transmission Control Framework for Continuous Media*. PhD thesis, University of North Carolina, 1997
- [83] Tang, J. & Isaacs, E. *Why do users like video: studies of multimedia-supported collaboration*. Computer Supported Cooperative Work, 1, 1993.
- [84] Thorngren, B. *Studier: Lokalisering*. Ekonomiska Forskninginstitutet, Stockholm, 1972.
- [85] Watson, A. & Sasse, M. *Evaluating audio and video quality in low-cost multimedia conferencing systems*, 1996



- 
- [86] Watson, A. & Sasse, M. *Measuring Percieved Quality of Speech and Video in Multimedia Conferencing Applications*. In Proceedings of ACM Multimedia, 1998
- [87] Watson, A. *Evaluating real-time multimedia audio and video quality*, 1997
- [88] Watzlawick, P. & Beavin, J. H. *Einige formale Aspekte der Kommunikation*. In Soziologie der Kommunikation: Eine Textauswahl zur Einführung, B. Badura and K. Gloy, Eds. Stuttgart: Frommann, 1966
- [89] Watzlawick, P., Bavelas, J. B. & Jackson, D. D. *Pragmatics of Human Communication*. W. W, Norton & Co., New York, 1967
- [90] Whittaker, S. *Rethinking Video as a Technology for Interpersonal Communications: Theory and Design Implications*. Journal of Uman-Computer Systems, 1999
- [91] Wilson, G. & Sasse, M. *Do users always know what's good for them*. In Proceedings of HCI 2000, Sunderland UK, September 1998
- [92] Bitkom: *Wege in die Informationsgesellschaft. Status quo und Perspektiven Deutschlands im internationalen Vergleich*. Edition 2003. <http://www.bitkom.org/>
- [93] ESPON Project 1.2.2: *Telecommunication And Energy Services And Networks: Territorial Trends And Basic Supply Of Infrastructure For Territorial Cohesion*. 2002  
[http://www.espon.lu/online/documentation/projects/thematic/96/tor\\_1.2.2.pdf](http://www.espon.lu/online/documentation/projects/thematic/96/tor_1.2.2.pdf)
- [94] Graduiertenkolleg *Technisierung und Gesellschaft* der TU Darmstadt.  
<http://www.ifs.tu-darmstadt.de/gradkoll/Beschreibung2.html>
- [95] Heise Online News: *Forscher formulieren "Megatrends" der Netz-Gesellschaft*. 2003.  
<http://www.heise.de/bin/nt.print/newsticker/data/wst-22.01.03-003/?id=364e6063&todo=print>

- [96] International Labour Office Geneva: *Report for discussion at the Tripartite Meeting on Employment, Employability and Equal Opportunities in the Postal and Telecommunications Services*. Genf, Mai 2002.  
<http://www.ilo.org/public/english/dialogue/sector/techmeet/tmpts02/tmpts-r.pdf>
- [97] Light Reading - Networking the Telecom Industry: *ITU: Internet Gap Widens*. 2003.  
[http://www.lightreading.com/document.asp?doc\\_id=12943](http://www.lightreading.com/document.asp?doc_id=12943)
- [98] Pressemitteilung: *Deutsche Telekom untermauert Innovationsführerschaft*. CeBIT 2001.  
[http://www.telekom3.de/de-p/pres/2-pr/2001/03-m/010322-innovationsfuehrerschaft-ar,templateId=\\_2Fdt\\_2Fweb\\_2Fmod\\_2FModul-PM3-print-load.jsp.html](http://www.telekom3.de/de-p/pres/2-pr/2001/03-m/010322-innovationsfuehrerschaft-ar,templateId=_2Fdt_2Fweb_2Fmod_2FModul-PM3-print-load.jsp.html)
- [99] RF design: ITU offers insights into 2002 communications trends. Online Exclusive, 2002.  
<http://rfdesign.com/newsarticle.asp?Newsarticleid=300370&SiteID=31&magazineid=4&mode=print>
- [100] Yamaguchi, H., Wada, M. & Yamamoto, H. *A 64 kbit/s Integrated Visual Communication System, - New Communication Medium for the ISDN*. IEEE Selected Areas in Communications, November 1986
- [101] Zhang, L., Deering, S., Estrin, D., Shenker, S. & Zappala, D. *RSVP: A New Resource ReSerVation Protocol*, IEEE Network Magazine, Vol. 7, No. 5, September 1993, S. 8-19
- [102] Zuberbühler, H.-J. *Rapid Evaluation of Perceptual Thresholds*. The Best-Pest Calculator: A web-based application for non-expert users", software-manual published on <http://www.psychophysics.ethz.ch/>, 2000
- [103] Zuberbühler, H.-J. *Quality Aspects of Multimodal Communication: User Perception and Acceptance Thresholds*. Dissertation ETH Zürich Nr. 15124, 2003

# Glossar

---

## Definitionen

### *Best-Effort-Prinzip (best effort delivery)*

Bezeichnung für Netzwerktechnologien die keine Datensicherungsmechanismen in der Sicherungsschicht haben. Wörtlich übersetzt heisst *Best Effort Delivery* „Auslieferung nach bestem Bemühen“. In den TCP/IP-Protokollen funktioniert diese Technik zufriedenstellend, weil das IP-Protokoll davon ausgeht, dass das zugrunde liegende Netz eine nicht gesicherte Verbindung verbindungslos zur Verfügung stellt. Das IP-Protokoll zusammen mit dem UDP-Protokoll stellt einen Best-Effort-Delivery-Service für Anwenderprogramme bereit.

Da die Internet-Protokolle ursprünglich nur für den Transport von zeitunkritischen Daten gedacht waren, so etwa für E-Mails oder File-Transfer, ist das *Best-Effort-Prinzip* für die Übertragung von zeitkritischen Daten wie für VoIP oder Videokonferenzen nicht mehr aufrecht zu halten. Aus diesem Grund wurden in den Standards für die Dienstgüte Datenpakete in drei Klassen eingeteilt.

Das Best-Effort-Prinzip soll von IntServ abgelöst werden, da es den Echtzeit-Anwendungen nicht gerecht wird.

### *Dialog*

Dialog stammt ab vom Griechischen *dialogos* (*dia* = *durch*, *logos* = *Vernunft*). Es hat nichts zu tun mit dem griechischen *di* (*zwei*, *doppelt*; vgl. *Duolog*). Dialog ist 'Gespräch im Gegensatz zu Monolog'; an einem Dialog können zwei oder mehr Gesprächspartner Teil haben.

## *Internet*

Das Internet ist der weltweit grösste Netzverbund, der jedem Teilnehmer eine nahezu grenzenlose Informations- und Kommunikationsinfrastruktur zur Verfügung stellt.

Es handelt sich um ein dezentral organisiertes, globales Rechnernetz, das aus sehr vielen miteinander verbundenen lokalen und nationalen Netzen besteht. Der Informationsaustausch zwischen den einzelnen Rechnern erfolgt auf der Grundlage des TCP/IP-Protokolls.

## *Multimedia*

Als Begriff ist Multimedia vieldeutig: Die wörtliche Übersetzung »mehrere Medien« lässt offen, welche gemeint sind. Als »Bausteine« für die Integration in Multimedia-Dokumente kommen Texte, Grafiken, Festbilder, Digitalfotos, Audio, Video, Animationen, Simulationen und Interaktionen in Frage.

Text ist nach wie eine der wichtigsten Grundelemente, da andere Medienobjekte in Text eingebettet werden können. Grafiken sind fester Bestandteil von Dokumenten. Das WYSIWYG-Prinzip hat aber auch den Textteil deutlich verändert - durch Schriften in typografischer Qualität.

Festbilder gehören bei den verschiedenen Window-Systemen auch schon zum festen Bestand. Allerdings handelt es sich dabei um Schwarzweissbilder oder um Grau- bzw. Farbbilder mit 256 verschiedenen Werten, das heisst 8 bit pro Pixel. Schwieriger ist die Integration bei grösserer Pixeltiefe. Digitale Fotos: In ihnen wird die volle Farbinformation - 8 bit für jede RGB-Farbe (Rot, Grün, Blau) - gespeichert.

Audio-Elemente in ein Dokument zu integrieren ist schon deshalb eine anspruchsvolle Aufgabe, weil sie - wie auch Video-Komponenten - eine zeitliche Dimension haben. Meist deutet in einem Dokument ein kleines Audio-Piktogramm darauf hin, dass hier eine Audio-Annotation vorliegt. Ein Doppelklick mit der Maus spielt die Aufnahme ab. Video lässt sich über ein eigenes Fenster in ein Dokument integrieren. Dieses Window zeigt ein Titelbild, und wiederum löst ein Doppelklick das Abspielen aus, diesmal des Video-Clips.

Der spannendste Teil fehlt in dieser Aufzählung noch. Erst durch folgende Features wird Multimedia zum zukunftsweisenden Instrument professioneller Informationsverarbeitung: Animation sind Texte, Grafiken oder Festbilder in Bewegung. Die festen Bilder, wie wir sie etwa von Dia-Präsentationen gewöhnt

sind, werden bei Multimedia-Präsentationen animiert - nämlich mit Bewegung versehen. Interaktion ist von entscheidender Bedeutung. Anders als beim Fernsehen ist der Benutzer nicht mehr in eine passive Rolle gedrängt. Er greift vielmehr aktiv in das Geschehen ein: Je nach Reaktion oder Antwort wird ein anderer Weg verfolgt. Interaktives Material erlaubt eine Vielzahl von möglichen Pfaden durch das Spiel, den Lehrstoff, den Informationspool.

Simulation in Multimedia-Dokumente einzubauen heisst, ein Computerprogramm als Objekt von dort aufrufbar zu machen. Das kann beispielsweise ein mathematisches Modell sein, das dem Benutzer erlaubt, diverse Parameter zu manipulieren und die resultierenden Effekte zu studieren. Zielrichtung ist es, die vollständige Integration der Computerwelt in Multimedia-Dokumente zu ermöglichen. Gerade die Animation, Interaktion und Simulation stellen die eigentliche Rechtfertigung für Multimedia dar. Untersuchungen haben ergeben, dass bei Präsentationen mit nur einem Medium - Sprache oder Text - lediglich 20 Prozent der Information behalten werden. Verbindet man mehrere Medien in einer Präsentation, verdoppelt sich der Prozentsatz. Lässt man aber Interaktion zu und verwendet Animationen sowie Simulationen, erhöht sich der Wert des Erinnerten auf mehr als 70 Prozent.

Wo die Multimedia-Zukunft bereits begonnen hat, sind die Bereiche Ausbildung, Verkaufsförderung bei direktem Kundenkontakt und elektronische Bildkonferenzen im Büro. Grosse Firmen lassen heute schon interaktive Lehrfilme zu Schulungszwecken entwickeln, bei denen der Benutzer selbst entscheiden kann, in welchen Schritten er durch das Lernprogramm geht. Ähnlich ist das Verfahren bei Touristikunternehmen, Autohäusern, Immobilienhändlern, Grafikbüros und Werbeagenturen: Gemeinsam mit dem beratenden Personal entscheidet sich der Kunde am Bildschirm für die Eigentumswohnung, den Neuwagen oder das Feriendomizil seiner Wahl. Er sieht, was er bekommt, am Schirm und hat zugleich die Daten dazu. Einen grossen Stellenwert wird die Multimedia eindeutig im kreativen Bereich einnehmen. Sowohl für Produktion, Entwicklung als auch für Präsentation greifen immer mehr Agenturen, Grafiker und Designer auf multimediale Werkzeuge zurück.

Komplette Bilddatenbanken mit Ton, zum Beispiel im medizinisch-dokumentarischen Bereich oder als Privatbibliothek auf CD-ROM, markieren zur Zeit die Schwelle der Zukunft. Die »persönliche Zeitung«, die täglich nach den informellen Bedürfnissen des Anwenders zusammengestellt und auf einer wiederbeschreibbaren CD-ROM abgespeichert wird, ist bereits heute möglich. Multimedia eröffnet eine ganz neue Ebene der Informationsverarbeitung: Videosequenzen, Texte, Toninformationen, Computer-Grafiken und -Daten

können frei nach Belieben gemixt, geändert oder in anderer Form aufbereitet werden. Dies bietet weitreichende, individuelle Anwendungsmöglichkeiten, die bis vor kurzem nur professionellen Studios mit sechsstelligen Investitionszahlen vorbehalten waren.

### *Quality of Service*

Unter Dienstgüte (QoS) versteht man alle Verfahren die den Datenfluss in LANs und WANs so beeinflussen, dass der Dienst mit einer festgelegten Qualität beim Empfänger ankommt. Es handelt sich also um die Charakterisierung eines Dienstes, der für den Nutzer unmittelbar »sichtbar« ist und dessen Qualität er messen kann. Technisch handelt es sich um eine Parametrisierung von Protokollen zur Bestimmung des Übertragungsverhaltens für bestimmte Dienste.

### *Telekommunikation (TK)*

Begriff der klassischen Nachrichten- oder Kommunikationstechnik, die sich hauptsächlich mit Sprach- und Bewegtbild-Übertragung auf Basis analoger Elektronik beschäftigt, angefangen vom Telefon bis zum Nachrichtensatelliten. Sie beinhaltet jede Übertragung, Sendung oder den Empfang von Zeichen, Signalen, Schriftbildern oder Tönen und Nachrichten, gleich welcher Art, mittels Leitungen, Radio oder optischen sowie anderen, elektromagnetischen Systemen.

Um Telekommunikation durchführen zu können, müssen Übertragungswege vorhanden sein und Verbindungen zwischen den Teilnehmern hergestellt werden. Dies erfolgt durch Telekommunikationsnetze. Je nach Vermittlungsprinzip unterscheidet man bei den Weitverkehrsnetzen zwischen Netzwerken mit Leitungsvermittlung und Datenpaketvermittlung, sowie zwischen herkömmlichen Netzen und solchen für Mobilfunk.

Neben den Netzen ist die Telekommunikation geprägt durch die Telekommunikationsdienste. Unter einem Telekommunikationsdienst ist ein bestimmtes Angebot eines Anbieters zur Durchführung der Kommunikation zu verstehen. Man unterscheidet zwischen den reinen Übertragungs- und Vermittlungsdiensten, den so genannten Transportdiensten oder Übermittlungsdiensten, den Telediensten, bei denen die Endgeräte berücksichtigt werden und den Mehrwertdiensten.

Im Konsumentenumfeld gibt es als neue Dienste die so genannten Verteildienste und die interaktiven Dienste. Zu der erstgenannten Gruppe gehören u.a. Pay-per-View, Pay-per-Channel, zu der zweiten Gruppe u.a. Video-on-Demand, Videogames, Teleshopping.

# Lebenslauf

---

## *Angaben zur Person*

Name: Braun  
Vornamen: Alexander Martin  
Adresse: Neugasse 97/17, 8005 Zürich  
E-Mail: Alexander.Braun@ethz.ch  
Geburt: 26. Juni 1973 in Lörrach (D)

## *Ausbildung*

1980-1984: Grundschule in Eimeldingen  
1984-1993: Kant Gymnasium in Weil am Rhein, Abschluss mit dem allgemeinbildenden Abitur  
1993-1998: Studium der Elektrotechnik an der ETH Zürich  
Vertiefung in Digital- und Kommunikationstechnik  
Semesterarbeiten: Digitales Satellitenfernsehen / MPE4 Audio  
Diplomarbeit: Qualitätskriterien bei digitalem Fernsehen und Video  
1998-2003: Doktorand am Institut TIK

## *Berufliche Erfahrung*

1998-2003: Wissenschaftlicher Mitarbeiter am Institut für technische Informatik und Kommunikationsnetze (TIK), Fachgruppe Systemtechnik, Prof. Dr. A. Kündig





# *TIK-Schriftenreihe*

---

- Nr. 1  
**Kausale Beschreibung von Systemen**  
Hans Peter Gisiger  
1993, ISBN 3-7281-2050-2
- Nr. 2  
**Compiling Sisal for the ADAM Architecture**  
Srdjan Mitrovic  
1993, ISBN 3-7281-2051-0
- Nr. 3  
**Compiling the Behaviour Part of LOTOS**  
Eric Emile Dubois  
1994, ISBN 3-7281-2054-5
- Nr. 4  
**The ADAM Architecture and its Simulation**  
Olivier Maquelin  
1994, ISBN 3-7281-2113-4
- Nr. 5  
**GIPSY: Ein Ansatz zum Entwurf integrierter Softwareentwicklungssysteme**  
Reto Marti  
1994, ISBN 3-7281-2114-2
- Nr. 6  
**Teilautomatische Herstellung von Codegeneratoren für Mikrokontroller**  
Hans Bärffuss  
1994, ISBN 3-7281-2121-5
- Nr. 7  
**SVOX: The Implementation of a Text-to-Speech System for German**  
Christof Traber  
1995, ISBN 3-7281-2239-4
- Nr. 8  
**Transparent Fault-Tolerance for Process Control Systems**  
Thomas Siegrist  
1995, ISBN 3-7281-2242-4
- Nr. 9  
**Routing in Weitraum-Rechnernetzen**  
Horst Hodel  
1996, ISBN 3-7281-2216-5
- Nr. 10  
**Execution Analysis of Real-Time Software Running on Heterogeneous Distributed Computer Systems**  
Andreas Danuser  
1995, ISBN 3-7281-2284-X
- Nr. 11  
**Prosodische Modifikation von Sprachsegmenten für die konkatenative Sprachsynthese**  
Beat Pfister  
1996, ISBN 3-7281-2316-1
- Nr. 12  
**A Framework for Dynamic Protocol Configuration**  
Thomas Peter Plagemann  
1996, ISBN 3-7281-2334-X
- Nr. 13  
**Execution Architecture of the Multithreaded ADAM Prototype**  
Philipp Färber  
1996, ISBN 3-7281-2384-6
- Nr. 14  
**Towards Computer Supported Audio Conferencing**  
Christoph A. Burkhardt  
1996, ISBN 3-7281-2386-2
- Nr. 15  
**Comparison of Classic and Hybrid HMM Approaches to Speech Recognition over Telephone Lines**  
Hans-Peter Hutter  
1997, ISBN 3-7281-2424-9
- Nr. 16  
**An Object Oriented Petri Net Approach to Embedded System Design**  
Robert Esser  
1997, ISBN 3-7281-2416-8
- Nr. 17  
**Theory of Evolutionary Algorithms and Application to System Synthesis**  
Tobias Blickle  
1997, ISBN 3-7281-2433-8
- Nr. 18  
**Eine objektorientierte Softwarearchitektur für Multimedia-Kommunikationssysteme**  
Daniel P. Ingold  
1997, ISBN 3-906469-00-X
- Nr. 19  
**Verifikation von Implementierungen abstrakter Datentypen für sicherheitskritische Systeme am Beispiel von Oberon**  
Daniel Schweizer  
1997, ISBN 3-906469-01-8
- Nr. 20  
**Group and Session Management for Collaborative Applications**  
Erik Wilde  
1997, ISBN 3-8265-2411-X
- Nr. 21  
**Project GIPSY: Facing the Challenge of Future Integrated Software Engineering Environments**  
Tobias Murer  
1997, ISBN 3-906469-02-6
- Nr. 22  
**Rapid-Prototyping eingebetteter Systeme**  
Ruedi Mattmann  
1997, 3-7281-2509-1

- Nr. 23  
**A Multipoint Communication Architecture for End-to-End Quality of Service Guarantees**  
Daniel Bauer  
1997, ISBN 3-8265-2638-4
- Nr. 24  
**On the Emulation of Impairments in ATM-Networks**  
Hans Kurmann  
1997, ISBN 3-7281-2590-3
- Nr.25  
**Emulation of LAN MAC Service for ATM Networks**  
Andreas X. Meier  
1997, ISBN 3-906469-04-2
- Nr.26  
**Controlling Segmental Duration in Speech Synthesis Systems**  
Marcel P. Riedi  
1997, ISBN 3-906469-05-0
- Nr. 27  
**Ein Komponentenmodell für die Verbindungssteuerung in Multimedia-Netzen**  
Stefan Frey  
1997, 3-906469-06-9
- Nr. 28  
**The Architecture of an Interactive Multimedia Communication System**  
Urs Röthlisberger  
1998, 3-906469-07-7
- Nr. 29  
**Internet-wide Software Component Development Process and Deployment Integration**  
Daniel Scherer  
1998, ISBN 3-906469-08-5
- Nr. 30  
**Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications**  
Eckart Zitzler  
1999, ISBN 3-8265-6831-1
- Nr. 31  
**Fast Longest Prefix Matching: Algorithms, Analysis, and Applications**  
Marcel Waldvogel  
1999, ISBN 3-8265-7312-9
- Nr. 32  
**Synchronisation in Distributed Multimedia Applications**  
Christina Class  
1999, ISBN 3-8265-6906-7
- Nr. 33  
**Parallelising Molecular Dynamics for Message Passing Systems**  
Martin Gerber  
1999, ISBN 3-8265-7225-4
- Nr. 34  
**Service Management in a Telecom Environment based on Active Network Technology**  
Markus Brunner  
1999, ISBN 3-8265-7583-0
- Nr. 35  
**Dynamic Security in Communication Systems**  
Germano Caronni  
1999, ISBN 3-8265-6809-5
- Nr. 36  
**Symbolic Methods Applied to Formal Verification and Synthesis in Embedded Systems Design**  
Karsten Strehl  
2000, ISBN 3-8265-7193-2
- Nr. 37  
**A Network Architecture Based on Market Principles**  
Georg Fankhauser  
2000, 3-8265-8093-1
- Nr. 38  
**Syntax and semantics of graph: An approach to the specification of visual notations for diskrete-event systems**  
Jörn Janneck  
2000, ISBN 3-8265-7688-8
- Nr. 39  
**Separate Connection and Functionality is the Pivot in Embedded System Design**  
Hans Otto Trutmann  
2000, ISBN 3-906469-09-3
- Nr. 40  
**On the Modeling and Evaluation of Real-Time Systems**  
Martin Naedele  
2000, ISBN 3-8265-7150-9
- Nr. 41  
**Algorithm-Architecture Trade-offs in Network Processor Design**  
Matthias Gries  
2001, ISBN 3-8265-9044-9
- Nr. 43  
**Modular Language Specification and Composition**  
Christoph Denzler  
2001, ISBN 3-8265-9157-7
- Nr. 44  
**Verbindungs- und Ressourcenmanagement in neuen Netzwerken**  
Patrik Erstermann  
2002, ISBN 3-906469-11-5
- Nr. 45  
**Communication Channel Synthesis for Heterogeneous Embedded Systems**  
Michael Eisenring  
2002, ISBN 3-7281-2867-8
- Nr. 46  
**SANGRIA - Secure ANon-nomous GRoup InfrAstructure**  
Nathalie Weiler  
2002, ISBN 3-8322-0904-2
- Nr. 47  
**Ein Datenakquisitionssystem für passive Messungen in Telekommunikationsnetzwerken**  
Bernard Stauffer  
2003, ISBN 3-8322-1735-5
- Nr. 48  
**A Proxy-Based Content Adaptation Framework**  
Marius Portmann  
2002, ISBN 3-906469-12-3
- Nr. 49  
**Ein Softwaresystem für die Verkehrsdatenanalyse in Kommunikationsnetzwerken**  
Oliver Lamparter  
2003, ISBN 3-8322-1565-4