Effizienz und Effektivität eines 3D-Telekonferenz-Systems

Michael Lüftl, Martin Rothbucher, Klaus Diepold

Technical Report

Effizienz und Effektivität eines 3D-Telekonferenz-Systems

Michael Lüftl, Martin Rothbucher, Klaus Diepold

6. November 2013

Lehrstuhl für Datenverarbeitung
Technische Universität München

Michael Lüftl, Martin Rothbucher, Klaus Diepold. Effizienz und Effektivität eines 3D-Telekonferenz-Systems. . Technische Universität München, LDV. 6. November 2013.

Dieses Werk ist unter einem Creative Commons Namensnennung 3.0 Deutschland Lizenzvertrag lizenziert. Um die Lizenz anzusehen, gehen Sie bitte zu `http://creativecommons.org/licenses/by/3.0/de/` oder schicken Sie einen Brief an Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

Inhaltsverzeichnis

1 Einleitung **7**

2 Grundlagen der auditiven Wahrnehmung **9**
2.1 Der Hörsinn . . . 9
2.2 Räumliches Hören . . . 11
2.3 Der Cocktailparty-Effekt . . . 13

3 Das Telekonferenzsystem am LDV **19**

4 Psychische Belastung **23**
4.1 Überblick und Einordnung . . . 23
4.2 Der Begriff psychische Belastung . . . 24
4.2.1 Arbeitswissenschaften . . . 25
4.2.2 Cognitive Load Theory . . . 26

5 Indikatoren psychischer Belastung **29**
5.1 Somatische Indikatoren . . . 30
5.2 Subjektive Schätzung . . . 31
5.2.1 Unidimensionale Skalen . . . 31
5.2.2 Multidimensionale Skalen . . . 34
5.3 Leistung . . . 35
5.3.1 Primäraufgaben . . . 35
5.3.2 Sekundäraufgaben . . . 36

6 Versuchsaufbau **39**
6.1 Überblick . . . 39
6.2 Stimuli . . . 40
6.2.1 Konferenz-Korpus (**?**) . . . 40
6.2.2 Aufnahme . . . 41
6.2.3 Sprechertrennung . . . 42
6.2.4 Wiedergabe . . . 43
6.3 Indikatoren / Messung . . . 44
6.3.1 Primäraufgabe . . . 45
6.3.2 Subjektive Schätzung . . . 46
6.3.3 Reaktionstest (Teil ②) . . . 47

6.4 Ablauf 47
6.5 Hypothesen 49

7 Ergebnisse **51**
7.1 Teil ① 52
7.2 Teil ② 56

8 Fazit **59**

Literaturverzeichnis **63**

Abstract

Eine Telekonferenz ist nicht dasselbe wie ein echtes Gespräch. Zum Beispiel ist es manchmal schwer zu erkennen, wer gerade spricht. Deswegen forscht man am Lehrstuhl für Datenverarbeitung der TUM an einem Telekonferenzsystem mit binauraler Wiedergabe. Es soll nun geklärt werden, ob ein solches „3D-System“ einen Vorteil gegenüber einem klassischen „Mono-System“ aufweist. Hierfür können neben Tests zur Nutzerpräferenz (Quality of Experience) auch andere Kriterien herangezogen werden. Eine Möglichkeit ist zu fragen, ob die psychische Belastung beim Teilnehmer geringer ist und die Kommunikation damit effizienter und effektiver abläuft?

Diese Arbeit leistet hierzu zwei Beiträge: Zum einen wird die Bedeutung des Begriffs psychische Belastung geklärt und Möglichkeiten zur Messung bei Höraufgaben gezeigt. Zum anderen wird ein Hörversuch durchgeführt, der mögliche Vorteile nachweisen soll. In diesem Versuch hören die Teilnehmer Konferenzen und beantworten im Anschluss Fragen zum Inhalt und zu ihrer gefühlten Anstrengung.

Die Ergebnisse zeigen eine signifikante Verbesserung verglichen mit dem Mono-System. Die Teilnehmer müssen sich weniger konzentrieren und erkennen leichter, wer gerade spricht. Außerdem ist im Nachhinein klarer, wer welchen Beitrag geleistet hat.

—

Conducting a teleconference is not the same as holding a real meeting. For instance, sometimes it may be difficult to figure out who is talking. Therefore a teleconferencing system is being developed at the Institute for Data Processing at TUM, that utilizes spatial sound. Hence the question is whether such a “3D-system“ is advantageous compared to a conventional “Mono-system“. To draw such a comparison one can ask users for their preference (Quality of Experience) — or use other measures. One might ask whether the cognitive load of interlocutors is lower and the communication consequently is more efficient and effective.

This thesis makes two contributions: First, the meaning of cognitive load is elucidated and means of measurement are presented. Second, a listening test is conducted to show possible advantages. In this test participants listen to conference recordings and subsequently answer questions about the topic and about their percieved listening effort.

Results show significant improvements over the mono-system. Participants have to concentrate less and are able to identify the current speaker more easily. Afterwards they also have a better picture of who said what.

1 Einleitung

Unsere Geschäftswelt ist globalisiert. Viele Menschen arbeiten heute zusammen, ohne im eigentlichen Sinn *zusammen* zu arbeiten. Sie arbeiten an verschiedenen Orten, sollen aber trotzdem so miteinander kommunizieren als wären sie im gleichen Raum: Effizient und effektiv. Sie planen gemeinsam Projekte, führen Verhandlungen und lösen Probleme, und das obwohl sie manchmal durch Ozeane getrennt sind.

Ein großer Teil der Kommunikation verläuft schriftlich: durch Emails und Chats, in Foren und über Projektmanagement-Tools. Einiges lässt sich jedoch einfacher im Gespräch klären, am besten mit allen beteiligten Personen und am besten an einem Ort. Oft ist das nicht machbar, dann behilft man sich mit Telefonkonferenzen, oder kürzer: *Telekonferenzen*.

Telekonferenzsysteme ermöglichen Gruppen ein Gespräch über weite Strecken. Leider ist das nicht dasselbe als würde man sich wirklich gegenüber sitzen:

> „Universally, in all our fieldwork, people complained about the quality of communication over audio and video conferences. Participants in audio conferences had *trouble figuring out who is talking* or what is being referred to.“
>
> (**?**, S.152, Meine Hervorh.)

Gründe für diese Probleme gibt es viele. Betrachten wir den Fall einer Telekonferenz mit gewöhnlichen Telefonen. Man kann die anderen Teilnehmer nur *hören*, aber nicht sehen. Wichtige nichtsprachliche Elemente wie Mimik und Gestik gehen damit schon einmal verloren. Man kann nicht sehen wie es jemandem unter den Nägeln brennt etwas zu sagen oder wie jemand seine Aussagen durch einen strengen Blick an eine bestimmte Person richtet. Teilnehmer, die länger nicht sprechen, können sogar aus der Wahrnehmung verschwinden.

Aber auch das, was man hört, ist nicht vergleichbar mit einem „echten“ Gespräch. Ein Unterschied ist, dass uns Telefone nur ein Mono-Signal bieten, unsere beiden Ohren in der Realität allerdings zu viel mehr fähig sind: Unser Hörsystem kann Schallquellen nämlich räumlich wahrnehmen und auf diese Weise trennen. So können wir uns in einer belebten Tischrunde einfach mit einzelnen Personen unterhalten, obwohl gleichzeitig auch andere reden (**Kapitel 2**). Dieses Verhalten, das bei einer angeregten Diskussion am Tisch normal ist, würde in einer Telekonferenz ziemlich schnell in ein Chaos ausarten. Zu diesen prinzipiellen Problemen kommen dann noch Verschlechterungen und Verzögerungen durch die Datenübertragung.

Natürlich gibt es technische Ansätze und Lösungen, um Telekonferenzen einem echten Treffen ebenbürtig zu machen. Jeder kennt wahrscheinlich Videotelefonie, die nicht

zuletzt durch Skype[1] für jeden verfügbar geworden ist. Allerdings wird auch an der Audio-Wiedergabe geforscht. Ein Ansatz ist die *3D-Telekonferenz* (eng.: *spatial audio conferencing / 3D telephony*). Hier stelle man sich einen Konferenzteilnehmer mit Kopfhörer vor, der die anderen Sprecher 3D hört, also räumlich verteilt wahrnimmt. Für ihn kann es sich so anhören als würden die anderen in einem Kreis um ihn herum sitzen, oder als wären alle vor ihm aufgereiht. Man nennt diese Art der Wiedergabe *binaural*.

An einem solchen System zur 3D-Telekonferenz wird auch am Lehrstuhl für Datenverarbeitung (LDV) der Technischen Universität München gearbeitet (**Kapitel 3**). Das LDV-System zielt dabei auf eine bestimmte Konstellation bei Telekonferenzen, bei der ein Teil der Teilnehmer gemeinsam am Tisch sitzt und weitere einzelne Teilnehmer an anderen Orten einbezogen werden sollen. Die abwesenden Teilnehmer sind in dieser Konstellation erst einmal benachteiligt. Das LDV-System gibt ihnen nun durch binaurale Wiedergabe zumindest teilweise das Gefühl vor Ort zu sein. Um dies zu ermöglichen ist spezielle Hardware zur Aufnahme und komplexe Software zu Sprechertrennung und Wiedergabe nötig.

Es stellt sich die Frage, ob ein solches 3D-System „besser“ ist und inwiefern sich ein solcher Unterschied äußert. Eine Antwort besteht darin, das System von verschiedenen Benutzern testen zu lassen und sie nach ihren Präferenzen zu fragen. Solche Tests geben dann Aufschluss über die sogenannte „Quality of Experience“.

Zusätzlich zu diesen subjektiven Präferenzen können allerdings auch noch andere Maßstäbe angelegt werden. So kann man danach fragen, ob die Kommunikation *effektiver* und *effizienter* abläuft. Diese beiden Merkmale können dem Begriff der *psychischen Belastung* untergeordnet werden (**Kapitel 4**). Der Begriff psychische Belastung (en. *mental workload / cognitive load*) ist vielfältig definiert, bezieht sich aber grob gesagt darauf, dass wir Menschen nur begrenzte Kapazitäten zur Informationsverarbeitung besitzen.

Es existieren verschiedene Ansätze um psychische Belastung zu erfassen (**Kapitel 5**). Diese reichen von einfachen Fragen („Wie sehr musstest du dich anstrengen?“) bis hin zu Messungen von Körperfunktionen, wie der z. B. Herzfrequenzvariabilität.

Eine Telekonferenz mit mehreren Teilnehmern kann unsere informatorischen Kapazitäten nun mehr oder weniger ausreizen. Der Kommunikationsvorgang kann damit mehr oder weniger anstrengend und effektiv für die Teilnehmer sein. Ziel dieser Arbeit ist es, herauszufinden ob die psychische Belastung bei dem 3D-Telekonferenz-System des LDV geringer ist als bei einem klassischen System. Dazu führe ich einen vergleichenden Hörversuch (**Kapitel 6**) durch, bei dem die Teilnehmer Konferenzaufnahmen hören und im Anschluss Inhaltsfragen beantworten.

Die Ergebnisse zeige ich in **Kapitel 7** und versuche in **Kapitel 8** ein Fazit aus diesen zu ziehen.

[1] `http://www.skype.com/de/`

2 Grundlagen der auditiven Wahrnehmung

Durch Hören nehmen wir unsere Umwelt war. Wir werden vor möglichen Gefahren gewarnt (Knistern im Wald) oder auf Dinge hingewiesen (schreiendes Baby). Hören dient der Objekterkennung (herunterfallendes Glas) und der Orientierung im Raum. Das Hören von Musik ermöglicht uns intensive emotionale Erlebnisse. Nicht zuletzt sind Hören und Sprechen Mittel für unsere alltägliche Kommunikation und damit Grundlage für soziale Kontakte (**?**, p. 123).

Dieses Kapitel beschreibt die Grundlagen der auditiven Wahrnehmung. Nach einer kurzen Einführung in die Funktionsweise des Ohrs und der neuralen Verarbeitung der Audiosignale (**2.1**), beschreibe ich, wie wir mit unseren beiden Ohren Richtungen, bzw. „Räumlichkeit“ wahrnehmen (**2.2**). Danach fokussiere ich mich auf psychologische Aspekte der auditiven Szenenanalyse und der selektiven Aufmerksamkeit, insbesondere im Zusammenhang mit dem sogenannten Cocktailparty-Effekt (**2.3**).

2.1 Der Hörsinn

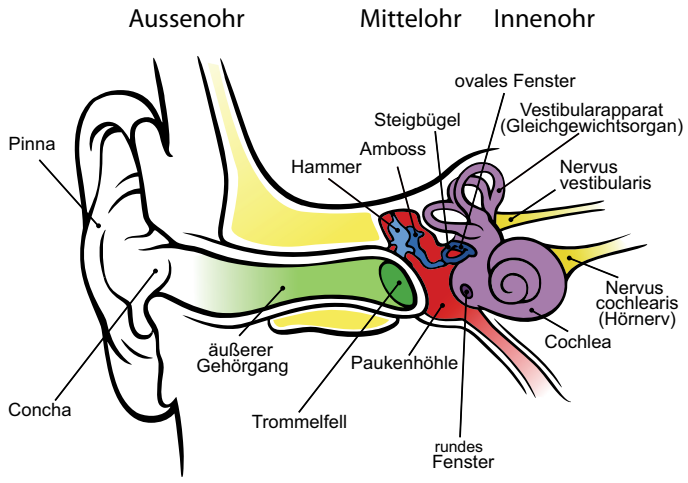


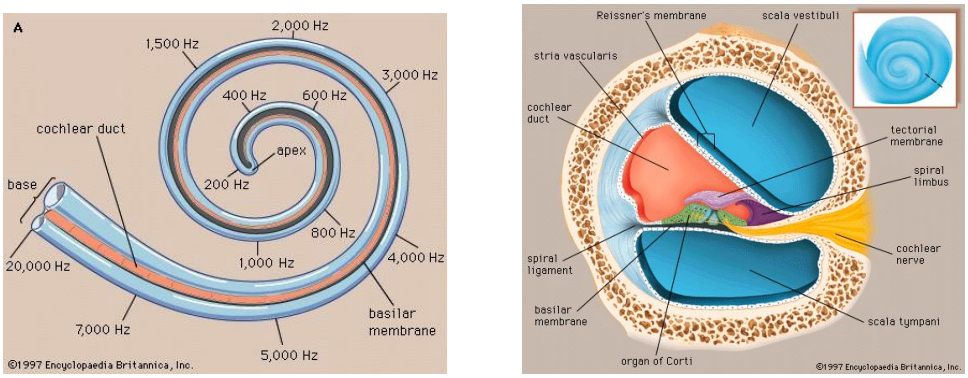
Abbildung 2.1: Anatomie des menschlichen Ohrs (Nach: **?**)

?, Ch. 13 bietet einen guten Überblick über die Funktionsweise des Ohrs, die ich hier aus Sicht des Schalls bzw. des auditorischen Signals beschreiben möchte.

Das menschliche Ohr ist anatomisch unterteilt in Außen-, Mittel- und Innenohr (Abb. 2.1). Ankommende Schallwellen treffen zuerst auf die *Ohrmuschel*, wandern dann durch den *äußeren Gehörgang* und treffen auf das *Trommelfell*. Auf diesem Weg verändert sich die Schallinformation bereits, da die Ohrmuschel wie ein richtungsabhängiges Filter funktioniert. Ausserdem verstärkt der der Gehörgang Frequenzen um 3 kHz[1].

Am Trommelfell beginnt das *Mittelohr*. Dieses erfüllt die Aufgabe die Schallwellen vom Medium Luft in das flüssige Medium des Innenohrs zu übertragen. An einem solchen Übergang zu einem Medium mit einer höheren akustischen Impedanz würde normalerweise der meiste Schall reflektiert werden. Allerdings sorgt ein System aus Knöchelchen (*Hammer, Amboss und Steigbügel*) mithilfe von Hebelwirkung dafür, dass die am Trommelfell eintreffende Energie weitgehend auf das Innenohr übertragen werden kann[2].

Das *Innenohr*, genauer gesagt die *Cochlea*, wandelt nun die Information der Schalldruckwellen in neuronale Signale um. Die Cochlea hat die Form einer Schnecke mit zwei parallelen flüssigkeitsgefüllten Röhren, die durch die *Basilarmembran* getrennt sind (Abb. 2.2). Die Druckwellen wandern vom ovalen Fenster am Ansatz bis zum eingerollten Ende und dann durch die zweite Röhre wieder zurück. Dabei wird die Basilarmembran zu Schwingungen angeregt. Diese Schwingungen werden von Haarzellen des *Cortischen Organs* erfasst und in neuronale Signale umgewandelt. Verschiedene Abschnitte entlang der Cochlea werden hierbei verschieden stark von verschiedenen Frequenzen angeregt, vergleichbar mit einer Frequenzweiche mit einzelnen Bandpässen. Diese *tonotopische Organisation* basiert sowohl auf passiven Resonanzeigenschaften der Cochlea, als auch auf aktiven Feedback-Vorgängen.



(a) **(b)**

Abbildung 2.2: (a) Tonotopische Aufteilung der Cochlea (schematisch) und (b) im Querschnitt (Quelle: **?**)

[1] In diesem Frequenzbereich befinden sich die, für das Sprachverständnis wichtigen, Plosivlaute.

[2] Die Gehörknöchelchen sind darüber hinaus mit Muskeln verbunden, die z. B. bei lauten Geräuschen unterbewusst die Übertragung dämpfen.

Die akustische Information verlässt das Ohr über den Hörnerv (*Nervus cochlearis*). Dessen Anzahl an Fasern entspricht ungefähr der Menge an Haarzellen in der Cochlea. Die tonotopische Organisation der Cochlea setzt sich hier also fort und jede einzelne Faser lässt sich vereinfacht einer *charakteristischen Frequenz* zuordnen.

Im weiteren Verlauf durchläuft die akustische Information nun einige Punkte im Stammhirn und mündet schließlich in den auditiven Kortex des Großhirns, wo sie weiteren kognitiven Prozessen, wie dem Sprachverständnis, zur Verfügung steht. Allerdings werden bereits im Stammhirn die Informationen beider Ohren kombiniert und Aspekte der lateralen Richtungswahrnehmung bestimmt.

2.2 Räumliches Hören

Wenn wir die Augen schließen, hören wir nicht nur *was* um uns herum passiert, sondern auch *wo*. Wir haben ein Gefühl für die Quelle des Schalls und die Beschaffenheit des Raums.

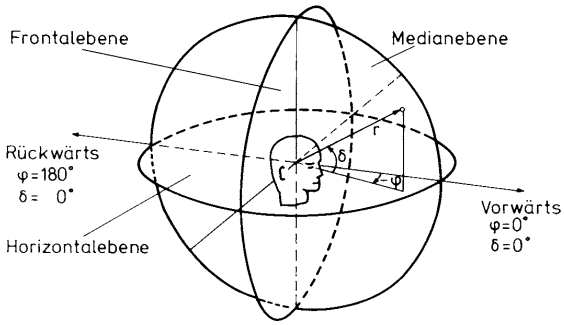


Abbildung 2.3: Kopfbezogenes Polarkoordinatensystem (Aus: **?**, *s. 88*)

Die Position einer Schallquelle wird beschrieben mit der Entfernung r und zwei Winkeln in einem polaren Koordinatensystem um den Kopf (Abb. 2.3): einem Links-Rechts-Anteil, dem *Azimut* ϕ und einem Oben-Unten-Anteil, der *Elevation* δ.

Die räumliche Wahrnehmung unserer akustischen Umgebung, also die Anzahl, Richtung und Entfernung verschiedener Schallquellen und die Beschaffenheit des Raums, ist von verschiedenen Aspekten geprägt. Ich möchte anhand **?**, Kapitel 3 und **?** einen Überblick über die Komplexität dieser Wahrnehmung geben:

Aspekte Räumlichen Hörens

Außenohr-Übertragungsfunktion (HRTF): Je nachdem aus welcher Richtung eine Schallwelle kommt legt sie am Kopf entlang, auf dem Weg zu den beiden Ohreingängen, verschiedene Wege zurück. Daraus ergeben sich zwischen linkem und rechtem Ohr naheliegenderweise *Laufzeit- und Pegelunterschiede* (Schall von links erreicht das linke Ohr zuerst und ist dort lauter als rechts).

Für sich genommen (Duplex Theorie, (**?**)) enthielten diese beiden Informationen nur Hinweise auf den Azimut einer Schallquelle, nicht aber dessen Elevation. Bei der Lokalisation würden sich also Mehrdeutigkeiten ergeben. Allerdings erzeugt das Außenohr-Kopf-System auch noch richtungsabhängige akustische Verzerrungen. Besonders das verschlungene und gerichtete Außenohr verstärkt oder dämpft verschiedene Frequenzen je nach Richtung unterschiedlich.

Aus Sicht der Signaltheorie bildet der Kopf hier also ein richtungsabhängiges, lineares und zeitinvariantes Filter-Paar, das mathematisch durch die *Außenohr-Übertragungsfunktion* (en.: *Head Related Transfer Function*) beschrieben wird. Für jede mögliche Position einer Schallquelle existiert also ein spezifisches Funktionenpaar $H_{l/r}(f, r, \phi, \delta)$, das diese frequenzabhängigen Veränderungen darstellt.[3] Abbildung 2.4 zeigt beispielhaft die Übertragungsfunktionen beider Ohren für eine Raumposition.

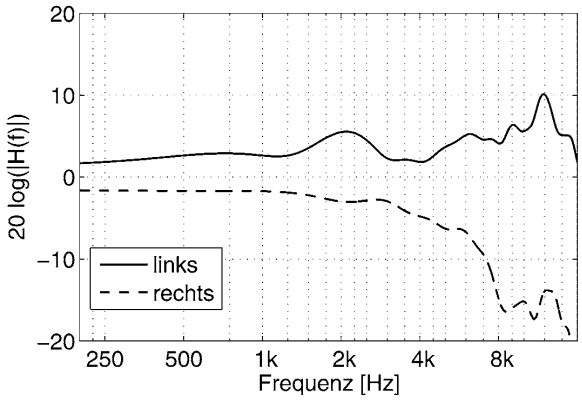


Abbildung 2.4: Beispiel für ein HRTF-Paar bei $90°$ Azimut und $0°$ Elevation, also senkrecht von links (Aus: **?**, *s. 91*)

Kopfbewegung: Durch Drehen des Kopfes verändern sich die Ohrsignale, so dass mehr Informationen zur Verfügung stehen, welche eine Lokalisation erleichtern (**?**).

[3] Es ist möglich die Entfernung r hier zu vernachlässigen und später durch eine einfache Pegeländerung zu berücksichtigen.

Akustische Eigenschaften des Raums: An das Ohr kommen nicht nur Schallwellen über den direkten Weg von der Quelle (Direktschall), sondern auch Anteile, die zuvor z. B. von Wänden reflektiert wurden (Raumschall). Was wir hören ist also immer auch geprägt von der Beschaffenheit des uns umgebenden Raums. Schallsituationen ohne Reflexionen (Freifelder) existieren in der Realität nur in besonderen Situationen und auch nur angenähert, so z. B. auf einer schneebedeckten Bergspitze oder in speziellen reflexionsarmen Kammern.

Visuelle Informationen: Unsere Hörwahrnehmung ist *multimodal* und wird besonders durch visuelle Reize mitgestaltet. Als Beispiel sei hier der Bauchredner-Effekt (**?**) genannt, also die Illusion eine Puppe spräche und nicht der *augen*-scheinlich stumme Bauchredner.

Wahrgenommene Entfernung: Die Entfernung, in der wir eine Schallquelle vermuten hängt natürlich vom Pegel des Schalls ab, der uns erreicht. Allerdings spielen ebenso die Reflexionseigenschaften des umgebenden Raums eine Rolle (**?**), oder unsere Erwartungshaltung bei bestimmten Geräuschen. Beispielsweise hängt die wahrgenommene Entfernung von Sprechern davon ab ob sie schreien oder flüstern. Bei ersterem wird sie über-, bei letzterem unterschätzt (**?**).

Künstliches räumliches Hören

Es gibt verschiedene Möglichkeiten um künstlich einen räumlichen Eindruck zu erzeugen. Eine triviale Möglichkeit ist das *Stereo Panning* bei dem man den zwei Stereo-Kanälen einfach unterschiedliche Pegel zuweist. Im Allgemeinen arbeitet man aber mit Wiedergabe über Kopfhörern und verwendet HRTF-Daten. Hierbei faltet man die Audiodaten mit der Impulsantwort, die zu einer beliebigen Koordinate gehört [4]. Diese Daten stammen meist von Messaufnahmen an einem speziellen Dummy mit Durchschnittsmaßen, wie in Abb. 2.5 dargestellt. Man kann die Übertragungsfunktionen aber auch in einem aufwändigeren Prozess an Personen direkt messen.

2.3 Der Cocktailparty-Effekt

Auditive Szenenanalyse

Normalerweise hören wir nicht nur *ein* Schallereignis, sondern mehrere gleichzeitig. Es ist jedoch für uns kein Problem den gleichzeitig einfallenden Schall sinnvoll aufzuteilen und zu interpretieren. Grundlage dafür ist eine Art kognitive Signaltrennung, die *Auditive Szenenanalyse* (**?**) genannt wird. Der Mensch trennt, bzw. gruppiert auditive Reize anhand

[4] Die Impulsantwort im Zeitbereich (*Head Related Impulse Response*) ist äquivalent zur HRTF im Frequenzbereich

Abbildung 2.5: *„KEMAR“*-Puppe zur Aufnahme von Durchschnitts-HRTFs

bestimmter Parameter, wie dem zeitlichen Abstand des Einsetzens, Frequenzunterschieden oder der räumlichen Herkunft.[5]

Der Cocktailparty-Effekt

Im Zusammenhang mit Telekonferenzen interessieren uns vor allem Situationen in denen die Schallquellen *Sprecher* sind. Stellen Sie sich also vor Sie befinden sich auf einer Feier. Es ist laut, die Leute um Sie herum unterhalten sich angeregt und es läuft Musik. Ungeachtet dieses Wirrwarrs an unterschiedlichen Stimmen und anderen Geräuschen können Sie sich trotzdem mit einer bestimmten Person unterhalten. Sie können alle anderen Gespräche ausblenden, auch wenn diese gleich laut sind, und sich auf Ihre Konversation konzentrieren. Sobald allerdings in der Nähe Ihr Name fällt, bemerken Sie das sofort und lauschen kurz nach dem Zusammenhang, um dann wieder zu Ihrem eigentlichen Gespräch zurückzukehren. Dieses Phänomen der selektiven auditiven Wahrnehmung von Sprache nennt man passenderweise *Cocktailparty-Effekt (CPE)*.

Experimente zum CPE

Experimente zum gleichzeitigen Verstehen verschiedener Sprecher wurden schon früh durchgeführt, wie z. B. bei dichotischen Hörtests von **?** und **?**: In einem Experiment hörten Testpersonen gleichzeitig auf beiden Ohren jeweils einen anderen Sprecher. Sie sollten sich dabei auf *ein* Ohr konzentrieren und das, was sie hörten mitsprechen. Im Ergebnis konnten sie sich auch nur noch an den Inhalt des fokussierten Ohrs erinnern, nicht aber an den Inhalt des ignorierten Ohrs. So bemerkten Testpersonen auch nicht, wenn auf dem ignorierten Ohr plötzlich Sprache rückwärts abgespielt wurde. Es fiel ihnen allerdings auf

[5] Ein grundlegendes Experiment von **?** kann man in ähnlicher Form auf `http://auditoryneuroscience.com/?q=topics/streaming-galloping-rhythm-paradigm` selbst ausprobieren.

wenn das Geschlecht wechselte oder plötzlich nur noch ein Sinuston zu hören war, oder aber wenn ihr eigener Name genannt wurde.

Der CPE ist vor allem ein binaurales Phänomen: In einem Experiment von **?** hörten die Versuchspersonen über drei Lautsprecher verschiedene Sprecher, die ihnen einfache Aufgaben stellten. Es wurden immer zwei Anfragen gleichzeitig wiedergegeben, wovon aber immer nur eine relevant für die VP war. Spieth stellte u.A. fest, dass sich die Leistung verbesserte wenn die Lautsprecher verteilt aufgestellt wurden ($-90°$, $0°$, $+90°$). Auch bei Raumsynthese über HRTFs (**?**) konnte ein besseres Verständnis als bei monauraler Wiedergabe nachgewiesen werden.

Es ist nicht überraschend, dass der CPE auch von visuellen Informationen beeinflusst wird. Treten Konflikte zwischen der visuellen und der akustischen Information auf, kann dies zu trügerischen Wahrnehmungen führen: Neben dem bereits genannten „Bauchredner-Effekt“ sei hier auch noch ein Experiment von **?** genannt. Dabei sahen die Versuchsteilnehmer ein Video einer Frau, die monoton Silben wiederholte. Sie *sahen* die Silbe `ga` (Lippenbewegung), *hörten* `ba` (Tonspur) und verstanden aber `da`. Fiel der Bild- oder Tonanteil weg verstanden sie das meiste korrekt. Wie in der Einleitung angesprochen, spielen beim Kommunikationsvorgang auch noch weitere Komponenten wie die Gestik, Mimik und Blickrichtung eine Rolle.

Theorien zum CPE

Es existieren verschiedene Theorien zum Cocktailparty-Effekt, beziehungsweise zur *selektiven Aufmerksamkeit* des Menschen allgemein (**?**). Drei bekannte Ansätze sind die:

- *Filtertheorie* von **?**
- *Attenuationstheorie* von **?**
- *Theorie der späten Selektion* von **?**

In der *Filtertheorie* (Abb. 2.6a) funktioniert Aufmerksamkeit nach einem Alles-oder-Nichts-Prinzip. Dabei wird bereits zu einem frühen Zeitpunkt in der Verarbeitungskette nur *ein* Informationsstrom durchgelassen, also z.B. nur die Informationen eines Ohres. Einen Schwachpunkt dieses Modells zeigt allerdings das vorhin angesprochene Experiment, bei dem Menschen ihren Namen am ignorierten Ohr hörten. Dies wäre im Rahmen der Theorie nur durch schnelles hin und her wechseln zu erklären.

Deshalb geht die *Attenuationstheorie* (Abb. 2.6b) davon aus, dass höheren kognitiven Prozessen alle Ströme gleichzeitig zur Verfügung stehen, diese aber davor verschieden gewichtet werden. Man kann, so gesehen, Informationen aus einem Ohr nicht komplett ausblenden, sondern nur abschwächen. Taucht dann ein Element mit niedrigem Aktivierungsniveau auf — wie der eigene Name — nimmt man diesen wahr.

Der Vollständigkeit halber sei auch noch die *Theorie der späten Selektion* genannt bei der *alle* Eingangsreize vollständig analysiert werden und zwar erst spät, auf Ebene der Handlungssteuerung (Abb. 2.6 c).

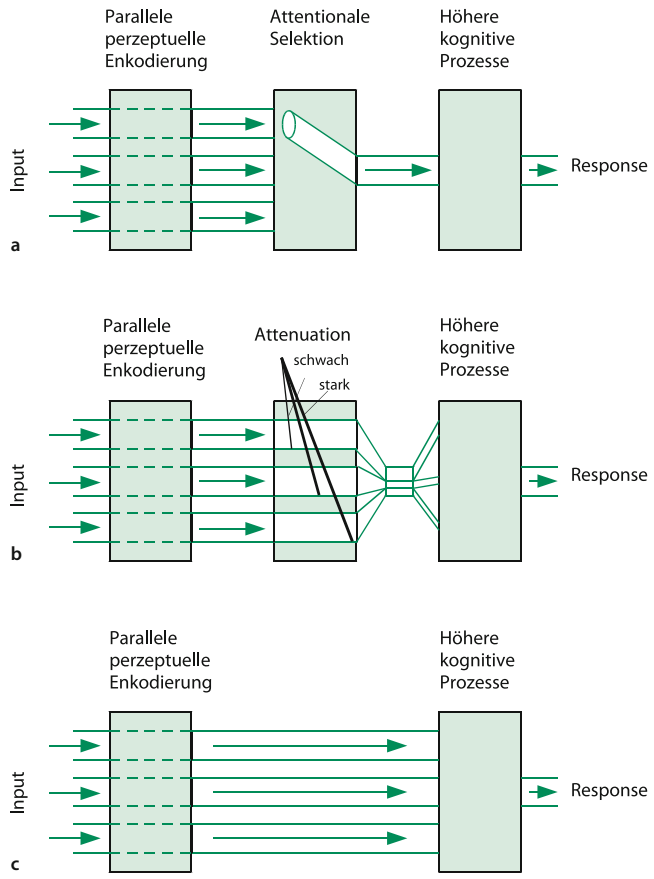


Abbildung 2.6: Modelle der selektiven Aufmerksamkeit: a) Filtertheorie b) Attenuationstheorie c) Theorie der späten Selektion (Aus **?**, *S. 182*)

Parameter des CPE

Für die Anwendung in einem Telekonferenz-System interessieren uns konkret die Rahmenbedingungen, die den CPE begünstigen und wie man diese technisch umsetzen kann. Ziel ist, dass der Teilnehmer seine natürliche Fähigkeit, mehrere Sprecher zu verarbeiten,

optimal nutzen kann. **?** fasst einige Möglichkeiten für sprachliche Kommunikationssysteme zusammen (eigene Reihenfolge):

1. „Provide spatial continuity within channels“
2. „Provide spatial disparity between channels“
3. „Associate visual images with audio streams“
4. „Provide a mechanism to “pull” one voice into focus“
5. „Provide F0 continuity“
6. „Micromodulate F0 to enhance voices“
7. „Filter streams into separate frequency bands“
8. „Pitch shift voices away from each other“
9. „Use different voices (synthetic or recorded)“
10. „Provide enough time for the user to fully fuse streams“
11. „Do not present too much information simultaneously“

Wie kann man diese Faktoren nun auf eine Telefonkonferenz anwenden? Auf den Ablauf der Kommunikation (9.-11.) lässt sich technisch kein Einfluss nehmen. Eigenschaften der Stimme (5. - 8.) könnten zwar technisch verändert werden sind jedoch auch charakteristisch für die einzelnen Sprecher. Punkt 3 weist in Richtung Videokonferenz. Punkt 4 wäre denkbar, erfordert jedoch einen zusätzliche Mehraufwand vom Teilnehmer, da er immer wieder einen Kanal zur Fokussierung auswählen müsste. Die Punkte 1 und 2 sind über HRTF-Synthese implementierbar und Teil des Telekonferenzsystems am LDV.

3 Das Telekonferenzsystem am LDV

In diesem Kapitel beschreibe ich das Telekonferenzsystem des LDV, genauer gesagt dessen aktuellen Forschungsstand. Abbildung 3.2 zeigt das Gesamtsystem an einem angestrebten Anwendungsfall: *Adam*, *Berta* und *Carl* sitzen in einer lockeren Runde an einem Tisch und führen ein Gespräch. In der Tischmitte steht ein akustisches Aufnahmegerät, die Konferenzspinne. An einem entfernten Ort befindet sich *Walter* und verfolgt die Konferenz über Kopfhörer. Dabei hört er die anderen Teilnehmer räumlich getrennt, und zwar jeden Sprecher an einer anderen, frei wählbaren Raumposition.

Zwischen Konferenztisch und *Walters* Ohr liegen die drei Komponenten des LDV-Systems:

1. Aufnahme Die am LDV entwickelte Konferenzspinne erfasst das Gespräch der Teilnehmer in der näheren Umgebung. Sie besteht aus acht Mikrofonen, die gleichmäßig im Kreis angeordnet sind (Abb. 3.1, Details in **?**). Zusätzlich zu den Audiosignalen wird noch die Temperatur erfasst und an einen Rechner zur Aufzeichnung weitergeleitet.

2. Sprechertrennung/-erkennung Die Audiokanäle werden dann von einem Algorithmus in einen Kanal pro Sprecher aufgeteilt. Dies erfordert zusätzlich kurze Trainingsdaten der einzelnen Sprecher und die aktuelle Raumtemperatur. Der Algorithmus ist mehrstufig aufgebaut und erfüllt drei zusammenhängende Funktionen:

1. *Lokalisierung* der verschiedenen Sprecher im Raum
2. *Separierung* der einzelnen Audio-Spuren
3. *Sprechererkennung*, also Zuordnung der Audio-Spuren zu den trainierten Sprechermodellen

Der komplexe Algorithmus und seine Entwicklung ist dokumentiert in **?**, aufbauend auf Arbeiten von **?**, **?** und **?**.

Im Ergebnis produziert der Algorithmus einen Kanal pro Sprecher. Diese Daten würden in einer realen Anwendung nun über ein Netzwerk übertragen werden. Entsprechende Details finden sich in **?**.

3. Wiedergabe Die einzelnen Kanäle werden über Stereo-Kopfhörer binaural wiedergegeben. Die verschiedenen Sprecher werden also durch HRTFs auf unterschiedliche räumliche Positionen gesetzt. Hierbei ergeben sich Alternativen aus der Wahl der HRTF:

- Personenspezifisch

- Aus anthropometrischen Merkmalen errechnet
- Aus einer Datenbank ausgewählt
- Durchschnitts-HRTF („KEMAR“)

Bei der Wiedergabe kann zusätzlich die Kopfbewegung durch ein Tracking-System erfasst werden. Wenn der Nutzer dann seinen Kopf dreht, kann die relative räumliche Platzierung der Quellen angepasst werden, so dass die Positionen der Sprecher, aus Sicht des Benutzers, gleich wahrgenommen werden. Hierzu wurde eine neuartige Möglichkeit des Headtrackings am LDV entwickelt, die eine handelsübliche Webcam zur Erfassung der Kopfposition verwendet (**?**).

Es sei angemerkt, dass der *aktuelle* Entwicklungsstand noch keinen Online-Betrieb, wie im vorherigen Szenario, zulässt. Ein Grund ist, dass der Algorithmus zur Sprechertrennung noch nicht echtzeitfähig ist. Das Zusammenspiel der Komponenten kann zur Zeit also nur simuliert werden.

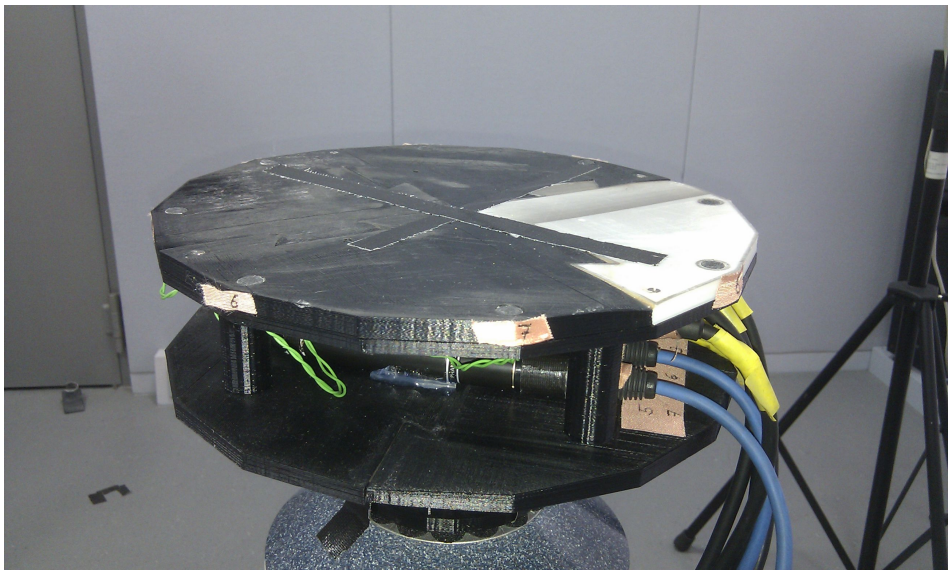


Abbildung 3.1: Am LDV entwickelte Konferenzspinne mit acht Mikrofonen

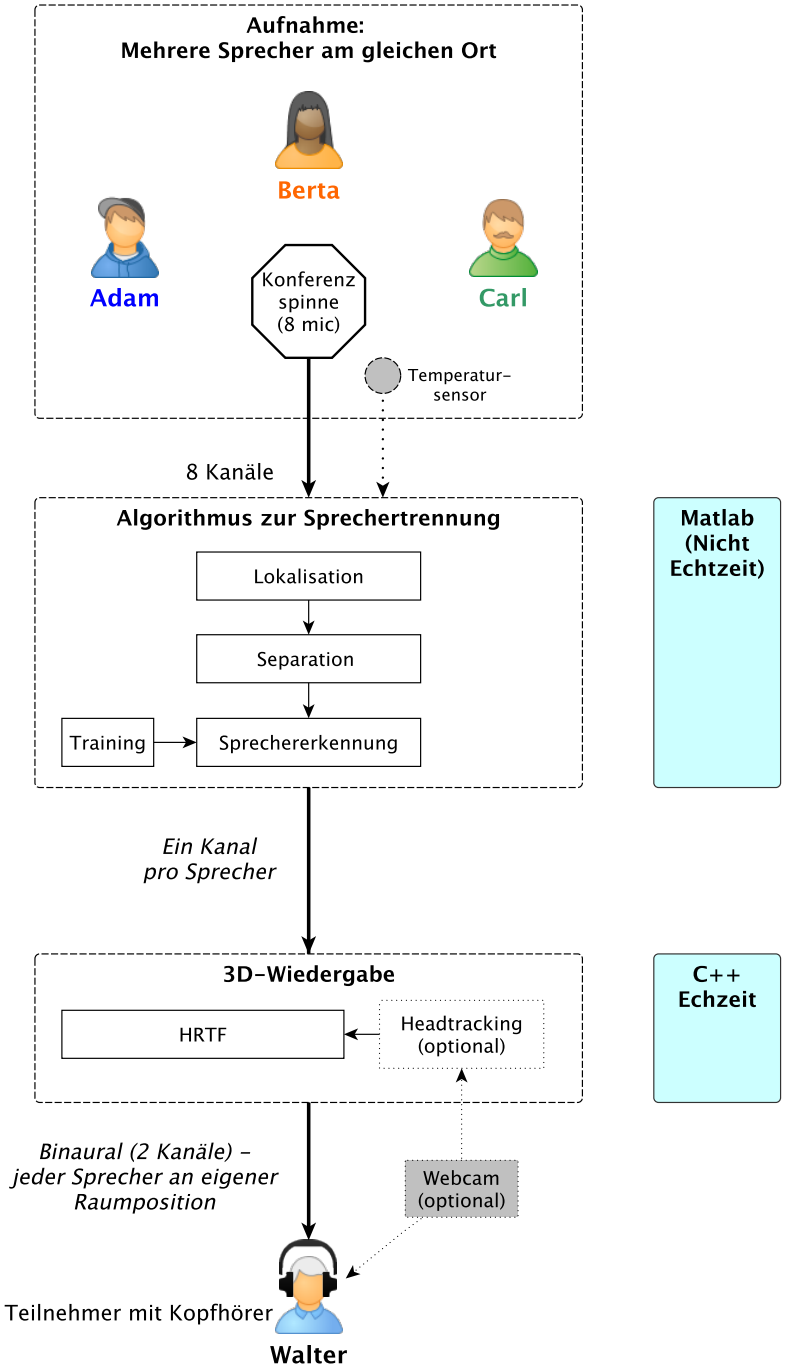


Abbildung 3.2: Das LDV-Telekonferenzsystem im Überblick. Rechts dargestellt, der aktuelle Implementierungsstand der Software

4 Psychische Belastung

In diesem Kapitel zeige ich, was psychische Belastung ist und inwiefern sie einen Ansatzpunkt zur Evaluation darstellt. Zu Beginn möchte ich aufzeigen, wie der Begriff mit den Konzepten *Effektivität* und *Effizienz* und anderen Evaluationsmöglichkeiten zusammenhängt (**4.1**). Danach beschreibe ich verschiedene Definitionen von psychischer Belastung (**4.2**).

4.1 Überblick und Einordnung

Bezugspunkt unserer Evaluation soll der Mensch sein, also der Nutzer für den ein System letztendlich entwickelt wird. In diesem Zusammenhang spricht man oft von der *Qualität*, deren Definition allerdings nicht immer eindeutig ist und einem zeitlichen Wandel unterliegt. Ich orientiere mich an der Definition aus **?**, S. 6:

> *Quality*: „Is the outcome of an individual's comparison and judgment process. It includes perception, reflection about the perception, and the description of the outcome. [...]“

Für den konkreten Fall der Interaktion mit einem technischen System ist weiter die *Quality of Experience* definiert:

> *Quality of Experience (QoE)* „is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state.“

Die QoE technischer Systeme betrachtet also weniger technische Aspekte, als die Qualitätswahrnehmung des Nutzers. Genauer gesagt setzt man beim Nutzer voraus, dass dieser einen Prozess der Wahrnehmung, des Vergleichs mit seiner Erwartung und einer abschließenden Beurteilung vornimmt (**?** und **?**).

Quality of Experience ist berechtigterweise ein wichtiger Ansatzpunkt zur Evaluation eines Telekommunikationssystems. Neben der Qualität existiert allerdings auch noch der Begriff der *Usability* (Gebrauchstauglichkeit), der im ISO Standard 9241 Teil 11 definiert ist:

> *Usability*: „The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. “ (aus **?**, S. 8)

Usability wird also in drei Teilbereiche aufgeschlüsselt:

> *Effectiveness*: „The accuracy and completeness with which specified users can achieve specified goals in specified environments.“
>
> *Efficiency*: „The resources expended in relation to the accuracy and completeness of the goals achieved.“
>
> *Satisfaction*: „The comfort and acceptability of the system to its users and other people affected by its use.“

Es fällt auf, dass sich die Konzepte der Usability und QoE nicht klar voneinander trennen lassen. Sie überlappen sich insbesondere beim Begriff Satisfaction. Eine gute Usability äussert sich letztendlich in einer guten QoE.

Allerdings setzt die QoE an einem anderen Punkt an, da sie nach dem Ergebnis eines Beurteilungsprozesses fragt. Prozesse, die ein Nutzer nicht bewusst bewertet, werden also gar nicht oder nur indirekt erfasst. Ein Beispiel ist die Effektivität einer Kommunikation. Diese könnte sich beispielsweise darin äußern, dass sich ein Teilnehmer besser oder schlechter an den Inhalt einer Telekonferenz erinnern kann. Ein anderes Beispiel ist die Effektivität, die nach den aufgewendeten Ressourcen fragt. Bei einer Telekonferenz muss sich ein Benutzer konzentrieren um mehreren Sprechern zu folgen. Wie stark er sich dabei konzentrieren muss oder wie stark er ermüdet, kann man als kognitive Ressource betrachten.

Diese Arbeit versucht ausgehend von den Konzepten *Effektivität* und *Effizienz* eine Bewertung des 3D-Telekonferenzsystems durchzuführen. Dafür fasse ich diese beiden Konzepte unter dem Gerüst der psychischen Belastung zusammen.

In einer parallelen Arbeit nimmt Thomas Volk die Sichtweise der Quality of Experience ein. Wir sehen unsere Arbeiten nicht strikt getrennt, sondern als zwei verschiedene Ausgangspunkte, um letztendlich ein fundiertes Urteil treffen zu können.

4.2 Der Begriff psychische Belastung

Als psychische Belastung bezeichnet man eine *informatorische* bzw. *kognitive* Beanspruchung, die Auftritt wenn eine bestimmte Aufgabe unter bestimmten Rahmenbedingungen zu verrichten ist. Diese kann, muss aber nicht mit ausgeprägten Emotionen einhergehen. Die englischen Begriffe „mental workload“ und „cognitive load“ beschreiben den Sachverhalt warscheinlich nüchterner und treffender. In der Literatur findet man eine Unzahl weiterer Begriffe, die oft synonym verwandt werden, wie „cognitive effort“, „mental effort“ oder „mental strain“ „attentional effort“ oder spezifisch bei Höraufgaben „listening effort“ oder „ease of listening“.

Eine exakte, allgemein anerkannte und über die verschiedenen Anwendungsfelder gleich verwendete Definition psychischer Belastung existiert nicht. Im Zusammenhang mit Mehrpersonen-Telekonferenzen erscheint mir folgende Definition passend:

> „... *mental workload* may be viewed as the difference between the capacities of the information processing system that are required for task performance to satisfy performance expectations and the capacity available at any given time.“
>
> (Gopher and Donchin (1986), entn. aus **?**, S. 2, Meine Hervorh.)

Im Bereich der Audio-Übertragungssysteme ist die Frage nach der kognitiven Belastung, anders als beim Qualitätsempfinden, noch ein relativ junges Feld. So wurde z.B. die Auditory Cognitive Neuroscience Society erst 2009 gegründet (**?**). Vergleichbare Evaluationen finden sich am ehesten noch in der Audiologie, also bei Forschungen über das menschliche Gehör, Hörschäden und Hörgeräten (**??**).

Psychische Belastung kann man von zwei Seiten analysieren. Eine Seite fragt nach der Anwendung, also wo Belastung auftritt und wie man sie verringern kann. Die andere Seite fragt nach den theoretischen Hintergründen:

- Allgemein, im Sinne der Ergonomie bzw. Arbeitswissenschaft
- Spezifisch, im Sinne der *Cognitive Load Theory*

4.2.1 Arbeitswissenschaften

Eine Telekonferenz kann als *Arbeitstätigkeit*, ein TK-System als *Arbeitsmittel* gesehen werden. Es sollte damit auch den Anforderungen der *Ergonomie*, bzw. der *Arbeitsphysiologie* gerecht werden. Über diese geben **?** und **?** einen Überblick.

> *Ergonomie* „ist die Lehre von der menschengerechten Gestaltung der Arbeit. Sie wird durch medizinisch-biologische und *psychologische Kriterien* des Ausgleichs zwischen menschengerechten sowie technisch bzw. organisatorisch realisierbaren Anforderungen und Belastungen der Arbeit geprägt.“
>
> (**?**, S. 481, Meine Hervorh.)

Ein ergonomisches Arbeitsmittel muss also nicht nur körperlich gut benutzbar sein, sondern soll den Menschen auch psychisch möglichst gering belasten. [1]

Auf übermäßige Belastung reagiert der Mensch allgemein mit einer schlechteren Leistung und verringerter Leistungsbereitschaft. Je nach Art der Überlastung und der Reaktion des Menschen unterscheidet man zwischen verschiedenen *Kompensationsaktivitäten* (Tabelle 4.1).

In einer TK-Situation sind je nach Inhalt des Gesprächs alle aufgezählten Kompensationsaktivitäten denkbar. Technisch beeinflussbar sind meiner Meinung nach insbesondere aber die Psychische Sättigung, Ermüdung und in geringerem Maße Stress. Letzterer

[1] Eigentlich müsste man hier von „Beanspruchung“ reden, da in der Literatur zwischen der „Belastung“ als äußere Einwirkung und der „Beanspruchung“ als subjektive Folge unterschieden wird. In dieser Arbeit verwende ich im Weiteren vereinfacht den Begriff Belastung

Kompensationsaktivität	*Erlebenssymptome*	*Beispielsituation*
Ermüdung	Müdigkeit, Abgespanntheit, Erschöpfung	Nach mehreren Stunden (erfolgreichen) Lernens
Monotonie	Schläfrigkeit, Langeweile, Müdigkeit	Langes Überwachen eines Radarschirms
Stress	Angst, Sorge, Unsicherheit	Beginn einer schwierigen Prüfung
Psychische Sättigung	Widerwillen, Gereiztheit	Level in einem PC-Spiel zu schwer

Tabelle 4.1: Kompensationsaktivitäten psychischer Belastung (angelehnt an **?**, S.446, mit eigenen Beispielen)

resultiert wahrscheinlich eher aus dem Gesprächsinhalt. Monotonie setzt eine gewisse Einförmigkeit und Reizarmut voraus, die in einem Gespräch wahrscheinlich nicht auftritt.

4.2.2 Cognitive Load Theory

Die *Cognitive Load Theory* ist ein theoretischer Ansatz zur Erklärung bestimmter Arten psychischer Belastung. Sie beschreibt ein Modell der menschlichen Informationsverarbeitung und zeigt auf, wie Lernvorgänge von bestimmten kognitiven Kapazitäten ermöglicht bzw. begrenzt werden (**?**, Ch. 2).

Ausgangspunkt der CLT ist eine begrenzte Kapazität des menschlichen Arbeitsgedächtnisses. Die Theorie leitet daraus Empfehlungen für die Gestaltung von Lernsituationen ab, mit dem Ziel das Arbeitsgedächtnis nicht zu überlasten. Dabei stützt sie sich auf ein kognitionspsychologisches Modell, wie der Mensch Informationen verarbeitet und lernt.

Modell der menschlichen Informationsverarbeitung

Den Prozess der menschlichen Informationsverarbeitung unterteilt man vereinfacht in drei Systeme (**?**, Ch. 1 und **?**, Ch. 31. Grafik 4.1 zeigt diese Gedächtnissysteme und ihren Zusammenhang[2].

Sinnesinformationen landen zuerst einmal in einem Kurzzeitpuffer, dem *Sensorischen Gedächtnis* mit einer sehr großen Kapazität. Sie können dort jedoch nur Sekundenbruchteile verweilen. Welche Informationen davon das Arbeitsgedächtnis erreichen, also bewusst *wahrgenommen* werden, bestimmt unsere *Aufmerksamkeit*. Dieser Vorgang ist leider sehr komplex und bei weitem nicht vollständig verstanden. Fest steht, dass der Mensch nicht alle Informationen im sensorischen Gedächtnis wahrnehmen kann, sondern sich selektiv auf bestimmte Informationen konzentrieren muss.

[2] Ich möchte dieses Thema hier nur anschneiden. Das Modell, wie ich es präsentiere, ist eine starke Vereinfachung des kompletten theoretischen Modells, und dieses wiederum eine Vereinfachung der Realität.

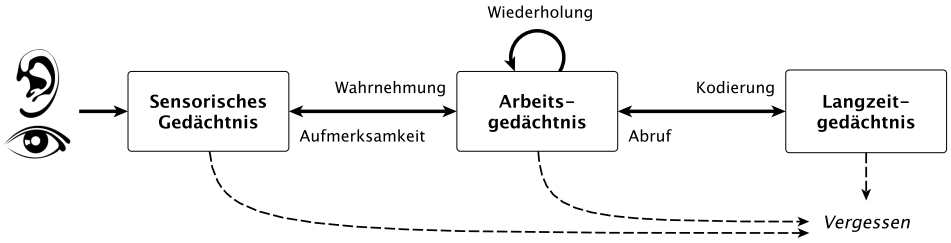


Abbildung 4.1: Modell der menschlichen Informationsverarbeitung (Nach **?**, *S. 6 und* **?**, *S. 792*)

Im *Arbeitsgedächtnis* kann nur eine begrenzte Menge an Informationen gleichzeitig gehalten werden – und dies auch nur für eine bestimmte Zeit. Ein klassisches Beispiel ist die Tatsache, dass man sich nur ca. sieben Ziffern einer Telefonnummer auf einen Schlag merken kann. Durch Gruppieren der Ziffern („089 34 24 44“) kann man aber auch längere Nummern erfassen. Allgemein kann also eine bestimmte Menge an Sinneinheiten gespeichert werden.

Für die Aufteilung des Arbeitsgedächtnisses selbst existiert ein weiteres Model von **?**. Demnach besteht es aus drei Teilen:

Zentrale Exekutive: Steuerung der Aufmerksamkeit

Phonologische Schleife: Zwischenspeicher für sprachliche und akustische Informationen

Räumlich, visueller Notizblock: Zwischenspeicher für visuelle und räumliche Informationen

Es existieren also getrennte Bereiche für sprachliche und räumliche Informationen. Wie **?** darlegt, kann dies – zumindest teilweise – als Erklärung für eine Überlegenheit binauraler Sprachwiedergabe gesehen werden. Dabei nutzt der Teilnehmer neben der phonologischen Schleife auch den räumlich visuellen Notizblock. Die Hörleistung wird so auf zwei Systeme verteilt und somit effizienter.

Letzten Endes erinnern wir uns nur an Informationen, die im *Langzeitgedächtnis* verankert werden. Es unterliegt nicht den gleichen Einschränkungen hinsichtlich Kapazität und Verweildauer wie das Arbeitsgedächtnis. Zur Art und Weise wie Wissen im Langzeitgedächtnis repräsentiert und verknüpft ist, existieren verschiedene Theorien, die den Rahmen hier allerdings sprengen würden.

Arten psychischer Belastung in der CLT

In der CLT teilt man die psychische Belastung des Lernvorgans, also die Inanspruchnahme des Arbeitsgedächtnisses, in drei Teile (**?**, Ch. 2):

Intrinsic Load wird verursacht durch die natürliche, intrinsische Komplexität der zu lernenden Informationen. Sie ist also durch die Informationen vorgegeben und damit unveränderbar.

Germane Load (dt.: relevante B.) ist die sinnvolle Belastung, die der Lernende aufbringt um Informationen ins Langzeitgedächtnis zu übernehmen, sie also zu „lernen“. Sie ist wünschenswert und soll gefördert werden.

Extraneous Load (dt.: unwesentliche B.) ist der Anteil der Belastung der vermieden werden soll und kann, indem Informationen besser präsentiert werden. Die Extraneous L. beansprucht Kapazitäten, die besser der Intrinsic– oder Germane Load zur Verfügung stehen sollten.

Psychische Überbelastung tritt dann ein, wenn die drei Anteile zusammen die Kapazität des Arbeitsgedächtnisses ausreizen.

5 Indikatoren psychischer Belastung

In diesem Kapitel zeige ich mögliche Methoden, psychische Belastung zu messen. Wir haben gesehen, dass der Begriff – unabhängig der unterschiedlichen Definitionen – verschiedene Faktoren und Auswirkungen unter sich vereint. Diese Multidimensionalität und Uneindeutigkeit äußert sich auch in der Messbarkeit. Die Methoden messen jeweils verschiedene Aspekte und korrelieren deswegen nicht zwangsläufig miteinander.

Die Messmethoden können anhand bestimmter Charakteristiken unterschieden werden. Die wichtigsten nach **?**, Ch. 3 sind:

- *Ansprechbarkeit/Sensitivität:* Erfasst das Verfahren mit hinreichender Genauigkeit den Bereich der Intensität, in dem sich meine Beanspruchung bewegt? Können vielleicht nur extreme Schwankungen erfasst werden?
- *Beeinflussung der eigentlichen Primäraufgabe:* Die Messmethode selbst sollte wenig zur psychischen Belastung beim Versuch beitragen.
- *Bandbreite und Zuverlässigkeit:* Wie zuverlässig sind die Messwerte? Wie hoch ist deren Vorhersagekraft für ähnliche Experimente?

Weitere Kriterien nennen **?** & **?**:

- *Diagnostizität:* Psychische Belastung beinhaltet verschiedene Dimensionen. Wie ist die Zusammensetzung der psychischen Belastung? Welcher Aspekt der Aufgabe trägt zur Belastung wie stark bei?
- *Akzeptanz:* Die Methode sollte von den VP angenommen werden.
- *Zeitauflösung:* Im Idealfall kontinuierlich, um kurzfristige Belastungssprünge zu erfassen.
- *Aufwand und Kosten:* für Equipment, Verarbeitung und Bewertung.

Man ordnet die Messmethoden in drei Kategorien:

- *Somatisch:* als physiologische Reaktion des Körpers (**5.1**)
- *Subjektive Schätzung:* wie die Person die Belastung erlebt (**5.2**)
- *Leistung:* in Primär- oder zusätzlicher Sekundäraufgabe (**5.3**)

5.1 Somatische Indikatoren

Hier möchte ich zuerst auf die somatischen Indikatoren der Beanspruchungsobjektivierung eingehen. Somatische Reaktionen können sich in verschiedenen physiologischen Systemen äußern, wie dem vegetativen Nervensystem, dem hormonellen System oder dem Stoffwechsel.

Aufgrund der schnelleren Ansprechzeit eignen sich zur Evaluation eines TK-Systems am ehesten Reaktionen des vegetativen Nervensystems. Grenzt man dann weiter ein nach der Eignung für Labor-Untersuchungen ergeben sich nach **?**, S.451-457 und **?**, S.157-S.168 folgende mögliche Beanspruchungsparameter:

- Herzfrequenz/-Variabilität
- Blutdruck (brachial oder peripher kontinuierlich)
- Atmung (Frequenz/-Variabilität, In-/Exspirationszeit, Volumen)
- Elektrodermale Aktivität (Hautwiderstand)
- Lidschlussfrequenz
- Muskelaktiviertheit
- Pupillenerweiterung

Einen Eindruck von der Größe der Indikatoren in der Praxis liefert der sog. „Pauli-Test“, ein Konzentrationstest bei dem fortlaufend zahlen addiert werden müssen. Dabei kommen VP auf Herzfrequenzunterschiede von ca. 10 $\frac{1}{min}$ und Variationsunterschiede von ca. 10 ms zwischen Ruhe und Belastung (**?**, S.163)

Die Vorteile somatischer Indikatoren sind, dass die VP zur Messung nicht trainiert werden muss, dass die Werte kontinuierlich erfasst werden können und dass die Werte physikalisch gemessen werden können.

Dagegen stehen allerdings einige Nachteile: Die Interpretation somatischer Daten ist schwierig. Viele unerwünschte Faktoren beeinflussen das Ergebnis, wie z. B. Alter und Gesundheit der Testperson, deren Verhalten zum Versuchszeitpunkt (Schlaf, Kaffee, Mahlzeiten) oder der individuelle Biorhythmus. Die Werte können, für sich genommen, nicht zwischen Personen verglichen werden. Man benötigt ein individuelles Bezugsniveau. Eine Schwierigkeit besteht in der Wahl eben dieses Ruhepunktes: Versuchspersonen sind *vor* einem Versuch vielleicht schon angespannt, *danach* aber ermüdet.

Es sei hier darauf hingewiesen, dass die genannten Probleme im Allgemeinen auch bei den anderen Methoden auftreten. Ich führe sie hier auf um zu zeigen, dass eine vermeintlich objektive, weil physikalische, Messung nicht zwangsläufig zu absolut validen Werten führt.

Laut **?** wurden psychophysiologische Methoden bisher nur selten bei Hörversuchen angewandt. Es ist auch unklar ob eine realitätsnahe TK-Situation überhaupt „belastend“ genug ist um eine Kompensationsaktivität hervorzurufen und damit messbare Unterschiede zu erzeugen (Sensitivität). Nicht zuletzt erfordert die Messung spezielle Geräte und teilweise besonderes medizinisches Fachwissen. Ich werde somatische Parameter in dieser Arbeit nicht weiter verfolgen.

5.2 Subjektive Schätzung

Belastung wird von der Versuchsperson auch bewusst wahrgenommen. Um diese Wahrnehmung zu erfassen wurden verschiedene standardisierte Methoden, bzw. Skalen entwickelt. Befragungen sind wahrscheinlich die verbreitetste Methode psychische Belastung zu bestimmen, sicher auch weil sie relativ einfach durchzuführen sind. Man kann die Skalen in unidimensionale und multidimensionale Skalen einteilen.

5.2.1 Unidimensionale Skalen

ITU 800: Listening Effort Scale

Die *International Telecommunication Union (ITU)* gibt Empfehlungen für subjektive Qualitätstests (**?**). Teil davon ist auch die fünfstufige, eindimensionale, diskrete „Listening-effort scale“ (Abb. 5.1).

Effort required to understand the meanings of sentences	
Complete relaxation possible; no effort required	5 ◯
Attention necessary; no appreciable effort required	4 ◯
Moderate effort required	3 ◯
Considerable effort required	2 ◯
No meaning understood with any feasible effort	1 ◯

Abbildung 5.1: ITU Listening-effort scale (Nach: **?**, S. 19)

Sie ist an sich eine Kategorieskala, allerdings wird zur Auswertung trotzdem ein Mittelwert berechnet, der sog. *Mean Opinion Score*. In ihrer Formulierung bezieht sich die Skala nur auf Tests zur Satzverständlichkeit.

Skala aus der Hörgerätetechnik

Eine ähnliche Skala wird in der Hörgerätetechnik verwendet (z.b. in **??**). Diese Skala besteht aus 13 Stufen, von denen jede zweite verbal verankert ist, allerdings mit knapperen Beschreibungen als in der ITU-Skala.

How much effort does it require to listen and understand the sencences		Score
○	Extreme effort	12
○		11
○	Much effort	10
○		9
○	Considerable effort	8
○		7
○	Moderate effort	6
○		5
○	Little effort	4
○		3
○	Very little effort	2
○		1
○	No effort	0

Abbildung 5.2: Skala nach **?**, wie sie in der Hörgerätetechnik eingesetzt wird

SEA

Kategorieskalen, wie der ITU-Skala wird implizit ein Intervallniveau zugeschrieben. Allerdings kann man sich zurecht fragen, ob in der Wahrnehmung der VP die verbalen Anker äquidistant sind. Spätestens bei einer Übersetzung ins Deutsche steht man also vor dem Problem semantischer Unterschiede (**?**).

Die Skala zur *Erfassung subjektiv erlebter Anstrengung (SEA)* von **?** wurde empirisch so konstruiert, dass die Position der verbalen Attribute möglichst der semantisch wahrgenommenen Verteilung entspricht (Abb. 5.3). So liegen z. B. „kaum-“ und „etwas anstrengend“ näher beieinander als „einigermaßen-“ und „ziemlich anstrengend“.

Leider konnte Eilers die Validität der Skala nicht hinreichend genau beweisen.

Weitere Skalen

Im Zusammenhang mit 3D-Telekonferenzen wurden bereits Evaluationen zur Höranstrengung durchgeführt. In **?** wurden den Teilnehmern nach dem Hören Skalen vorgelegt, deren Fragen sich auf konkrete Schwierigkeiten einer Telekonferenz beziehen (Abb. 5.4). Die

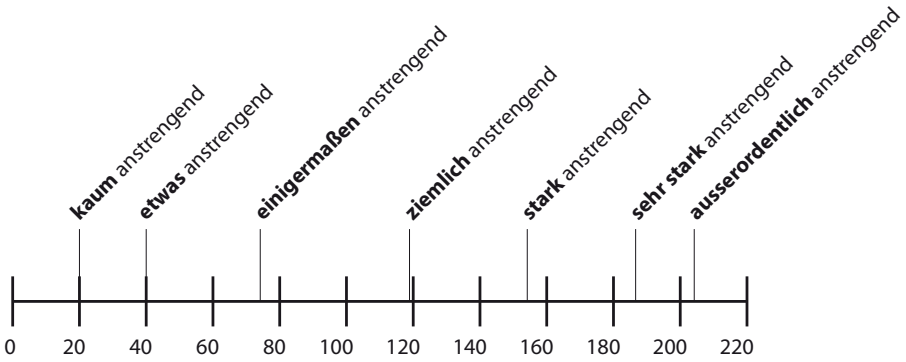


Abbildung 5.3: SEA Skala (nach: **?**)

verwendeten Skalen sind dabei links und rechts dünn erweitert (**?**). Damit soll verhindert werden, dass die Versuchspersonen Extremwerte vermeiden und so die Skala nicht voll ausnutzen (*Contraction Bias* (**?**)).

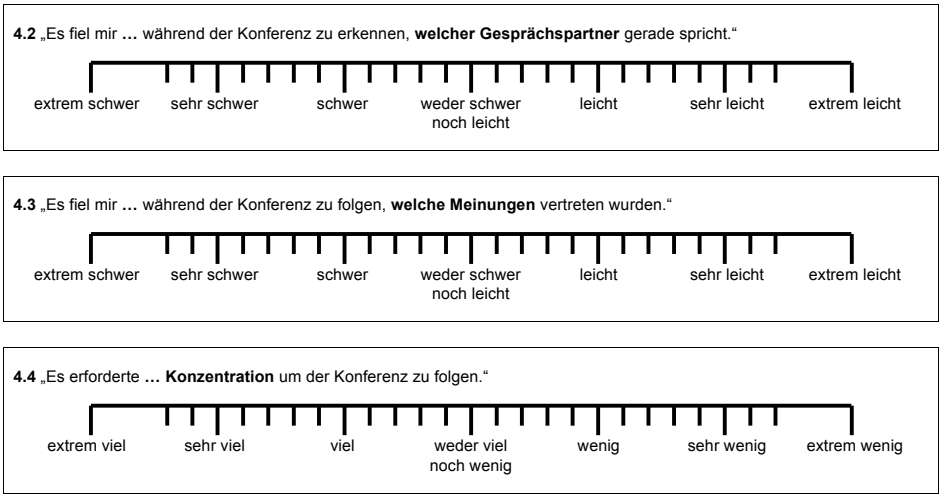


Abbildung 5.4: Fragen zur psychischen Belastung, wie sie im Fragebogen von **?** enthalten sind

Ähnliche Fragen wurden auch in anderen Untersuchungen zu 3D-Telekonferenzsystemen verwendet. **?** fragte: „During the conference, determining which conferee was speaking was: Easy – Difficult “ und „How much of your attention was focused on determining who was speaking: Most – None “, jeweils anhand einer 7-stufigen Skala. Auch **?** verwendet sinngemäß die gleichen Fragen.

MCH

Die *Modified Cooper Harper Scale (MCH)* von **?** ist eine Skala im Stil eines Flowcharts (Abb. 5.5). Dabei arbeitet sich die VP bei der Bewertung vom Startknoten zuerst durch mehrere Ja/Nein-Fragen, um sich dann letztendlich für eine von drei möglichen Stufen zu entscheiden.

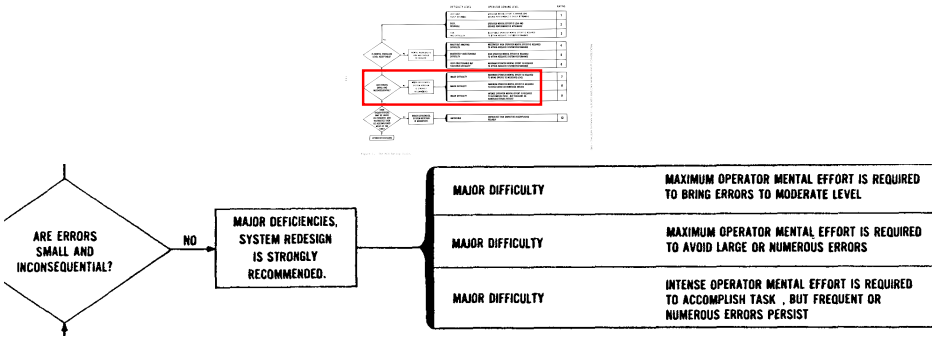


Abbildung 5.5: Ausschnit des MCH (aus: **?**, S. 133)

5.2.2 Multidimensionale Skalen

Multidimensionale Skalen versuchen die Belastung in Teilaspekte aufzuschlüsseln. Als Beispiel möchte ich hier nur den *NASA-TLX* aufführen:

NASA-TLX

In anderen Branchen, als der Telekommunikation, haben Skalen zur Bestimmung psychischer Belastung lange Tradition. So hat die *NASA* den *Task Load Index* entwickelt (**??**). Der *NASA-TLX* hat sein ursprüngliches Anwendungsfeld Luft und Raumfahrt schnell überschritten und wird nun größtenteils zur Evaluation von audiovisuellen Displays oder Eingabegeräten verwendet.

Der NASA-TLX verwendet *sechs* Skalen um verschiedene mögliche Dimensionen der Anstrengung zu erfassen (siehe Abb. 5.6 auf S. 38). Durch ein Gewichtungsschema kann ein zusammenfassender Wert berechnet werden. Hierbei ist interessant, dass die Gewichte der Dimensionen auf die einzelnen VP zugeschnitten werden. Dazu ist eine zusätzliche Prozedur nach dem Versuch nötig. Hier bewerten die VP die sechs Dimensionen, durch paarweise Vergleiche dahingehend, wie sehr sie jeweils ihr Belastungsempfinden geprägt haben. Damit sollen zum einen die Varianz zwischen den VP verringert und zum anderen unterschiedliche Aufgaben vergleichbar gemacht werden.

In einem audiologischen Experiment verwendet **?** diese Skala um die Belastung bei einer Mehrsprecher-Situation zu untersuchen.

Die höhere Diagnostizität multidimensionaler Skalen bringt auch Probleme mit sich. So muss man Versuchspersonen vor dem Test ausgiebig mit den Skalen vertraut machen. Nur so kann sichergestellt werden, dass die Subskalen durchgehend gleich verstanden werden. Des weiteren sind manche Dimensionen, wie der „Physical Demand“ des NASA-TLX nicht in jedem Fall sinnvoll.

5.3 Leistung

Unter den Leistungsbasierten Indikatoren verstehe ich die Leistung, die eine Versuchsperson bei Testaufgaben erbringt. Man kann folgende Arten von Aufgaben unterscheiden:

- Primäraufgaben (z. B. Erinnerungsvermögen nach Telekonferenz)
- Sekundäraufgaben („Multitasking-Fähigkeit“)

5.3.1 Primäraufgaben

Es ist nicht einfach zu sagen was die Primäraufgabe einer Kommunikation ist, was also eine effektive Kommunikation ausmacht. Kommunikation ist schließlich ein interaktiver, sozialer Vorgang, der auch nichtsprachlichen Komponenten beinhaltet. Deswegen vereinfachen wir im Weiteren den Begriff und definieren Kommunikation als einen einseitigen Vorgang des Hörens und Verarbeitens von Sprache.[1]

Im Allgemeinen unterscheiden sich die Primäraufgaben – und damit die verwendeten Stimuli – in der externen Validität, also ihrer Natürlichkeit bzw. Verallgemeinerungsfähigkeit, und der internen Validität, also der Zuverlässigkeit und Robustheit der Ergebnisse. Diese beiden Aspekte schließen sich teilweise aus und müssen deswegen abgewogen werden.

Sprachverständlichkeit

Eine naheliegende und verbreitete Primäraufgabe bei Telekommunikationstechnologien ist das Testen der *Sprachverständlichkeit*. Es werden einzelne Worte (**?**) oder ganze Sätze (**?**) vorgespielt und abgefragt. Als Maß wird die *Speech Reception Threshold (SRT)* bestimmt, das Pegelverhältnis zwischen Sprache und Störgeräusch bei dem gerade 50% der Worte vom Probanden verstanden wurden. Ein Beispiel für hierbei verwendete Testdaten ist der Oldenburger Satztest, mit Sätzen der festen grammatikalischen Form `Name - Verb - Zahlwort - Adjektiv - Objekt`.

[1] Hier sei natürlich trotzdem auf realitätsnähere, aber aufwendigere *Konversationstests* hingewiesen, wie bei (**?**) und (**?**)

Eine andere, interaktive Methode kommt aus dem Militär – der *Coordinate Response Measure (CRM)* (**?**). Hierbei hört der Proband verschiedene Anweisungen der Art "Ready, `Call Sign`, go to `color number` now". Wenn das gesprochene Rufzeichen (`Call Sign`) seinem eigenen, vorher zugewiesenen, Rufzeichen entspricht muss er mit der gehörten Nummer und Farbe antworten. Anhand der Erkennungsrate oder der Reaktionszeit kann dann auf die Verständlichkeit geschlossen werden. Der CRM passt zu Aufgaben aus dem militärischen Bereich, wie dem Abhören von Funk, aber weniger zu Situationen wie sie in geschäftlichen Telekonferenzen auftreten.

Sprachverständnis

Eine weiteres Merkmal einer effektiven Kommunikation ist sicherlich auch, ob sich Teilnehmer *nach* einer Konferenz an deren Inhalt erinnern können, oder daran, wer was gesagt hat. Im Gegensatz zur Sprach*verständlichkeit* bezeichnet **?** dies als *Verständnis* (eng. *Comprehension*).

Solche Verständnis-Versuche wurden bereits bei anderen 3D-Telekonferenzsystemen durchgeführt. Dabei fragte man die Probanden im Anschluss der TK u.A. danach, welcher Sprecher bestimmte Aussagen getroffen hat und wie sicher sie sich ihrer Erinnerung sind. Auch wurden die Kandidaten gebeten die Standpunkte der Sprecher schriftlich zusammenzufassen (**???**).

Solche Erinnerungstests weisen eine hohe externe Validität auf, oder anders gesagt: der Vorteil einer hohen Leistung in einer solchen Primäraufgabe leuchtet ein. Zur Messung der psychischen Belastung sind sie alleine aber nicht ausreichend. Dies liegt daran, dass Menschen in Stresssituationen ihre Leistung erst einmal durch eine größere Anstrengung kompensieren können (**?**).

5.3.2 Sekundäraufgaben

Neben der Primäraufgabe kann man der VP auch eine künstliche Zweitaufgabe stellen, die zeitgleich bearbeitet werden soll. Man kann so beispielsweise Rückschlüsse auf freie kognitive Kapazitäten und damit die psychische Belastung ziehen. Dies macht besonders dann Sinn, wenn Personen trotz höherer Anstrengung – oder besser gesagt *wegen* höherer Anstrengung – ihre Leistung in einer Primäraufgabe hoch halten können.

Verschiedene Arten von Sekundäraufgaben sind denkbar. In der Regel sind diese möglichst einfach gehalten, um schnell erlernbar zu sein, und um die Primäraufgabe nicht zu sehr zu beeinflussen.

Nach **?** können Sekundäraufgaben auf zwei verschiedene Arten eingesetzt werden. Eine, bereits angesprochene Methode sieht vor, dass die VP versuchen soll, sich hauptsächlich auf die Primäraufgabe zu konzentrieren, um dort konstant gute Leistung zu erbringen. Je mehr man sich dabei für die Primäraufgabe anstrengen muss, desto weniger Kapazitäten bleiben für die Sekundäraufgabe, deren Leistung sich dementsprechend verringert (**?**).

Bei dieser Methode kann die Leistung in der Sekundäraufgabe als Multitasking-Fähigkeit interpretiert werden.

Die andere Art zielt darauf ab, durch die Sekundäraufgabe die Leistung in der Primäraufgabe absichtlich zu verringern. Auf diese Weise verschiebt man den Sensitivitäts-Bereich der Primäraufgabe. Diese Methode macht sich also den eigentlich größten Nachteil von Sekundäraufgaben zunutze, nämlich, den Versuch zu beeinflussen und damit auch die Werte anderer Indikatoren zu verändern.

Beispiele für Sekundäraufgaben sind:

- In einem Experiment zu binauraler TK-Wiedergabe von **?** sollten die VP als Sekundäraufgabe *Buchstaben zählen*, also beispielsweise alle ‚e'-s in einem Exzerpt von Kafka's „Verwandlung". Im Ergebnis erhöhte die binaurale Wiedergabe die Sekundärleistung („e's per second ") nur wenn den Teilnehmern die Richtung des Schlüsselworts vorher gesagt wurde. Die Primäraufgabe hier war das Heraushören eines Schlüsselworts aus einer TK.

- Bei einer Untersuchung zu Hörschäden von **?** sollten sich die VP *Zahlenreihen einprägen*, und bei der darauffolgenden Höraufgabe möglichst nicht vergessen.

- In einer allgemeinen Studie zur CLT wurde von der VP verlangt einen langsamen *Farbwechsel am Bildschirm erkennen* (**?**).

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Abbildung 5.6: NASA Task Load Index (*Quelle:* `http://humansystems.arc.nasa.gov/groups/TLX/paperpencil.html`)

6 Versuchsaufbau

6.1 Überblick

In einem Hörexperiment versuche ich die psychische Belastung zu erfassen. Dabei hören die Versuchspersonen realistische Konferenzaufnahmen über Kopfhörer und beantworten danach Fragen zum Inhalt der Konferenz und zu ihrer gefühlten Anstrengung.

Ziel der Evaluation ist es, das am LDV entwickelte System mit einem „klassischen“ Fall zu vergleichen. Diese zwei Testfälle unterscheiden sich darin wie die Stimuli des Hörversuchs erstellt werden (Tab. 6.1):

	LDV-System (B)	*„Klassisch“* (M)
Aufnahme	Konferenzspinne: 8 Kanäle	Mono: 1 Kanal der Konferenzspinne
Verarbeitung	Algorithmus zur Sprechertrennung	-
Wiedergabe	Binaural: ohne Tracking, KEMAR-HRTF	Diotisch (Mono)

Tabelle 6.1: Zu vergleichende Testfälle und die Entstehung der Stimuli.

Der Test ist aufgeteilt in zwei eigenständige, aufeinanderfolgende Tests: Teil ① und ② (Tab. 6.2). Die Idee ist, im zweiten Test eine Situation mit höherer Belastung zu erzeugen. Die höhere Belastung besteht einerseits darin, dass die Versuchspersonen bereits ca. eine Stunde Teil ① bearbeitet haben und andererseits, dass mit Teil ② eine zusätzliche Sekundäraufgabe hinzukommt. Dabei müssen die Versuchspersonen während des Hörens der Konferenzen zusätzlich einen Reaktionstest bearbeiten.[1]

Einen schnellen Überblick über den Versuch liefert die Anleitung für die Teilnehmer auf Seite 50.

[1] Anfänglich war geplant, für Teil ② auch die langen Konferenzen mit mehr Sprechern zu verwenden. Verbunden mit der Sekundäraufgabe wäre dies für die VP allerdings *zu* anstrengend und langwierig geworden. Dieser Kompromiss verringert natürlich die Belastung in Teil ②, und damit leider auch dessen Aussagekraft.

	Teil ①	*Teil ②*
Dauer	$\sim$ 1h	$\sim$ 30 min
Sekundäraufgabe	-	Reaktionstest
Konferenzen	3 · 3 Sprecher	3 · 2 Spr.
	3 · 4 Spr.	1 · 3 Spr.

Tabelle 6.2: Die zwei Teile des Versuchs

6.2 Stimuli

6.2.1 Konferenz-Korpus (?)

Als Grundlage für die Stimuli kann ich dankenswerterweise auf realistische Konferenzaufnahmen von **?** zurückgreifen. Die Konferenzen wurden von professionellen Sprechern auf deutsch aufgenommen und sind relativ natürlich, weisen also eine gewisse externe Validität auf. Dennoch sind sie untereinander, in ihrer inhaltlichen Struktur, vergleichbar. Der Korpus beinhaltet 13 verschiedene Konferenzen zu verschiedenen Themen, die sich in Anzahl der Sprecher und ihrer Länge unterscheiden. In meinem Versuch verwende ich 11 Konferenzen davon (Tabelle 6.3).

						Lautsprecher			
	ID	*Thema*	*# Sprecher*	*Länge*	*Orig. ID*	*L1*	*L2*	*L3*	*L4*
Einführung	**0**	-	• • •	1:07	7*	-	-	-	-
Teil ①	**4**	Konferenz	• • •	6:17	1	B	C	A	
	5	Meeting	• • •	5:51	2	A	B	C	
	6	Jubiläum	• • •	5:02	10	F		D	E
	7	CD	• • • •	8:13	3	C	D	A	B
	8	Stadtfest	• • • •	8:14	8	C	D	B	A
	9	Malereien	• • • •	8:19	9	E	F	A	B
Teil ②	**1**	Messe	• •	3:53	4	D		C	
	2	Projekt	• •	3:51	12		E		F
	3	Film	• •	3:47	13		F		E
	10	Auto	• • •	6:06	11	D		F	E

Tabelle 6.3: Die im Experiment verwendeten Konferenzen. L1-L4 bezeichnen die verwendeten Lautsprecherpositionen bei meinen Aufnahmen (Kapitel 6.2.2). Konferenz 0 ist ein Zusammenschnitt aus der Vorlage 7 des Korpus. Format: 44,1 kHz, 32-bit, wav.

Alle Konferenzen haben inhaltlich einen ähnlichen Aufbau. Der erste Sprecher übernimmt dabei immer die Diskussionsleitung. Die Konferenzen bestehen aus vier Phasen: Vorstellungsrunde, Problemlösung, Informationsaustausch und Verabschiedung. Die Problemlösungsphase stellt dabei den längsten und wichtigsten Teil dar. In ihr werden nacheinander verschiedene Probleme abgearbeitet. Jeder Teilnehmer ist dabei gleich stark vertreten. Es wurde sichergestellt, dass auch die Konferenzen mit verschiedener Anzahl an Sprechern vergleichbar sind. So kommt mit jedem neuen Teilnehmern auch immer die gleiche Menge an Information hinzu.

Zur Veranschaulichung ist hier eine kurze Transkription von ca. 20 Sekunden aus der Mitte von Konferenz Nr. 3 („Film“):

HR. SILBERLING: *... würd ich gerne wissen welche speziellen Anforderungen seitens des Drehbuchs gibt's den da?*
HR. FINDBERGER: Äh des ist ganz klar, also wenn Sie das Drehbuch gelesen haben . . . oder hätten, ja. Es ist ganz klar was benötigt wird: Es ist zum einen – ist es roter Sand...
HR. SILBERLING: *Mhhmm.*
HR. FINDBERGER: ...Hügel...
HR. SILBERLING: *Ja.*
HR. FINDBERGER: ...und auch ne Ebene Fläche...
HR. SILBERLING: *Mhhmm.*
HR. FINDBERGER: Also es sind drei Dinge. Ja?
HR. SILBERLING: *Mhhmm.*
HR. FINDBERGER: Wie gesagt ich wiederhol gern nochmal: ...

Bei den ursprünglichen Aufnahmen wurde für die Sprecher eine TK-Situation simuliert. Dabei saßen die Sprecher gemeinsam in einem reflexionsfreien Raum, ohne sich aber zu sehen oder sich direkt hören zu können. Sie hörten sich untereinander nur über Kopfhörer unter Verwendung eines 3D-Rendering-Systems der TU Berlin. Nach den Aufnahmen wurden die Audiodateien noch zu- und umgeschnitten um der gewünschten inhaltlichen Struktur zu entsprechen. Überlappungen, also Momente in denen mehrere Sprecher gleichzeitig sprechen, machen im Mittel ca. 5% der Konferenzlänge aus.

6.2.2 Aufnahme

Die zur Verfügung gestellten Konferenzen habe ich mit der Konferenzspinne aufgenommen. Aus Sicht der Spinne wurde somit eine realitätsnahe Konferenz simuliert. Dafür wurden vier Lautsprecher in gleichem Abstand um die Konferenzspinne aufgestellt (Abb. 6.1) und die Testkonferenzen darüber abgespielt. Jeder Lautsprecher entsprach dabei einem realen Sprecher. Die Verteilung der Sprecher ist in Tabelle 6.3 dargestellt.

Als Aufnahmeraum wählten wir ein Labor in dem normalerweise subjektive Videotests durchgeführt werden. Der Raum wurde einem echofreien Raum vorgezogen, da er etwas

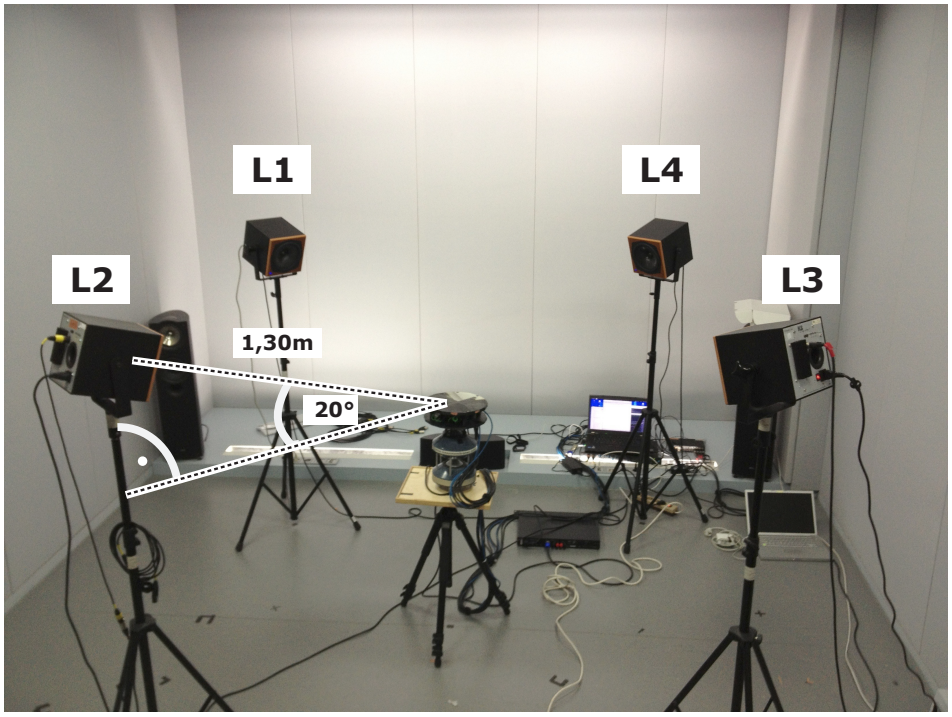


Abbildung 6.1: Aufnahme der Testkonferenzen über die Konferenzspinne. Die Lautsprecher wurden dabei in gleichem Abstand (r = 1,30m) und gleicher Elevation ($\delta = 20°$) um die Konferenzspinne aufgestellt ($\Delta\phi = 90°$)

näher an den akustischen Gegebenheiten eines Büroraums ist, aber dennoch genormt ist (ITU-R BT.500-12).

Das Aufnahme-Setup ist in Bild 6.2 dargestellt. Vor der Wiedergabe wurde die Abtastrate der Samples von 44,1 kHz auf 48 kHz hochgerechnet. Zeitgleich zur Audio-Aufnahme wurde die Raumtemperatur durch einen Sensor erfasst.

6.2.3 Sprechertrennung

Die aufgenommenen 8-kanaligen Konferenzen wurden dann mit dem Separierungsalgorithmus bearbeitet, um einzelne Sprecher-Kanäle für die Wiedergabe zu erhalten. Dazu wurde zuerst für jede Konferenz ein Sprechermodell trainiert. Als Trainingsdaten wurden für jeden Sprecher Abschnitte von 10–14 s Länge aus den Vorstellungsrunden der jeweiligen Konferenzen geschnitten. Jede Konferenz hat so für jeden Sprecher ihr eigenes Model. Nach dem Training erfolgte die eigentliche Separierung durch den Algorithmus. Das Format der Dateien vor der letztendlichen Wiedergabe ist wav (PCM) mit 16bit bei 48000 Hz.

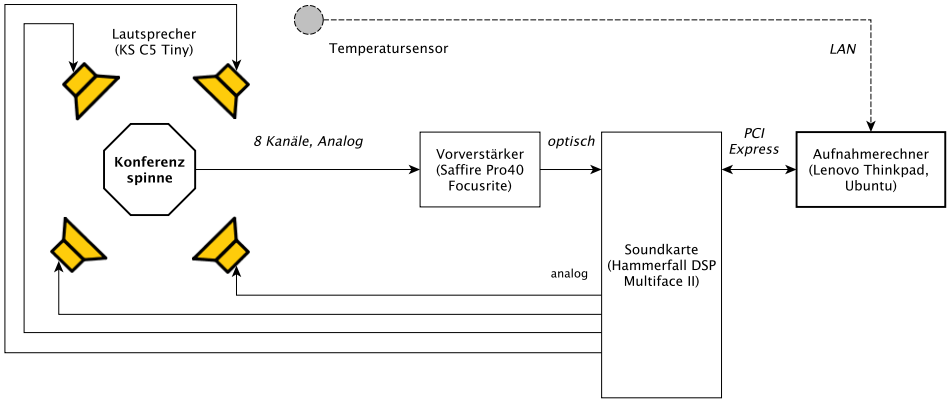


Abbildung 6.2: Equipment und Verschaltung bei der Aufnahme (Siehe auch: Foto 6.1)

6.2.4 Wiedergabe

Testfall B: Binaurale Widergabe

Die Wiedergabe beim Versuch erfolgt *ohne* das Face-Tracking-System und mit Standard-HRTFs (KEMAR). Diese Wahl gründet sich auf Ergebnissen eines Versuchs zur QoE am LDV-System[2]. Dort wurden verschiedene HRTF-Datensätze und Tracking-Methoden verglichen. Festgestellt wurden u.A. zwei Dinge:

1. Zum einen bevorzugten die Teilnehmer die Wiedergabe *mit* Tracking – allerdings nur wenn dafür ein ausgereiftes Infrarot-Tracking Produkt verwendet wurde. Dem Webcam-Tracking wurde immer noch der Fall komplett ohne Tracking vorgezogen.

2. Im Falle der HRTFs schnitten individuell gemessene HRTFs nur leicht besser ab als die Standard-Variante (KEMAR). Ohne Tracking waren die individuellen sogar schlechter.

Die Vorteile individuell gemessener HRTFs kommen also nur bei einem sehr guten Tracking-System zur Geltung. Ein solches, teures System kann allerdings kein Teil einer praktikablen Lösung sein, wie sie in dieser Arbeit bewertet werden soll.

Die Positionen der virtuellen Sprecher sind im vorderen Halbkreis auf Ohrhöhe (Elevation $\delta = 0$) verteilt (Abb. 6.4). Es werden keine Sprecher hinter dem Kopf platziert, da ohne Tracking sonst warscheinlich Front-Back-Confusions entstünden. Die genauen Winkel - besonders bei vier Sprechern - sind gewählt, da es leichter fällt Sprachquellen zu trennen, die vor dem Kopf liegen, als eher seitlich angeordnete (**?**). Extreme und damit unnatürliche Richtungen, wie senkrecht von links oder rechts, werden vermieden.

[2] Es handelt sich um *vorläufige* Ergebnisse vom 31.7.2013

Abbildung 6.3: VP beim Versuch (Das Trackingsystem wird nicht benutzt)

Testfall M: Mono Wiedergabe

Für das Vergleichssystem *M* wird aus den gleichen Aufnahmen mit der Konferenzspinne der erste Kanal verwendet und auf beiden Ohren wiedergegeben.

6.3 Indikatoren / Messung

Zur Bestimmung der psychischen Belastung verwende ich im Experiment drei Methoden (Tabelle 6.4): Eine subjektive Befragung durch zwei Skalen, die Leistung in einer Primäraufgabe (Erinnerungstest) und die Leistung in einer Sekundäraufgabe (Reaktionstest).

Messmethode	*Beschreibung des Werts*	*Symbol*
Primäraufgabe (Erinnerungstest)	Anteil korrekt zugeordneter Zitate	K
	Gewissheit der Antwort	G
Subjektiv (2 Skalen)	Allgemeine Konzentration	SK
	Schwierigkeit Sprechererkennung	SS
Sekundäraufgabe (Reaktionstest)	Reaktionszeit (Mittel)	T

Tabelle 6.4: Überblick über die erfassten Werte

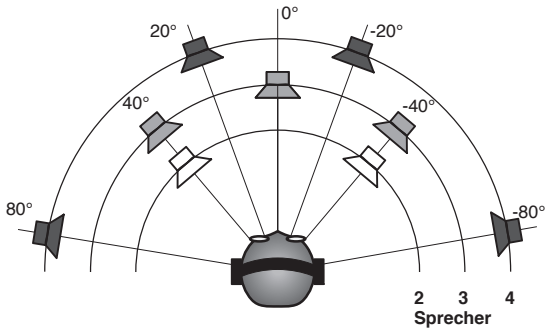


Abbildung 6.4: Positionen der Sprecher bei der binauralen Wiedergabe (Sicht von oben)

6.3.1 Primäraufgabe

Die Primäraufgabe ist ein Erinnerungstest. Die Teilnehmer sind angewiesen der Konferenz so zu folgen als wären sie „stille Teilnehmer“, die sich ein Bild von den Sprechern machen sollen.

Nach dem Hören einer Konferenz werden den Teilnehmern dann nacheinander Zitate gezeigt, die sie den Sprechern zuordnen sollen. Pro Konferenz werden 16 solcher Fragen gestellt. Dabei gibt es immer zwei Antwortmöglichkeiten (Abb. 6.5). Zu ihrer Antwort muss die VP noch angeben, wie sicher sie sich dieser ist.

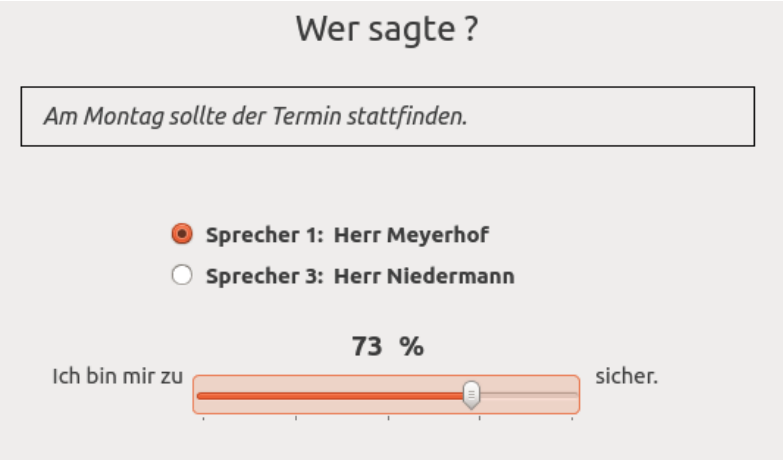


Abbildung 6.5: Inhaltsfrage der Art „Wer sagte ... ?“ (Screenshot)

Die verwendeten Zitate sind ebenfalls aus dem Konferenz-Datensatz von **?**.
Es werden folgende Werte erfasst:

Beschreibung	*Symbol*	*Wertebereich*
Anteil korrekt zugeordneter Zitate	K	0–1
Gewissheit der Antwort	G	0–1

Tabelle 6.5: Erfasste Werte bei der Inhaltsfrage

6.3.2 Subjektive Schätzung

Neben den Inhaltsfragen geben die Teilnehmer auch Auskunft über die empfundene Anstrengung. Dazu bekommen sie zwei Aussagen mit Lücken:

- „Es erforderte ___ Konzentration, der Konferenz zu folgen“
- „Es fiel mir ___ während der Konferenz zu erkennen, welcher Gesprächspartner gerade spricht“

Die Lückenwörter sind auf den Skalen als verbale Anker eingetragen (Abb. 6.6).

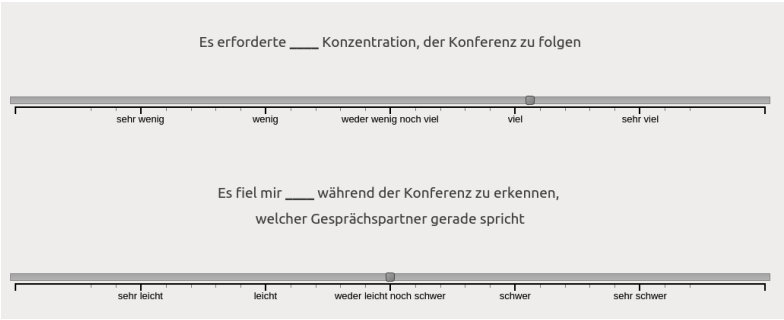


Abbildung 6.6: Skalen zur wahrgenommenen Anstrengung (Screenshot)

Für jede Skala wird damit ein Wert erfasst:

Beschreibung	*Symbol*	*Wertebereich*
Allgemeine Konzentration	SK	0–1
Schwierigkeit Sprechererkennung	SS	0–1

Tabelle 6.6: Erfasste Werte bei den Skalen

6.3.3 Reaktionstest (Teil ②)

In Teil ② bearbeiten die Versuchspersonen, gleichzeitig zum Hören, einen Reaktionstest. Bei diesem soll die VP reagieren, sobald eine schwarze Fläche ausgeblendet wird. Die Fläche „verschwindet“ im Schnitt alle vier Sekunden und der Benutzer soll mit der Maus drauf klicken, sobald er dies bemerkt. Dies wird optisch bestätigt. Nach 1500 ms wird ein Timeout ausgelößt und die Box taucht wieder auf. Hat der Benutzer bis dahin nicht reagiert wird dies als versäumtes Ereigniss gewertet. Klickt der Benutzer vor dem Verschwinden der Box, wird dies als überstürzter Klick gewertet.

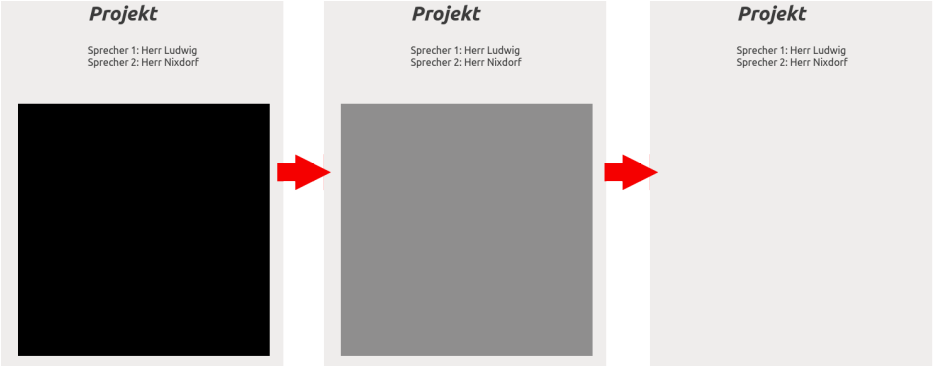


Abbildung 6.7: Reaktionstest (Screenshot)

Beim Reaktionstest werden folgende Werte erfasst:

Beschreibung	*Symbol*	*Einheit*	*Wertebereich*
Reaktionszeit (Mittel)	T	s	0–1500
Zusätzlich:			
Anteil versäumter Ereignisse (an allen Ereignissen)	NM	%	0–100
Anteil überstürzter Klicks (an allen Klicks)	NF	%	0–100

Tabelle 6.7: Erfasste Werte beim Reaktionstest

6.4 Ablauf

Der Test dauert insgesamt 1,5 h und wird im Audiolabor des LDV durchgeführt, einem annähernd echofreien Raum.

Am Anfang jedes Versuchs wird die VP eingewiesen. Dazu liest sie zuerst eine schriftliche Anleitung (Abb. 6.8). Diese wird im Anschluss mündlich besprochen. Danach kann die

VP den Reaktionstest ausprobieren und einen kompletten Trainingsblock durcharbeiten, bestehend aus Konferenz und Anschlussfragen. Kommen dabei Fragen auf, werden diese direkt geklärt. Die VP wird nun alleine gelassen um den eigentlichen Test zu bearbeiten.

Das Experimentaldesign sieht vor, dass jede VP jede *Konferenz (1-9)* genau einmal hört und dass dabei beide *Testzustände (*B, M*)* gleich oft angewandt werden. Die Reihenfolge der Stimuli und Testzustände ist in Tabelle 6.8 beschrieben. Die Abfolge der Testzustände wird für jeden Kandidaten zufällig permutiert. Zur Durchmischung der Reihenfolge der Konferenzen (1-9) wird ein lateinisches Quadrat verwendet. Dies wird für die Teil ① und Teil ② jeweils unabhängig vorgenommen.

		Teil ①						Teil ②			
VP	*Einführung*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*9*
1	0	4B	5M	9M	6B	8B	7M	1B	2M	10M	3B
2	0	5M	6M	4M	7B	9B	8B	2B	3M	1M	10B
3	0	6B	7B	5M	8M	4B	9M	3B	10M	2B	1M
4	0	7M	8M	6B	9M	5B	4B	10M	1B	3B	2M
5	0	8B	9M	7M	4B	6M	5B	1M	2B	10B	3M
6	0	9B	4M	8B	5B	7M	6M	2M	3B	1M	10B
7	0	4B	5B	9M	6B	8M	7M	3B	10M	2B	1M
8	0	5M	6M	4M	7B	9B	8B	10M	1B	3B	2M
9	0	6M	7M	5B	8M	4B	9B	1M	2B	10M	3B
10	0	7B	8M	6M	9M	5B	4B	2B	3M	1B	10M
11	0	8B	9M	7M	4B	6B	5M	3B	10M	2M	1B
12	0	9M	4B	8M	5B	7B	6M	10B	1M	3M	2B
13	0	4M	5B	9M	6M	8B	7B	1B	2B	10M	3M
14	0	5B	6B	4B	7M	9M	8M	2B	3M	1M	10B
15	0	6M	7M	5M	8B	4B	9B	3B	10M	2B	1M
16	0	7M	8B	6M	9M	5B	4B	10M	1M	3B	2B
17	0	8B	9M	7M	4B	6M	5B	1B	2B	10M	3M
18	0	9M	4B	8B	5M	7B	6M	2B	3M	1M	10B
19	0	4B	5M	9M	6M	8B	7B	3M	10M	2B	1B
20	0	5B	6M	4M	7B	9B	8M	10B	1M	3M	2B
21	0	6M	7M	5B	8B	4B	9M	1B	2M	10B	3M

Tabelle 6.8: Reihenfolge der Konferenzen (1-9) und Testzustände (B: Binaural, M: Mono) für jede Versuchsperson. Gestrichelte Linien deuten die lateinischen Quadrate an.

6.5 Hypothesen

Grundsätzlich gilt bei den späteren statistischen Tests *eine* Nullhypothese für alle Variablen, nämlich dass sich die Erwartungswerte in den Testfällen B und M nicht unterscheiden. Die Alternativhypothese ist somit, dass ein signifikanter Unterschied besteht. Obwohl diese Hypothese ungerichtet ist, erwarte ich dennoch gewisse Tendenzen:

Erinnerungstest

Ich gehe davon aus, dass die Versuchspersonen in Fall B insgesamt eine höhere Leistung erzielen und sich dabei ihrer Antworten sicherer sind.

$$K_B > K_M; \quad G_B > G_M$$

Subjektive Schätzung

Ich erwarte, dass die VP aussagen, bei B-Konferenzen weniger Konzentration zu benötigen und dass es ihnen leichter fällt zu erkennen wer gerade spricht.

$$SK_B < SK_M; \quad SS_B < SS_M$$

Reaktionstest

Zum Reaktionstest habe ich keine Hypothese. Der Zweck dieser Sekundäraufgabe ist eher die Belastung der Situation allgemein zu erhöhen.

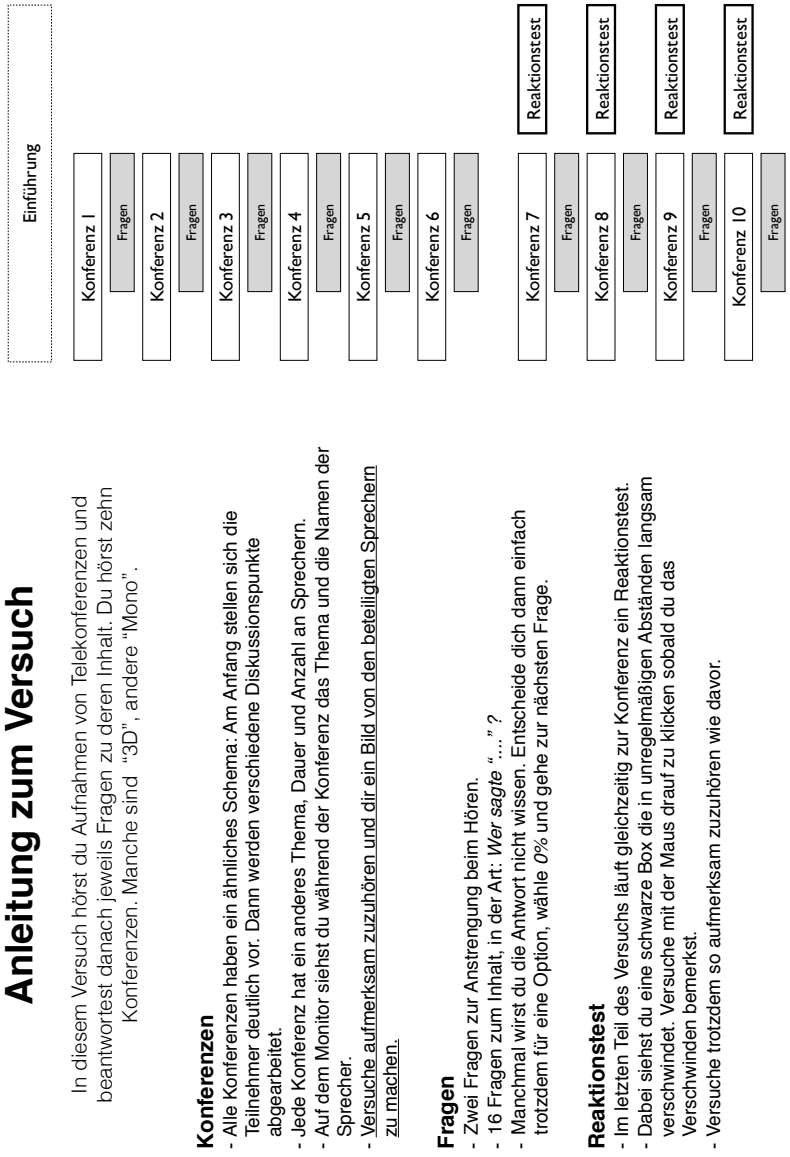


Abbildung 6.8: Anleitung für die Versuchsteilnehmer

7 Ergebnisse

Die Versuche wurden durchgeführt mit 21 Studenten und Mitarbeitern der TU München, die alle schon Erfahrungen mit Hörexperimenten gemacht hatten. Im Schnitt benötigten die VP für den Versuch 1 h und 45 min.

Im Folgenden sind die Ergebnisse dargestellt. Zuerst wird Teil ① ausführlich besprochen und dann, in kürzerer Form, Teil ②.

Es wird jeweils der binaurale mit dem Mono-Fall verglichen. Um hier statistisch auffällige Unterschiede zu erkennen, wird eine Varianzanalyse (ANOVA[1]) durchgeführt. Dieses Verfahren sagt aus, ob sich die Mittelwerte zweier Gruppen signifikant unterscheiden. Dabei wird die Varianz der Gruppen-Mittelwerte ins Verhältnis gesetzt zu der Varianz innerhalb der Gruppen (F-Wert). Am Ende erhält man einen Wert p. Dieser ist die Wahrscheinlichkeit, dass ein Experiment – unter Annahme der Nullhypothese – einen solchen, oder extremeren, Unterschied liefert. Ein möglichst kleiner Wert spricht also gegen die Nullhypothese der Gleichheit. (**?**, S. 192)

Erklärungen zu den Boxplots finden sich in Tabelle 7.1:

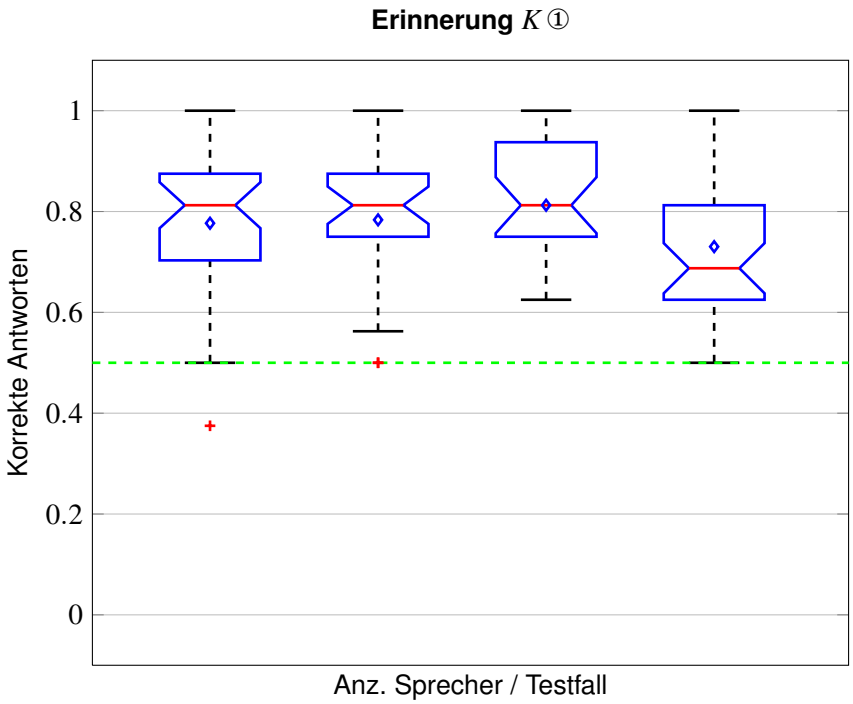
Symbol		
Plot	*Tab.*	*Beschreibung*
>—<		Median und 95%-Konfidenzintervall
▭		IQR $[Q_{.25}, Q_{.75}]$, also 50% der Werte
⊥		Whisker (Maximal $1{,}5 \cdot IQR$)
+		Ausreißer (Werte über/unter den Whiskern)
◇	$\bar{x}$	Arithmetischer Mittelwert
	$\Delta\bar{x}$	Differenz der Mittelwerte
	σ	Standardabweichung
	F	F-Statistik der Varianzanalyse
	p	p-Wert der Varianzanalyse
		$p < 0.05$: **statistisch signifikant**

Tabelle 7.1: Erklärung zu den Boxplots und Tabellen

[1] Es wird die „anova1“-Funktion von Matlab verwendet. Genau genommen handelt es sich – bei einem Vergleich von (nur) zwei Gruppen – um einen „t-test“

7.1 Teil ①

Erinnerung (Sprecherzuordnung)



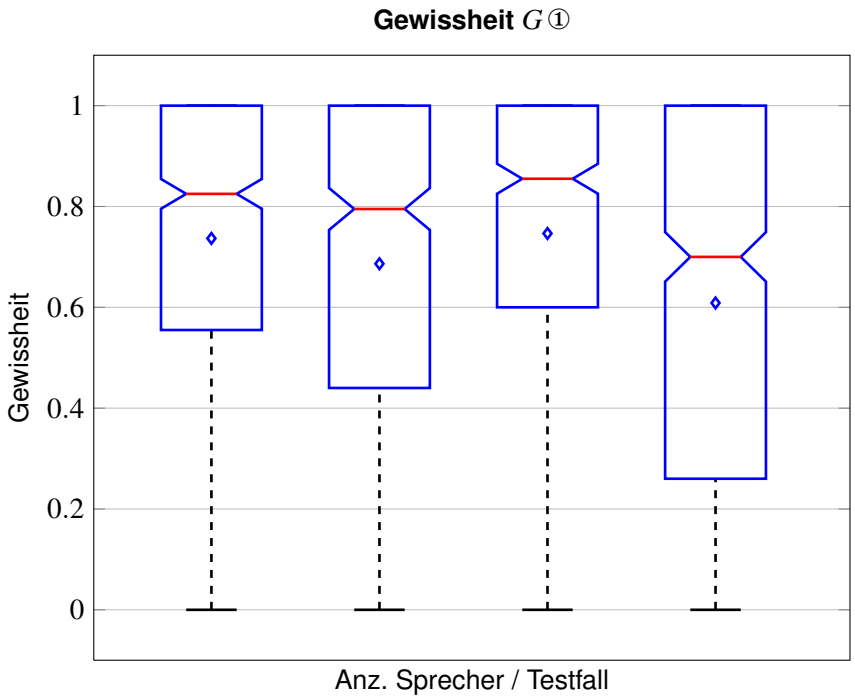
	3 Binaural	3 Mono		4 Binaural	4 Mono
$\bar{x}$	0.777	0.784	$\bar{x}$	0.813	0.730
σ	0.149	0.138	σ	0.125	0.121
$\Delta\bar{x}$	0.007		$\Delta\bar{x}$	-0.082	
F	0.03		F	6.96	
p	0.855		p	**0.011**	

Abbildung 7.1: Anteil korrekt beantworteter Fragen

Die Fragen zu den Sprecherzuordnungen wurden allgemein sehr gut beantwortet. Natürlich ist hier zu beachten, dass dies Multiple-Choice-Fragen mit immer zwei Antwortmöglichkeiten waren. So läge man auch nur durch Raten bei 50% (- - -) der Fragen richtig.

Es kann ein signifikanter Unterschied im Fall von vier Sprechern nachgewiesen werden. Dabei lagen die VP durch binaurale Widergabe um 8,3 Prozentpunkte öfter richtig.

Gewissheit



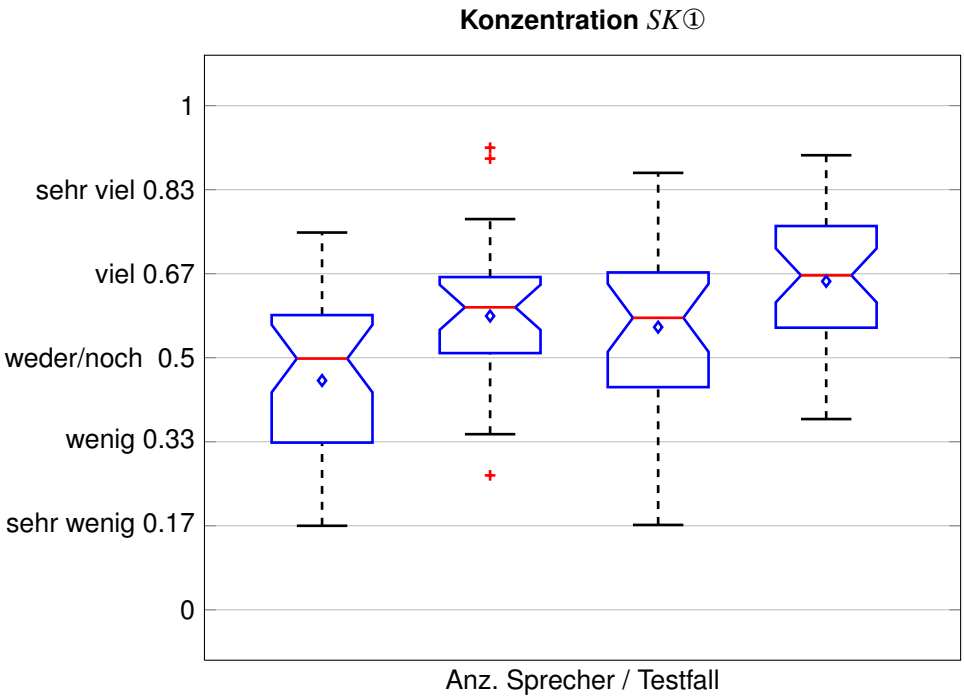
	3 BINAURAL	3 MONO		4 BINAURAL	4 MONO
$\bar{x}$	0.737	0.686	$\bar{x}$	0.747	0.609
σ	0.306	0.337	σ	0.302	0.360
$\Delta\bar{x}$	-0.050		$\Delta\bar{x}$	-0.138	
F	(6.16)		F	(41.94)	
p	**(0.013)**		p	**(<0.001)**	

Abbildung 7.2: Gewissheit bei Inhaltsfragen

Es ist erkennbar, dass sich die VP sowohl im 3- als auch im 4-Sprecher-Fall ihrer Antworten bei binauraler Wiedergabe tendenziell sicherer sind. Analog zu den Erinnerungsfragen ist dieser Unterschied bei 4 Sprechern größer.

Bemerkung: In diesem Fall vermeide ich Aussagen zur statistischen Signifikanz. Die Skala besitzt wahrscheinlich nicht das erforderliche Niveau und die Daten sind nicht normalverteilt. Grund dafür ist, dass die VP explizit angewiesen wurden, 0% zu wählen, falls sie die Antwort nicht wissen.

Skala zur Konzentration



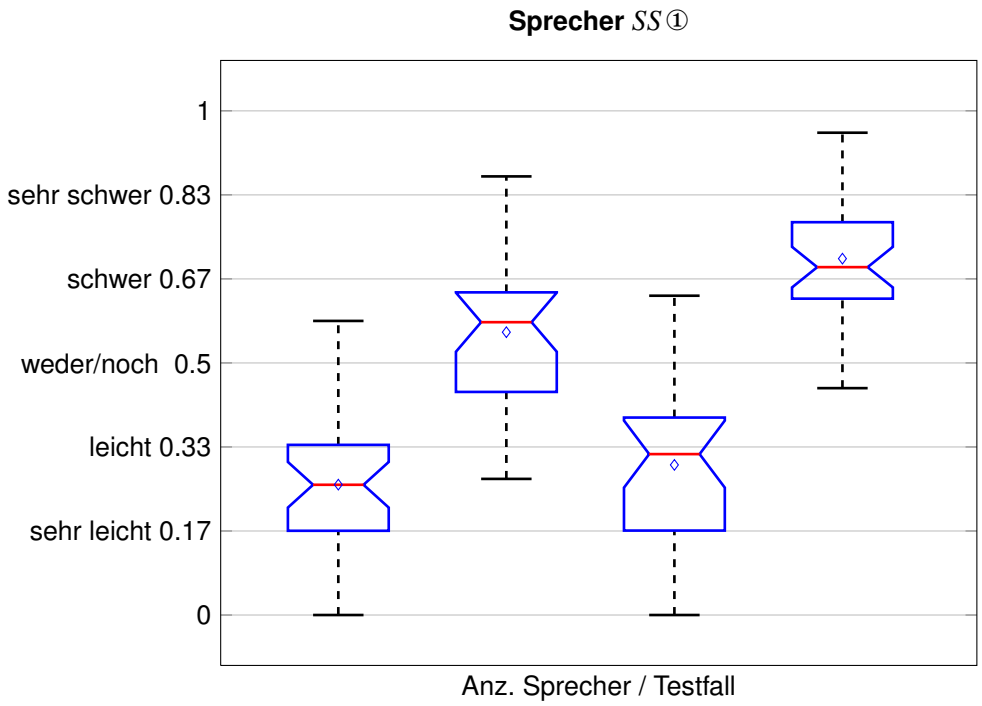
	3 BINAURAL	3 MONO		4 BINAURAL	4 MONO
$\bar{x}$	0.455	0.583	$\bar{x}$	0.561	0.652
σ	0.166	0.156	σ	0.175	0.148
$\Delta\bar{x}$	0.128		$\Delta\bar{x}$	0.091	
F	9.78		F	4.99	
p	**0.003**		p	**0.029**	

Abbildung 7.3: Skala zur allgemeinen Konzentration

Die Angaben zur Konzentration bewegen sich eher im Mittelfeld. Insgesamt benötigten die Versuchspersonen also weder „sehr wenig“ noch „sehr viel “ Konzentration um den Konferenzen zu folgen.

Es war jedoch signifikant weniger Konzentration erforderlich um den binauralen Konferenzen zu folgen.

Skala zur Sprechererkennung



	3 Binaural	3 Mono		4 Binaural	4 Mono
$\bar{x}$	0.259	0.561	$\bar{x}$	0.298	0.707
σ	0.126	0.171	σ	0.154	0.115
$\Delta\bar{x}$	0.302		$\Delta\bar{x}$	0.409	
F	65.22		F	146.34	
p	**<0.001**		p	**<0.001**	

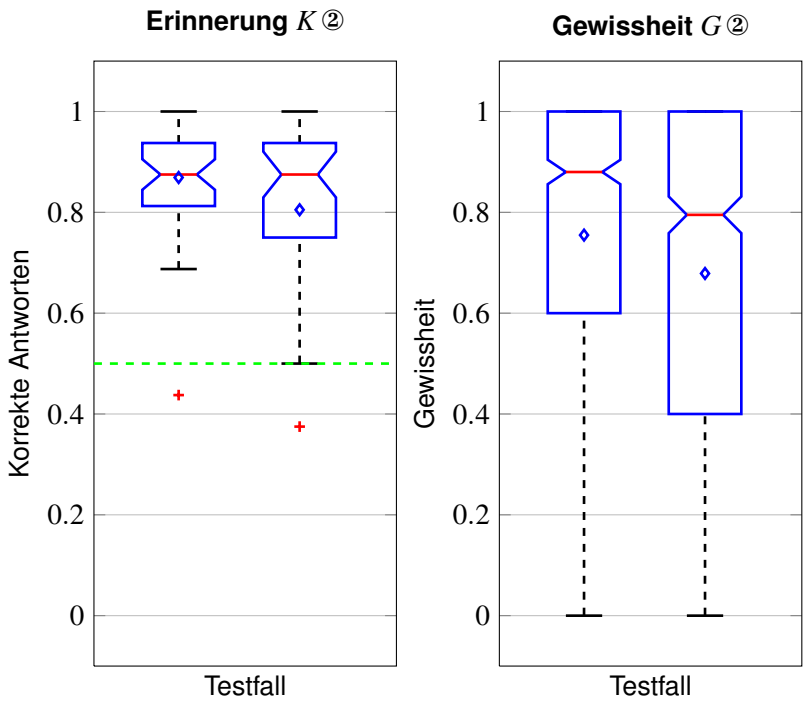
Abbildung 7.4: Skala zur Sprechererkennung

Es fiel den Versuchspersonen deutlich leichter, zu erkennen, wer spricht, wenn die binaurale Wiedergabe verwendet wurde. In binauralen Konferenzen fiel es ihnen „(sehr) leicht“, in Mono-Konferenzen „schwer“.

7.2 Teil ②

Der zweite Teil wird nicht nach Sprechern aufgeschlüsselt, da die Sprecheranzahl hier nicht gleich verteilt ist (3 · 2 Spr., 1 · 3 Spr.) und somit keine unabhängige Variable darstellt.

Erinnerung (Sprecherzuordnung) und Gewissheit



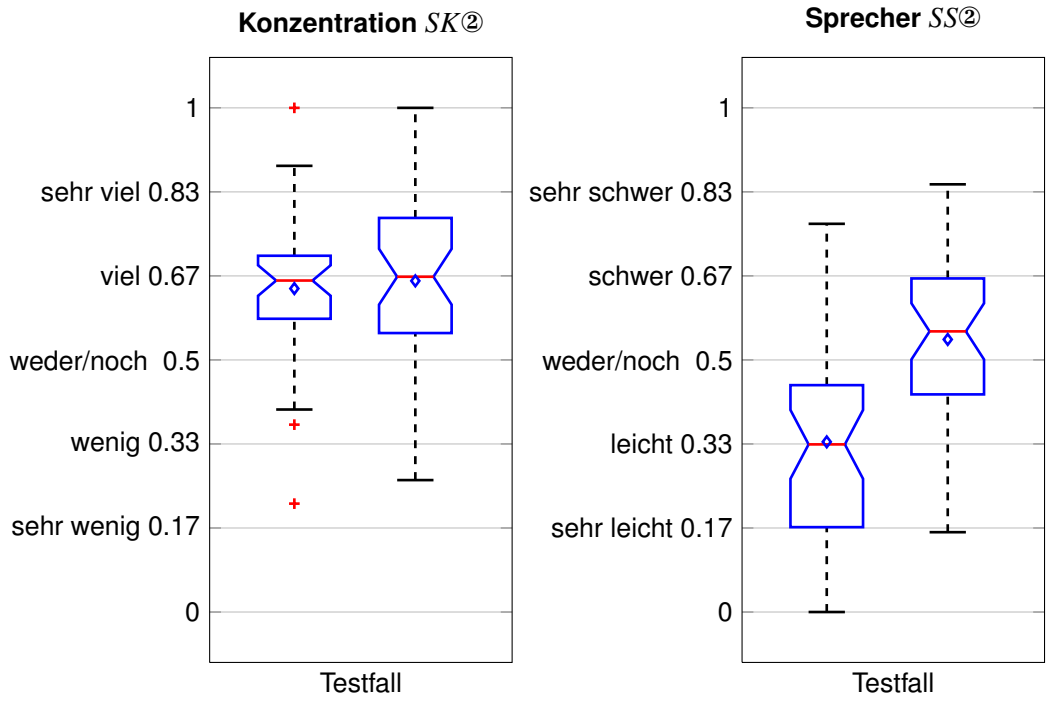
	BINAURAL	MONO
$\bar{x}$	0.869	0.805
σ	0.107	0.143
$\Delta\bar{x}$	-0.064	
F	5.4	
p	**0.023**	

	BINAURAL	MONO
$\bar{x}$	0.755	0.679
σ	0.230	0.342
$\Delta\bar{x}$	-0.076	
F	(18.82)	
p	**(<0.001)**	

Abbildung 7.5: a) Anteil korrekt beantworteter Fragen b) Gewissheit bei Inhaltsfragen

Im Mono-Fall konnten die VP signifikant weniger Sätze korrekt zuordnen und waren sich ihrer Antworten außerdem unsicherer. Allerdings ist dieser Unterschied nicht sehr ausgeprägt.

Skala zur Konzentration und zur Sprechererkennung



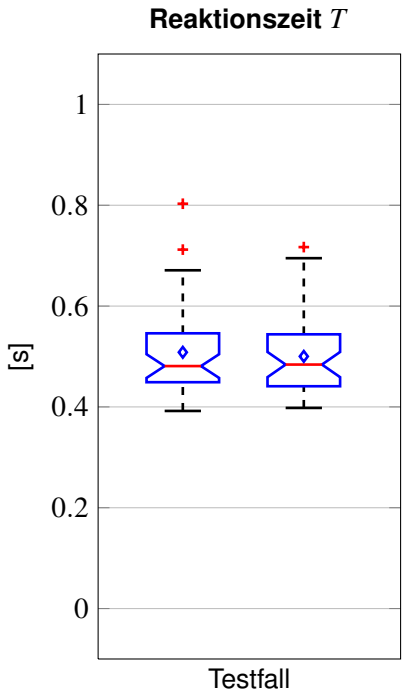
	BINAURAL	MONO
$\bar{x}$	0.642	0.657
σ	0.149	0.177
$\Delta\bar{x}$	0.015	
F	0.18	
p	0.670	

	BINAURAL	MONO
$\bar{x}$	0.338	0.541
σ	0.185	0.169
$\Delta\bar{x}$	0.203	
F	27.58	
p	**<0.001**	

Abbildung 7.6: a) Skala zur allgemeinen Konzentration b) Skala zur Sprechererkennung

Für Mono- und Binaural-Konferenzen war ähnlich viel Konzentration nötig. Dabei fiel es den Teilnehmern im binauralen Fall allerdings trotzdem deutlich leichter, zu erkennen. wer gerade spricht.

Reaktionszeit



	BINAURAL	MONO
$\bar{x}$	0.508	0.500
σ	0.087	0.078
$\Delta\bar{x}$	-0.008	
F	0.2	
p	0.652	

Abbildung 7.7: Reaktionszeit

Die Reaktionszeit ist in beiden Fällen gleich.[2]

[2] Die zusätzlichen Parameter zum Reaktionstest wurden nicht ausgewertet.

8 Fazit

Ziel dieser Arbeit war es, herauszufinden ob das 3D-Telekonferenzsystem des LDV die psychische Belastung bei der Kommunikation verringert und somit effizienter und effektiver als mit einem Mono-Vergleichssystem macht.

Im ersten Schritt wurden dazu die Hintergründe der beiden relevanten Themenfelder beleuchtet: Räumliches Hören und Psychische Belastung. Letztere wurde aus zwei Sichtweisen beschrieben, nämlich der Arbeitswissenschaft und der *Cognitive Load Theory*. Danach wurden mögliche Messverfahren vorgestellt (somatisch, subjektiv und leistungsbasiert) und in Hinblick auf den geplanten Versuch beurteilt. Verwendet wurden davon letztendlich zwei subjektive Indikatoren (Konzentration, Sprecher) und zwei leistungsbasierte Indikatoren (Erinnerungstest und Reaktionstest). Die Versuchs-Stimuli wurden erzeugt, indem realitätsnahe Konferenzmitschnitte über die Konferenzspinne aufgenommen und durch den Separierungsalgorithmus bearbeitet wurden.

Aus Sicht der Versuchsteilnehmer ergab sich im Versuch damit ein durchaus realistisches Szenario: Sie sollten als „stille Teilnehmer“ den Konferenzen folgen, um danach Auskunft darüber zu geben, welcher Sprecher bestimmte Aussagen getroffen hatte und wie sicher sie sich ihrer Erinnerung waren. Des Weiteren gaben sie an, wie sehr sie sich während der Konferenz konzentrieren mussten und wie leicht es ihnen fiel, zu erkennen, wer in jedem Moment sprach.

Die Ergebnisse (Teil ①) zeigen meist einen Vorteil des LDV-Systems gegenüber dem Mono-System. Allerdings fällt dieser je nach Indikator unterschiedlich stark aus. Die *subjektiven* Indikatoren zeigen, dass 3D-Telekonferenzen gefühlt etwas weniger Konzentration erfordern und dass es deutlich leichter fällt, zu erkennen, wer gerade spricht. Letzterer Unterschied ist zwar nicht unbedingt überraschend aber sehr stark ausgeprägt. Die Ergebnisse des *objektiven* Erinnerungstests, also der Frage, wer bestimmte Aussagen getroffen hat, sind allerdings nicht so eindeutig. Denn auch wenn sich die Versuchspersonen im 3D-Fall ihrer Erinnerung tendenziell sicherer sind, so zeigen sie doch objektiv eine ähnlich gute Leistung. Nur bei mehreren Sprechern (4) ergeben sich dann auch in den Erinnerungstests signifikant bessere Ergebnisse durch binaurale Wiedergabe.

Im zweiten Versuchsteil (②) sollten erschwerte Bedingungen mit höherer Grundbelastung geschaffen werden, indem eine zusätzliche parallele Reaktionsaufgabe bearbeitet werden musste. Der Durchführbarkeit wegen, konnten hier allerdings nur kürzere Konferenzen mit wenigen Sprechern verwendet werden, so dass die Prämisse einer *insgesamt* höheren Belastung im Nachhinein nicht mehr unbedingt gilt. Nichtsdestotrotz gibt Teil ② Aufschluss über die psychische Belastung in einer *Multitasking*-Situation. Hier zeigt sich fast durchwegs ein signifikanter Vorteil des LDV-Systems. Nur die aufgewendete

Konzentration wird ähnlich hoch angegeben, was möglicherweise darauf zurück zu führen ist, dass diese Wahrnehmung von der Sekundäraufgabe dominiert wird.

Als Ergebnis dieser Arbeit kann festgehalten werden, dass beim 3D-Telekonferenzsystem des LDV deutlich leichter zu erkennen ist, wer gerade spricht und dass allgemein weniger Konzentration aufgewendet werden muss, um der Konferenz zu folgen. Das System kann also, soweit es die subjektiven Indikatoren betrifft, als *effizienter* betrachtet werden.

Auch eine bessere Gedächtnisleistung ist teilweise gegeben. Die Ergebnisse legen nahe, dass das 3D-Telekonferenzsystem, insbesondere bei mehr als drei Sprechern oder in Multitasking-Situationen, somit auch zu einer *effektiveren* Kommunikation führt.

Abbildungsverzeichnis

2.1 Anatomie des menschlichen Ohrs 9
2.2 Die Cochlea 10
2.3 Kopfbezogenes Kugelkoordinatensystem 11
2.4 HRTF Beispiel 12
2.5 KEMAR 14
2.6 Modelle der selektiven Aufmerksamkeit 16

3.1 Am LDV entwickelte Konferenzspinne mit acht Mikrofonen 20
3.2 LDV-Telekonferenzsystem 21

4.1 Menschliche Informationsverarbeitung 27

5.1 ITU-Skala 31
5.2 Skala aus Hörgerätetechnik 32
5.3 SEA-Skala 33
5.4 Skala zu Konferenzen 33
5.5 MCH-Skala 34
5.6 NASA-TLX 38

6.1 Aufnahme über Konferenzspinne 42
6.2 Aufnahme-Equipment 43
6.3 Photo vom Versuch 44
6.4 Sprecherpositionen 45
6.5 Screenshot: Inhaltsfrage 45
6.6 Screenshot: Skalen 46
6.7 Screenshot: Reaktionstest 47
6.8 Versuchsanleitung 50

7.1 ① Erinnerung 52
7.2 ① Gewissheit 53
7.3 ① Skala: Konzentration 54
7.4 ① Skala: Sprecher 55
7.5 ② a)Erinnerung b) Gewissheit 56
7.6 ② a) Skala: Konzentration b) Skala: Sprecher 57
7.7 ② Reaktionszeit 58

Abkürzungsverzeichnis

CLT Cognitive Load Theory

CPE Cocktailparty-Effekt

HRTF Head Related Transfer Function

HRIR Head Related Impulse Response

LDV Lehrstuhl für Datenverarbeitung, an der Technischen Universität München

TK Telekonferenz

VP Versuchsperson

Dank

Natürlich war diese Arbeit nur möglich weil *viele andere* vor und neben mir, am und mit dem Telekonferenzsystem am LDV gearbeitet haben. Ungeachtet dessen möchte ich mich bei ein paar Personen für bestimmte Beiträge zu meiner Arbeit bedanken:

Zuerst einmal bei Janto für die Vorlagen, dann bei Thomas mit dem ich zusammen die Aufnahmen durchgeführt habe, bei Benedikt, für das Lösen von Bugs und seinen SVN-Server und bei Andrea und Korbi für die Korrektur dieser Arbeit. Nicht zuletzt natürlich bei Martin, dafür dass er als Betreuer immer vor Ort und ansprechbar war.