DAS LABOR FÜR KOGNITIVE SYSTEME AN DER BTU COTTBUS-SENFTENBERG

Christian Richter, Jens Lindemann, Ronald Römer, Matthias Wolff

Brandenburgische Technische Universität Cottbus-Senftenberg
christian.richter@b-tu.de

Kurzfassung: Der Beitrag soll einen Überblick über das Forschungslabor für kognitive Systeme (*Cognitive Systems Lab*, CSL) geben, das von unserer Professur seit dem Jahre 2011 aufgebaut wird. Das Labor befindet sich in einem akustisch ausgestatteten Spezialraum, dessen Nachhallzeit durch abnehmbare Absorberplatten variabel gestaltet werden kann. Eine zusätzliche schalldichte Kabine innerhalb des Raums kann wahlweise als Aufnahme- oder Technikraum zum Einsatz kommen. Das CSL ist für die Entwicklung von intuitiv bedienbaren, adaptiven und selbstlernenden technischen Systemen ausgelegt und steht neben unseren Mitarbeitern und Studenten auch Kooperationspartnern zur Verfügung. Im Beitrag beschreiben wir hauptsächlich die technische Ausstattung und die Möglichkeiten des Labors. Auf aktuelle Arbeiten im Labor gehen wir am Rande ein.

1 Übersicht

Das Labor bietet vier Arbeitsbereiche:

- ein Versuchsfeld zur Simulation und zum Test kognitiver Benutzerschnittstellen mit einem dreidimensionalen 64-kanaligen Mikrofonfeld, einem berührungsempfindlichen 70-Zoll-Monitor, HD- und 3D-Kameras sowie Spezial- und Studiomikrofonen,
- ein aktiver Ultraschallmessplatz,
- ein Audiomessplatz mit einem 4-kanaligen Messmikrofonfeld sowie
- ein Sprecherarbeitsplatz mit Studiomikrofon und –kopfhörer in einer schalldichten Kabine.

Aktuell stehen verschiedene Versuche auf dem Gebiet der Mensch-Maschine-Interaktion im Zentrum der Arbeit im CSL. Eine besondere Rolle spielen dabei Experimente nach der Wizard-of-Oz-Methode, für welche das Labor durch die räumliche Teilung und die technische Infrastruktur besonders geeignet ist. Wir untersuchen dabei unter anderem Aspekte des „Design-for-All“-Prinzips, welches zentrales Element des BMBF-Förderschwerpunkts „Mensch-Technik-Interaktion im demografischen Wandel“ ist. Dafür bietet das Labor die Möglichkeit einer multimodalen System-Benutzer-Interaktion, zum Beispiel in Form von akustischen (Spracherkennung), visuellen (Gestensteuerung, Gebärdenspracherkennung) oder haptischen (Touch-Gesten) Eingaben sowie visuellen (Bildschirm, Lichtzeichen) oder akustischen (Sprachsynthese, Earcons) Ausgaben. Außerdem erlaubt das System eine automatische Benutzeridentifikation (z. B. Personen- und Sprecherkennung), damit neben der Raumüberwachung auch Zugriffsberechtigungen für das System und individuelle Bedürfnisse der Anwender adäquat berücksichtigt werden können.

2 Versuchsfeld für kognitive Benutzerschnittstellen

2.1 Arbeitsplätze und technische Ausstattung

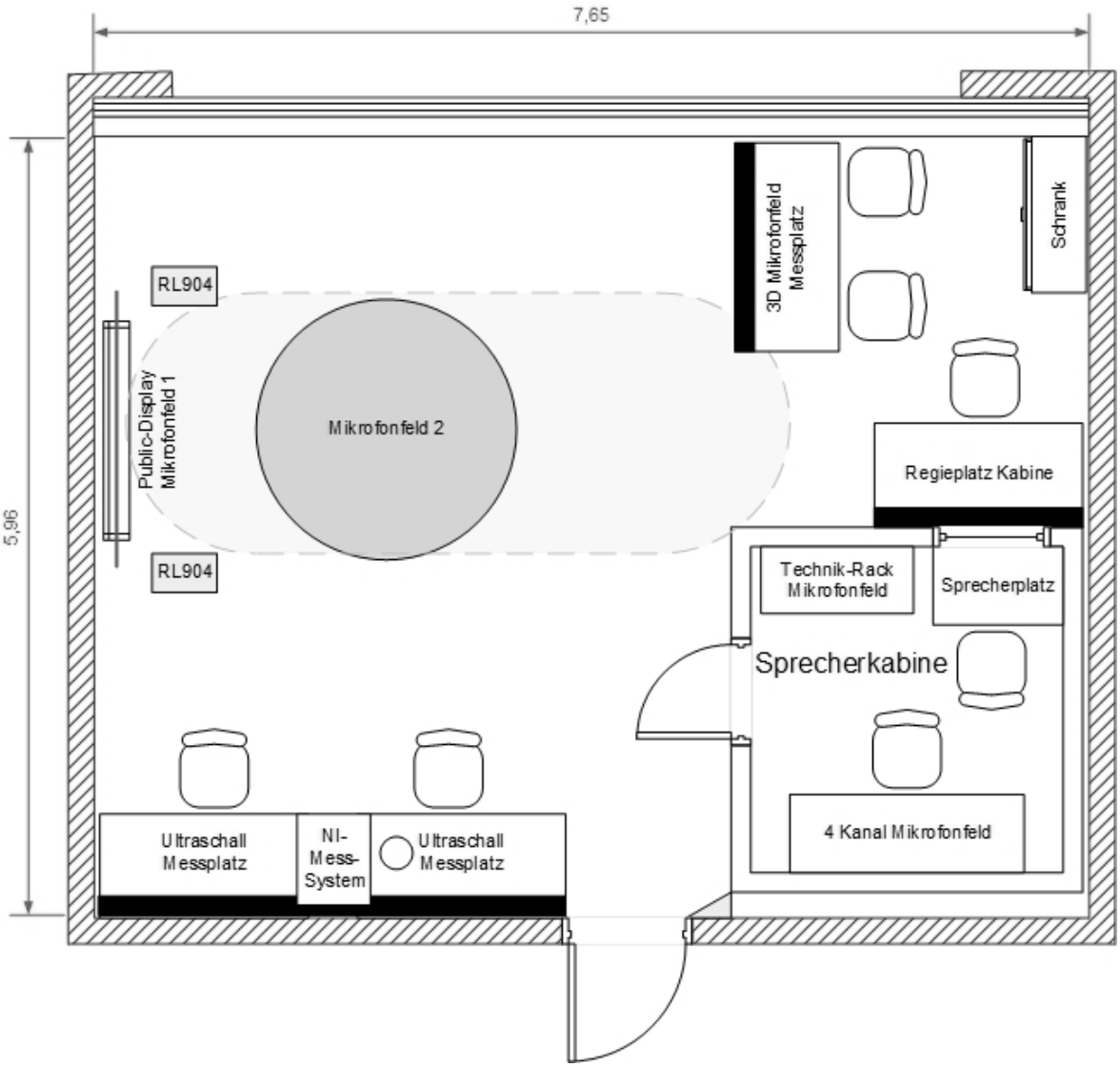

Bild 1: Grundriss des Labors für kognitive Systeme.

Der Grundriss in Bild 1 zeigt die Gliederung des Labors in mehrere Arbeitsbereiche. Im äußeren Bereich des Labors befinden sich neben den Ultraschallmessplätzen, auf die in Abschnitt 3 näher eingegangen wird, auch der Regieplatz sowie die Messfeldsteuerung. Ausgehend vom Regieplatz erfolgt die Überwachung und Administration der Versuche im inneren Laborbereich (Sprecherkabine).

Die 3D-Messfeldsteuerung besteht aus 2 Workstations mit Touchscreen und spezieller Benutzeroberfläche im „Star Trek“-Design. Die Oberfläche ist dem Library Computer Access and Retrieval System (LCARS) nachempfunden. Hier erfolgt die Überwachung sämtlicher Versuchsabläufe über ein Lokal Area Network. Das Public-Display dient in erster Linie den Probanden als zentrale Interaktionskomponente. Auch hier ist die LCARS-Oberfläche implementiert.

Das wichtigste Element des Versuchsfelds bildet ein dreidimensionales 64-kanaliges Mikrofonfeld mit 2 Anordnungen von je 32 Messmikrofonen vom Typ Beyerdynamik MM1. Das horizontale Feld (Spiralanordnung) ist auf einer Laufkatze montiert. Die Mikrofone des vertikalen Felds sind um den zentralen Monitor angeordnet.

Die Mikrofone haben folgende Eigenschaften:

- Linearer Frequenzgang bei einem Übertragungsbereich von 20 Hz bis 20 kHz,
- Halbkugelcharakteristik im eingebauten Zustand,
- Übertragungsfaktor 15 mV/Pa und maximaler Schalldruckpegel 128 dB.

Weitere Ausstattungsmerkmale sind:

- Achtfacher OctoPre Mikrofonverstärker,
- RME MADI 68–fach-A/D-Umsetzer,
- Audio-Datenakquise mit RME Hammerfall HDSP MADI,
- LED-Statusanzeige je Schallquelle (Pegel variabel RGB),
- Mikrofonfeld-Positioniereinrichtung mittels Schrittmotor und Lasersensor,
- 70“ Multitouch-Public-Display und LCARS-Benutzeroberfläche,
- Rohrrichtmikrofon Super CMIT 2 U SCHOEPS mit hoher Richtwirkung durch rein digitale Vorverarbeitung sowie RME AES-32 für Audio-Datenakquise,
- 2 Lautsprecher ME Geithain RL-904 (Regielautsprecher, geeignet für Audio, Video- und Filmstudios, maximaler Schalldruckpegel 106 dB / r=1 m, Übertragungsbereich 40 Hz bis 20 kHz ± 3 dB),
- Industriekamera IDE UI-5240CP-C-HQ (1.3 Megapixel CMOS-Sensor mit hoher Lichtempfindlichkeit, 50 Bilder/s, und Ethernet-Datenübertragung),
- 3D-Kamera vom Typ Kinect zur Erkennung von Gesten und drei Workstations mit Touch-Display.

2.2 Projekte und Anwendungen

2.2.1 Akustische Quellenortung

Eine Form der akustischen Quellenortung ist die Sprecherlokalisation. Ziel ist es, die genauen Raumkoordinaten des Sprechers zu ermitteln, um den Fokus für eventuelle Interaktionen unabhängig von der sonstigen akustischen Umgebung genau auf diesen Punkt auszurichten. Zum Zweck der Schalldetektion kommt das mehrkanalige Mikrofonfeld zum Einsatz. Aus den Abständen der Schallquelle (Sprecher) zu den einzelnen Mikrofonen ergeben sich unterschiedliche Laufzeiten für jeden Schallempfänger. Eine geeignete zeitverzögerte und gewichtete Überlagerung der Mikrofonsignale bewirkt eine Fokussierung des Mikrofonfelds auf einen beliebigen Raumpunkt. Die Mikrofonfeldsteuerung im Labor für kognitive Systeme tastet den Raum nach einer bestimmten Strategie ab [Bir14], ermittelt die Position der energiereichsten Schallquelle und fokussiert auf diese. Die große Anzahl von Mikrofonen ist hierbei eine besondere Herausforderung. Zukünftig sollen eine Sprachsignalerkennung sowie eine Unterstützung der Lokalisierung durch Kameras (siehe Abschnitt 2.2.2) integriert werden.

2.2.2 Optische Experimente

Als optische Sensoren für das kognitive System stehen uns im Labor derzeit eine Industrie- sowie eine Kinect-Kamera zu Verfügung, welche in die Umrandung des Public-Displays fest integriert sind. Diese dienen primär zur Informationsgewinnung für ein intelligentes, sprachverarbeitendes System.

Die Kinect-Kamera erzeugt mittels eines Projektors ein definiertes Punktmuster im Infrarotbereich. Damit ist es möglich, Tiefeninformationen zu den erfassten Objekten zu gewinnen. Für diesen Zweck ist die Kinect-Kamera neben einem klassischen Sensor für sichtbares Licht zusätzlich mit einem Infrarot-Sensor ausgestattet. Eine laufende Studienarbeit beschäftigt sich mit der Positionsbestimmung von Personen, welche aktuell im Labor anwesend sind. Diese zusätzlichen Informationen können zum Beispiel auch für die Optimierung der vorhandenen Sprecherlokalisation [Bir14] durch das System verwendet werden. Weiterhin ist es möglich, dass Personen, welche mit dem intelligenten System interagieren, Gesten verwenden. Erste Untersuchungen auf dem Gebiet der Gestenerkennung wurden hierfür bereits in einer Masterarbeit [Wer14] durchgeführt. Um gehörlosen Menschen bei der Nutzung kognitiver Systeme die Verwendung ihrer natürlichen Kommunikationsform zu ermöglichen, ist auch die Analyse von Gebärdenspracheingaben nötig. Aktuelle Systeme, welche eine visuelle Gebärdensprachenerkennung durchführen, arbeiten nach dem Prinzip der Lautsprachanalyse [Lin14], [Coo11]. Die zu analysierenden gebärdensprachlichen Zeichen werden mit den Händen artikuliert und durch bedeutungstragende nichtmanuelle Parameter, wie Gesichtsausdruck, Mundbild und Blickrichtung ergänzt.

2.2.3 Simulation und Test kognitiver Benutzerschnittstellen

Mit ihrer Fähigkeit zum Wahrnehmen ihrer Umgebung (über Sensoren), selbständigem Lernen, Entscheiden und Handeln (über Aktuatoren) ermöglichen kognitive Systeme unter anderem eine situationsgerechte Unterstützung des Menschen insbesondere bei unsicherer Information. Der routinierte Umgang mit derartigen Fähigkeiten ist insbesondere bei der menschlichen Kommunikation zu beobachten. Dabei kann speziell die sprachliche Kommunikation hervorgehoben werden, diese dient vor allem dem Austausch und dem Abgleich von Vorstellungen. Zu diesem Zweck muss der Mensch also in der Lage sein, zwischen den Gedanken der Kommunikationsteilnehmer zu vermitteln. Für die Mensch-Maschine-Kommunikation folgt daraus einerseits, dass auch die Maschine über Vorstellungen oder "Maschinengedanken" verfügen muss, die zu den Vorstellungen des menschlichen Kommunikationspartners kompatibel sein müssen. Andererseits können derartige Gedanken nur physikalisch - also über Signale - übertragen werden.

Bei der Sprachkommunikation treten eine Reihe von Problemen in Verbindung mit dem akustischen Übertragungskanal auf, so dass für den praktischen Einsatz anspruchsvolle Algorithmen zur Kompensation von störenden Einflüssen nötig sind. Um dies zu vermeiden und damit den wissenschaftlichen Fokus stärker auf die kommunikativen Aspekte zu richten, können einerseits Headset-Mikrofone zum Einsatz kommen, andererseits kann eine freie und natürliche Kommunikation durch Mikronfonfelder ermöglicht werden. Letzteres hat für uns ein höheres Gewicht, da die Erforschung von bedeutungsgetriebener Kommunikation nicht durch technische Randbedingungen eingeschränkt werden soll.

3 Audio- und Ultraschallmesssystem

3.1 Ultraschall-Messplatz

Für Messaufgaben zur zerstörungsfreien Prüfung im Ultraschallbereich dient unter anderem der EFRE-geförderte (Programm: Geräteinvestition für zukunftsfähige Hochschulen), nachfolgend beschriebene Laborplatz zur kognitiven Signalverarbeitung.

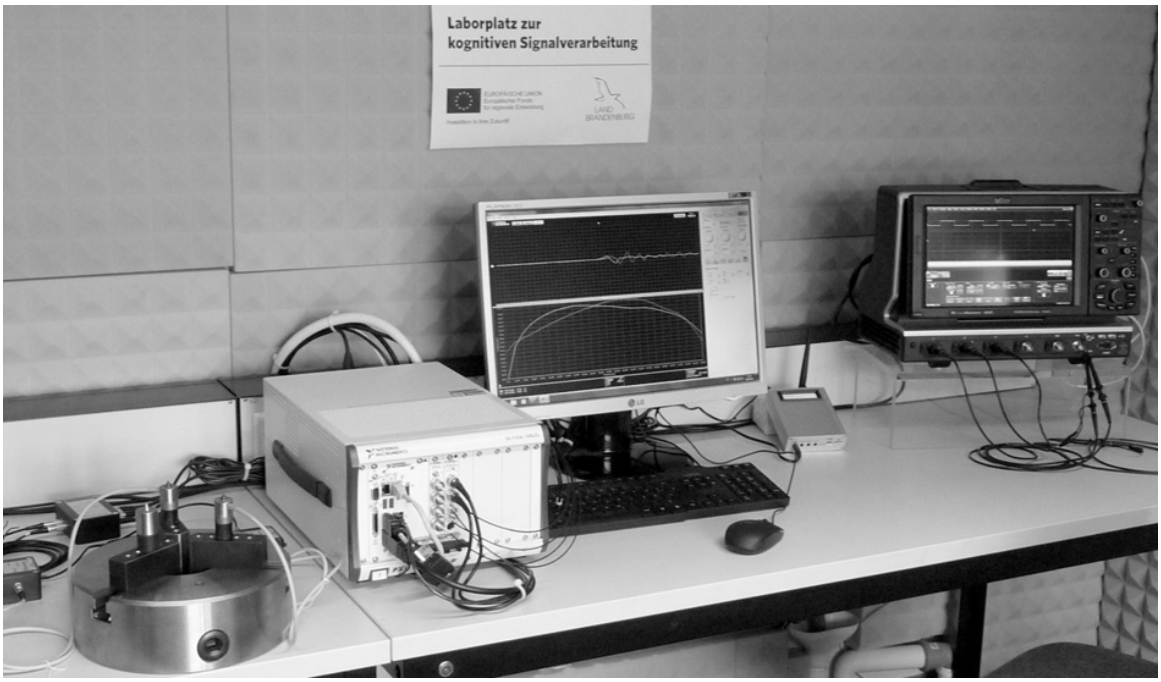

Bild 2: Messwert-Erfassungssystem von National Instruments und Oszilloskop von Teledyne LeCroy.

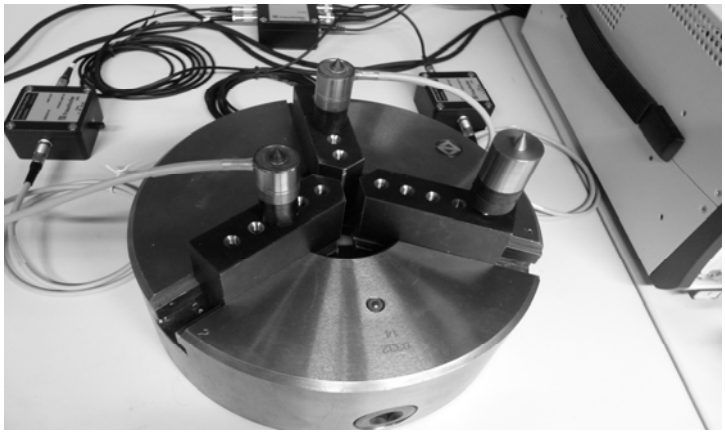

Bild 3: 3-Punkt-Aufnahme und Ultraschallsensoren.

Das Bild 3 zeigt den Messtisch mit Körperschallsensoren auf der Basis von Piezo-Wandler-Technologie. Die Sensoren besitzen einen Arbeitsbereich (Sensitivity) von 10-1000 kHz. Befestigt werden die Sensoren auf einer speziell vom (IKTS) Fraunhofer-Institut für Keramische Technologien und Systeme gefertigten 3-Punkt-Aufnahme. Die Anpassung an das Datenerfassungssystem von National Instruments erfolgt über einen Empfangsverstärker je Sensor und einen Konverter zum Anschluss an den NI PXIe-5122 Digitizer. Der Digitizer ermöglicht eine Datenakquise von 100 MS/s bei 14 Bit Verarbeitungsbreite. Die Signalverarbeitung und akustische Mustererkennung der aufgenommenen Ultraschallsignale erfolgt mit eigener Software im NI-Messsystem.

3.2 Vierkanal-Messmikrofonfeld

Als akustische Empfänger werden Kondensator-Messmikrofonkapseln MK 301 E der Firma Microtech Gefell eingesetzt. Die 1/4" Messmikrofonkapsel MK 301 E ist für akustische Messungen in der Forschung konzipiert und zeichnet sich entsprechend der Herstellerangaben durch eine hohe zeitliche Konstanz der elektroakustischen Parameter aus. Der Frequenzbereich beträgt 5 Hz bis 100 kHz, bei einem möglichen Schalldruckpegel von 158 dB.

Das Bild 4 zeigt die Anordnung von Messmikrofonen wahlweise im Stativ oder in einer Lochplatte. Letztere ermöglicht die genaue Reproduktion von Messanordnungen.

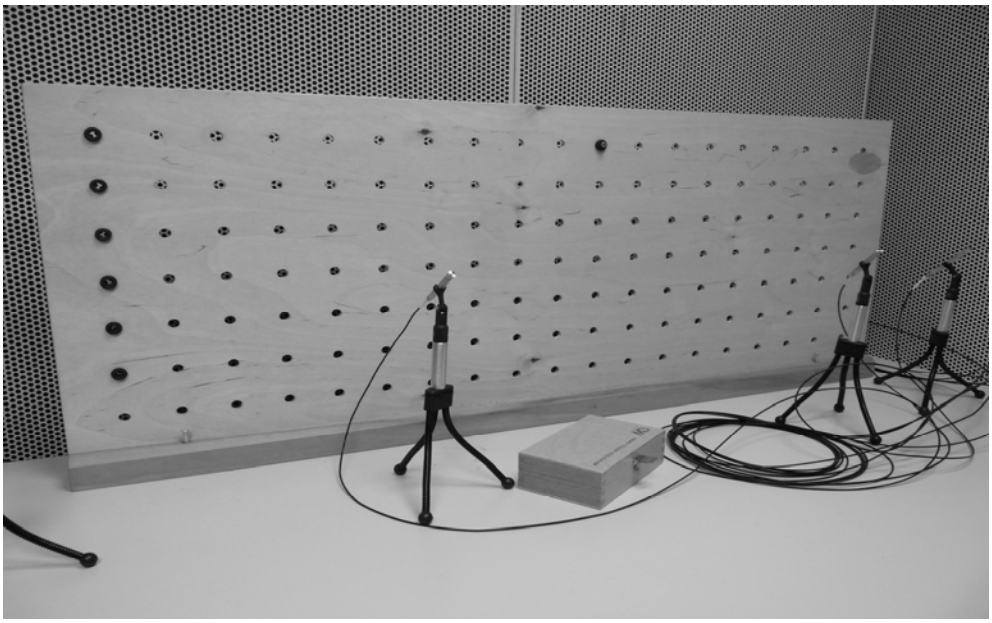

Bild 4: 4-kanaliges Messmikrofonfeld.

Als Backend dient das bereits unter Abschnitt 3.1 erwähnte Messwert-Erfassungssystem von NI. Mit Hilfe der 4 Eingänge des Dynamiksignalanalysators NI PXI-4462 können die Messwerte erfasst werden. Die IEPE-Konditionierung erfolgt gleichermaßen über den Dynamiksignalanalysator. Der NI PXI-4462 besitzt darüber hinaus folgende Parameter:

6 Verstärkerparameter für Eingangsbereiche von ±316 mV bis ±42,4 V, 4 simultan abgetastete Analogeingänge, bis zu 204,8 kS/s, flexible Antialiasing-Filter, A/D-Wandler mit 24 Bit Auflösung und 118 dB Dynamikbereich.

3.3 Projekte und Anwendungen

Das Akustikmesssystem wird für die Forschung zur akustischen Mustererkennung für technische Schallsignale sowie zur kognitiven Signalverarbeitung eingesetzt [Wol11, Abschn. 6.4]. Zurzeit finden Experimente im Rahmen eines BMBF-geförderten Projekts zur Entwicklung eines miniaturisierten Systems zur intelligenten Signalverarbeitung statt.

4 Sprecherarbeitsplatz in einer schalldichten Kabine

4.1 Technische Ausstattung

Der Sprecherarbeitsplatz dient sowohl zur Herstellung von Audioaufnahmen als auch zur Durchführung von Hörexperimenten. Dank der Auskleidung mit akustischen Dämmstoffen (Hersteller: Telekom Innovation Laboratories) ist die Kabine reflexionsarm und vom sonstigen Laborraum akustisch getrennt. Damit besteht auch die Möglichkeit geräuschintensive Geräte während der Messung im Außenlabor innerhalb dieser Kabine unterzubringen. Der Arbeitsplatz ist ausgestattet mit einem hochwertigen Studiomikrofon sowie Audiometrie-Kopfhörern. Der Aufbau von weiteren Experimenten kann variabel gestaltet werden.

Das Kondensatormikrofon der Firma Neumann vom Typ TLM 170 R ist ein umschaltbares Großmembranmikrofon (Kugel, Niere, Acht, Hyperniere und breite Niere), das sich durch seine linearen Übertragungseigenschaften auszeichnet. Der Übertragungsbereich bewegt sich von 20 Hz bis 20 kHz bei einem Geräuschpegelabstand von 74 dB. Es wird auch in einer großen Vielfalt von Anwendungsfällen bei Rundfunk, Film und Fernsehen eingesetzt und eignet sich hervorragend für HQ-Sprachaufnahmen.

Der Audiometrie-Kopfhörer Sennheiser HDA 200 ist ein geschlossener, dynamischer Kopfhörer, speziell für den Einsatz an Hochton-Audiometern. Er weist eine hohe passive Außengeräuschdämmung auf. Das frequenzabhängige Verhalten von 20 Hz bis 20 kHz ist annähernd linear. Die Reproduzierbarkeit der Messergebnisse ist entsprechend Herstellerangaben sehr gut.

Als weitere Signalquelle dient ein R&S UPL Audio Analyzer. Dessen Signalgenerator weist eine nur sehr geringe Verzerrung auf.

4.2 Projekte und Anwendungen

Die Produktion von hochwertigen Tonaufnahmen ist eine der wichtigsten Tätigkeiten, die innerhalb der Sprecherkabine vollzogen wird. Bei der Durchführung von Studienarbeiten, Laboren und Forschungsprojekten wird diese Möglichkeit genutzt. Besonders vorteilhaft erweist sich dabei der Sichtkontakt zum Regieplatz.

Die Kabine erlaubt unter anderem die Evaluation von Hörbeispielen unter Studiobedingungen. Während der Darbietung nimmt der Proband die subjektive Bewertung von Hörbeispielen vor. Dies kann ein Benotungstest wie zum Beispiel der Mean-Opinion-Score-Test (MOS) sein [Ric12]. Es können auch bestimmte Hörphänomene wie zum Beinspiel die akustische Verdeckung präsentiert werden.

Im Rahmen eines Laborversuchs im Lehrmodul Sprachverarbeitung werden hier audiometrische Messungen durchgeführt. Die geräuscharme Kabine eignet sich besonders für Ruhehörschwellenmessung mit Hilfe des genannten Audiometrie-Kopfhörers und des Audioanalysators. Die Abbildung der Frequenz in Abhängigkeit vom Schalldruckpegel wird automatisiert betrieben. Im Ergebnis erhält man die typische Ruhehörschwelle für jedes getestete Ohr.

Die Sprecherkabine wird weiterhin für Wizard-of-Oz-Experimente verwendet, bei denen der Versuchsleiter in der Kabine und für die Versuchsperson unsichtbar die Funktion eines technischen Systems simuliert [Kar12], [Foe14] und [Sto14].

Bild 5: Sprecherkabine mit Sprecher- bzw. Wizard-Arbeitsplatz und Fenster zum Regieplatz.

5 Literatur

[Bir14] M. Birth: Sprecherlokalisierung mit einem 3D-Mikrofonfeld. In R. Hoffmann (editor), Elektronische Sprachsignalverarbeitung 2014. ESSV 2014, Volume 71 of Studientexte zur Sprachkommunikation, pages 233–238, Dresden, Germany, März 2014. TUDpress, Dresden. ISBN: 978-3-944331-51-5.

[Wer14] T. Werner: HMM-basiertes Analyse-Synthese-System zur Verarbeitung von 3D-Trajektorien. Masterarbeit; Cottbus, 2014.

[Lin14] J. Lindemann: Semantische Verarbeitung von Gebärdensprache in intelligenten hierarchischen Sprachdialogsystemen. In R. Hoffmann (editor), Elektronische Sprachsignalverarbeitung 2014. ESSV 2014, volume 71 of Studientexte zur Sprachkommunikation, pages 118–125, Dresden, Germany, März 2014.

[Coo11] H. Cooper; B. Holt; R. Bowden: Sign Language Recognition. In: Visual Analysis of Humans, S. 539–562: Springer London; London, 2011.

[Ric12] C. Richter, T. Holfeld: Gestörte akustische Umgebungen und ihr Einfluss auf netzbasierte auditive Testverfahren. In M. Wolff (editor), Elektronische Sprachsignalverarbeitung 2012. ESSV 2012, volume 64 of Studientexte zur Sprachkommunikation, pages 275–282, Cottbus, Germany, Aug. 2012. TUDpress, Dresden. ISBN: 978-3-942710-81-7.

[Kar12]K. Karnagel, R. Römer, S. Rogge, J. Lindemann: Semantische Modellierung von System-Benutzer-Interaktionen am Beispiel einer Home-Entertainment Steuerung. In M. Wolff (editor), Elektronische Sprachsignalverarbeitung 2012. ESSV 2012, volume 64 of Studientexte zur Sprachkommunikation, pages 306–313, Cottbus, Germany, Aug. 2012. TUDpress, Dresden. ISBN: 978-3-942710-81-7.

[Wol11] M. Wolff: Akustische Mustererkennung. Dresden, TUDpress, 2011.

[Foe14] L. Förster: Rückweisung ungrammatischer Spracherkenner-Eingaben, Cottbus, 2014.

[Sto14] M. Stöhr: Wizard-of-Oz-Methode - Sprachsteuerung einer Musikbibliothek, Cottbus, 2014.