Technische Universität Berlin

Institut für Sprache und Kommunikation

Fachgebiet Audiokommunikation und –technologie

D-10587 Berlin

- Synchronitätsfenster und Sprachkompetenz -

Perzeptive Wahrnehmung von Bild und Ton in Abhängigkeit von Sprachkenntnissen

Masterarbeit

vorgelegt von:

Sabine Buckbesch

[redacted]

Erstgutachter: Prof. Dr. Stefan Weinzierl
Zweitgutachter: Dr. Hans-Joachim Maempel

23.April 2013

Technische Universität Berlin

FAKULTÄT I
GEISTESWISSENSCHAFTEN

Postanschrift: TU Berlin Fakultät I - Prüfungsausschuss, Sekr. FR 3-6
Franklinstr. 28/29, 10587 Berlin

Studiendekan
und
Vorsitzender der
Prüfungsausschüsse

Prof. Dr. phil. W. König

Tel.: (030) 314-24304
Fax: (030) 314-29396
E-Mail:
verena.salomo@tu-berlin.de

Eidesstattliche Erklärung

Ist jeder an der TU Berlin verfassten schriftlichen Arbeit eigenhändig unterzeichnet beizufügen!

Hiermit erkläre ich an Eides statt gegenüber der Fakultät I der Technischen Universität Berlin, dass die vorliegende, dieser Erklärung angefügte Arbeit selbstständig und nur unter Zuhilfenahme der im Literaturverzeichnis genannten Quellen und Hilfsmittel angefertigt wurde. Alle Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind kenntlich gemacht. Ich reiche die Arbeit erstmals als Prüfungsleistung ein.

Titel der schriftlichen Arbeit

_____

_____

VerfasserIn/VerfasserInnen*

Name Vorname Matr.-Nr.

_____

Betreuende/r DozentIn

Name Vorname

_____

Mit meiner Unterschrift bestätige ich, dass ich über fachübliche Zitierregeln unterrichtet worden bin und verstanden habe. Die im betroffenen Fachgebiet üblichen Zitiervorschriften sind eingehalten worden. Eine Überprüfung der Arbeit auf Plagiate mithilfe elektronischer Hilfsmittel darf vorgenommen werden.

_____

Ort, Datum Unterschrift**

*Bei Gruppenarbeiten sind die Unterschriften aller VerfasserInnen erforderlich.
**Durch die Unterschrift bürgen Sie für den vollumfänglichen Inhalt der Endversion dieser schriftlichen Arbeit.

Zusammenfassung

Die vorliegende Masterarbeit untersucht die Wahrnehmung von audiovisueller Asynchronität bzw. Synchronität in Abhängigkeit von der Sprachkompetenz des Zuschauers. Der alltägliche Einsatz von digitalen Medien und die internationale Informationsverbreitung in den unterschiedlichsten Sprachen, machen auf einen bisher zu weniger erforschten Aspekt der Sprachwahrnehmung aufmerksam. In früheren Untersuchungen (Navarra, 2010; Heide, 2009; van Wassenhove, 2007; Rudloff, 1997) wurden durch Verhaltensforschung oder bildgebende Verfahren Erkenntnisse über die Wahrnehmung von audiovisuellen Sprachsignalen und an der Verarbeitung beteiligte Gehirnregionen gewonnen.

Diese Arbeit strebt an eine Brücke zwischen den verschiedenen Ansätzen der Linguistik, Neurowissenschaft und empirischen Forschung der audiovisuellen Wahrnehmung zu schlagen. Gleichzeitig wird mit dem Einsatz von drei Sprachen ein Aspekt der Sprachsignalverarbeitung eingebunden, welcher bisher nur wenig erforscht wurde. In Anlehnung an die Arbeit von Frau Heide (2009) untersucht der im Rahmen dieser Masterarbeit durchgeführte Hörversuch das Erkennen von Asynchronität für die Sprachen Deutsch, Englisch und Persisch. 30 deutsche Muttersprachler bekamen die Aufgabe in einem Forced-Choice-Verfahren von jeweils paarweise dargebotene Videos die physikalisch synchrone Version erkennen. Die zeitlichen Latenzen bewegten sich dabei zwischen einer Tonvoranstellung von +120ms bis hin zu einer Tonverzögerung von -120ms in einem Intervall von 40ms.

Die Ergebnisse wiesen keine signifikanten Unterschiede im Bezug auf die Lage des Punktes subjektiver Synchronität (PSS) und des eben wahrnehmbaren Unterschieds (JND) auf. Die Ergebnisse von Heide (2009) konnten bestätigt werden und eine Verbreiterung des fremdsprachigen Synchronitätsfensters im Vergleich zum deutschen nachgewiesen werden. Die gewonnenen Daten konnten allerdings nur Tendenzen in den Randbereichen von +120ms und -120ms bestätigen. Dies lässt die Schlussfolgerung zu, dass in einer Fremdsprache auch bei stärkerer Bild-Ton-Latenz eine Integration audiovisueller Information erfolgt. Die fehlenden Signifikanzwerte beim PSS und JND lassen vermuten, dass die dargebotene Sprache und die damit verbundene Sprachkompetenz des Rezipienten keinen Einfluss auf die Wahrnehmung von audiovisueller Synchronität hat.

Danksagung

Ich möchte mich zu erst für die inhaltliche Unterstützung von Prof. Dr. Stefan Weinzierl und Dr. Hans-Joachim Maempel bedanken. Hervorzuheben ist die bereitwillige Hilfe von Martina Vrhovnik, Anita Goldswain und Mina Fallahi, die als jeweilige Muttersprachler maßgeblich an der Erstellung des audiovisuellen Materials beteiligt waren und ihr Einverständnis zur Verwendung für diese Masterarbeit gegeben haben. Stefan Schütz gilt mein herzlich Dank für die Erstellung des Programmes „Eiusdem Temporis“, welches als Benutzeroberfläche während des Hörversuchs gedient hat. Für ihre kontinuierliche Hilfestellung bei der statistischen Datenauswertung möchte ich mich bei Dr. Astrid Bartels bedanken. Und abschließend geht mein Dank an alle Teilnehmern, die an diesem Hörversuch teilgenommen haben und ohne die diese Masterarbeit nicht möglich gewesen wäre.

Inhalt

1. Einleitung und Aufgabenstellung 1
1.1 Synchronität 2
1.2 Hypothesen 6
2. Grundlagen 8
2.1 Grundlagen der audiovisuellen Signalverarbeitung 8
2.1.1 visuelle Signalverarbeitung 8
2.1.2 Auditive Signalverarbeitung 13
2.1.3 Zeitlicher Ablauf 17
2.2 Psycholinguistische Grundlagen 18
2.2.1 Sprachsignalwahrnehmung 18
2.2.2 Mentales Lexikon 21
2.2.3 Fremdsprache 23
2.2.4 Messmethoden 25
2.3 aktueller Forschungsstand 28
3. Experimentelle Versuchsdurchführung 33
3.1 Material 33
3.1.1 Sprachmaterial 33
3.1.2 Videomaterial 38
3.1.3 Probanden 39
3.1.4 Eiusdem Temporis 40
3.2 Versuchsaufbau 42
3.3 Versuchsdurchführung 44
3.4 Analyse 45
4. Ergebnisse 47
4.1 Darstellung der Ergebnisse 47
4.2 Diskussion 64
4.2.1 Diskussion der Hypothesen 64

4.2.2 Kritische Anmerkungen 72
5. Fazit und Ausblick 74
6. Verzeichnisse 76
6.1 Literaturverzeichnis 76
6.1.1 Internetquellen 80
6.1.2 Software 81
6.2 Abbildungsverzeichnis 82
6.3 Tabellenverzeichnis 84
7. Anhang 85
7.1 Farbwahrnehmung 85
7.2 Praat-Analyse 86
7.3 „Eiusdem Temporis" 87
7.4 Schaltbild 88

1. Einleitung und Aufgabenstellung

Seit der Erfindung des Stummfilms zu Beginn des 20.Jahrhunderts und der folgenden Weiterentwicklung zum Tonfilm im Jahre 1926 hat sich das Medium Film immer weiter ausgebreitet. In Form von Kinofilmen, Fernsehen und Internetvideos werden heutzutage täglich Informationen verbreitet, deren vollständige und korrekte Wahrnehmung von der Synchronität zwischen Bild und Ton abhängt. Eine asynchrone Wiedergabe kann leicht zu Korrekturbemühungen im Gehirn führen, wodurch die Konzentration auf den Bild- und/oder Toninhalt gestört wird. Doch genau wie es bei der Bildwiederholungsrate eine Schwelle gibt, ab dem das menschliche Auge kein Flimmern mehr wahrnimmt, existiert eine Bild-Ton-Latenz, die für die synchrone Reizwahrnehmung in einem akzeptablen Rahmen liegt.

Thema dieser Arbeit ist das Erkennen der bereits genannten Bild-Ton-Synchronität in Abhängigkeit von der im Film verwendeten Sprache und der entsprechenden Sprachkompetenz. Aus früheren Untersuchungen (Heide, 2009; Rudloff, 1997) ist bereits die Lage des Zeitintervalls, in dem der Mensch Abweichungen von der physikalischen Synchronität nicht wahrnimmt, bekannt. Dieses Intervall wird Synchronitätsfenster genannt. Die Forschungsfrage ist, ob das Synchronitätsfenster von der Kompetenz einer Sprache abhängig ist und sich somit in Abhängigkeit von dem Kenntnisstandes von Muttersprache und Fremdsprache verändert. Ebenso konnte aus empirischen Forschungen ein Punk subjektiver Synchronität ermittelt werden. Dieser Punkt stellt die zeitliche Latenz dar, zu der es nicht mehr möglich ist zu entscheiden, ob Bild oder Ton zuerst erscheint. In einem ähnlichen Rahmen wurden Versuche durchgeführt, um ein besseres Verständnis für die auditive und visuelle Informationsverarbeitung im Gehirn und die dazugehörigen Gehirnregionen zu erlangen. Entsprechende Untersuchungen werden im Abschnitt 2.3 "aktueller Forschungsstand" Erwähnung finden.

Die vorliegende Masterarbeit kombiniert so Elemente der auditiven und visuellen Wahrnehmung, der Informationsverarbeitung entsprechender Signale und psycholinguistische Modelle zur Spracherkennung miteinander. In den anfänglichen Kapiteln sollen die Basiselemente dieser Arbeit eingeführt werden. Um einen Eindruck vom menschlichen Synchronitätsfenster gewinnen zu können, soll in den folgenden Abschnitten zuerst auf die zwei Seiten der Bild-Ton-Synchronität eingegangen werden:

die Produktion des Materials sowie die Wiedergabe und damit verbundene Wahrnehmung des Materials. Da Bild- und Tonmaterial heutzutage häufigen Datentransfers unterzogen werden und auch bei der Nachbearbeitung und Wiedergabe mehrere Umwandlungen durchlaufen, kommt es zwangsläufig zu zeitlichen Versätzen, welche berücksichtigt werden müssen. Der Unterpunkt "Synchronität" soll daher einen Überblick über die technischen Aspekte während der Filmproduktion geben, während im Kapitel 2 auf die theoretischen Grundlagen der Psycholinguistik und der audiovisuellen Signalverarbeitung eingegangen wird. Um herauszufinden, ob die Sprache einen Einfluss auf die Wahrnehmung von Bild-Ton-Versätzen hat, wurde ein entsprechender Hörversuch durchgeführt, welcher im Kapitel 3 ausführlich beschrieben wird. Die Ergebnisse sind im Kapitel 4 dargestellt und werden im Folgenden diskutiert und interpretiert.

1.1 Synchronität

Um ein Verständnis für Synchronität und ihrem Gegenteil, der Asynchronität, zu entwickeln, soll am Anfang dieser Arbeit kurz der Begriff "synchron" definiert werden. Er setzt sich aus zwei griechischen Wortstämmen zusammen: "syn" (gleich) und "chronos" (Zeit). "Synchron" beschreibt also einen Punkt der Gleichzeitigkeit von Ereignissen. Im Fall von audiovisuellen Materialen bezieht es sich immer auf die Gleichzeitigkeit von Bild und Ton. Neben der physikalischen Synchronität wird auch die psychologische Synchronität unterschieden (Heide, 2009, S.3; nach Pöppel, 2000, S.27). Im Vergleich zur physikalischen Gleichzeitigkeit, bei der Bild und Ton ohne Versatz abgespielt werden, kann es bei der psychologischen Gleichzeitigkeit zu Abweichungen von Bild und Ton kommen, die für den Zuschauer jedoch subjektiv als optimal synchron wahrgenommen werden. Im Rahmen dieser Arbeit liegt die Konzentration allerdings nur auf dem Erkennen der physikalischen Synchronität. Diese zu gewährleisten ist das Hauptaugenmerk während einer Filmproduktion.

Schon bei den Dreharbeiten für einen Film muss auf die Synchronität von Bild- und Tonaufnahmen geachtet werden. Sie müssen nicht nur gleichzeitig sein, sondern auch die gleiche Laufzeitgeschwindigkeit besitzen. Fehler in der ersten Produktionsphase durch unterschiedliche Laufzeiten können später dazu führen, dass während der Post-Produktion Bild und Ton nicht synchronisiert werden können.
Zur Zeit des Stummfilms war es üblich den Film in den Kinos schneller abzuspielen, als er original aufgezeichnet wurde. Ein Grund dafür könnte sein, dass auf diese Weise

der Film schneller zu Ende war, wodurch er einem neuen Publikum früher präsentiert werden konnte und mehr Geld in die Kinokassen kam. Mit der Erfindung des Tonfilms änderte sich dies jedoch, da ein schnelleres Abspielen auch eine deutlich hörbare Pitch-Veränderung zur Folge hatte. Um den Takt zu halten, wurden für Magnetbandaufzeichnung mit paralleler Kamera Synchronmotoren mit einer Framerate von 24 Frames pro Sekunde verwendet. In den 60'er Jahren folgten dann Kamera und Audiorekorder mit Quarzoszillatoren, welche die Synchronität von Bild und Ton ermöglichten. Die Framerate musste in den folgenden Jahren für das Farbfernsehen angepasst werden und lag nun bei 25fps (PAL) bzw. 29,97fps (NTSC) (Holman, 2010, S.124). Da die meisten Filme in den USA jedoch mit 24 Frames aufgezeichnet werden, ergibt sich ein Problem für die Umwandung in das 30 Frame TV-Format. Um ein zu schnelles Abspielen und eine Pitcherhöhung zu vermeiden, kommt das 2:3 Pull-up-Verfahren zum Einsatz, bei dem abwechselnd zwei und drei Halbbilder eingefügt bzw. dupliziert werden. Allerdings gilt zu beachten, dass NTSC nicht genau mit 30 Frames sondern 29,97 Frames pro Sekunde läuft. Der Film wird um das verbleibende 0,1 % langsamer abgespielt, wodurch Synchronisationsprobleme entstehen können. In Europa wird ein Film entweder mit 25 Frames aufgezeichnet, sodass keine Umwandlung für das Fernsehen erfolgen muss, oder 24fps werden 4 % schneller abgespielt, um in einer Sekunde 25 Frames statt 24 abzubilden. Um die Synchronität zu wahren, muss in diesem Fall allerdings auch die Audiospur um 4 % schneller abgespielt werden, sodass Charakterstimmen, Geräusche und Hintergrundmusik eine Anhebung der Grundfrequenzen erfahren. Dies müsste in Nachbearbeitungen korrigiert werden (Holman, 2010, S.127).

Die genannten Beispiele sind nur zwei von unterschiedlichen SMPTE-Formaten. Der SMPTE-Code (Society of Motion Picture and Television Engineers) wurde für das Editieren von Filmen entwickelt und beinhaltet verschiedene Standards für die Zeitcodierung. Allgemein wird jedes Bild mit einem eindeutigen Zeitcode versehen, welcher im Format hh:mm:ss:ff festgehalten wird. H steht hierbei für "Stunde", m für "Minute", s für "Sekunde" und f für "Frame". Die Anzahl der Frames kann sich unterscheiden und liegt je nach Anwendung zwischen 23,976fps und 30fps. Der Timecode kann entweder längs in einem speziellen Timecode-Track aufgezeichnet werden (LTC - Longitudinal Time Code) oder vertikal direkt in die Videospur geschrieben werden (VITC - Vertical Intercal Time Code) (Heide, 2009, S.6-7). Für die Bearbeitung wird ein neuer Timecode erstellt, da durch das Editieren Lücken im Original entstehen können, welche für ein folgendes Lesegerät, einem Synchronizer,

nicht zu überbrücken sind. Eine Übersicht über die heute genutzten Frameraten, Timecodes und Abtastraten ist in der nachfolgen Tabelle zu finden:

Frame Rate	Where used	SMPTE Time Code Frame Counter	Audio sample rate, Hz
23,976	24p video	0-23, then 0	48000
24	Film 24p video alternate	0-23, then 0	48000 or 48048
25	Film or video for PAL television	0-24, then 0	48000
29,97 NDF	NTSC Video, some production	0-29, then 0	48000
29,97 DF	NTSC video, some production, and network delivery masters	0-29, then 0, dropping frame numbers at the start of each minute except for minutes 0, 10, 20, 30, 40, 50	48000
30 NDF	On production sound recorders for film to be posted NDF	0-29, then 0	48048
30 DF	On production sound recorders for long-form television and film projects	0-29, then 0, dropping frame numbers at the start of each minute except for minutes 0, 10, 20, 30, 40, 50	48048

Tabelle 1: Frameraten für Film und Video mit Timecodes und Sample Raten (Holman, 2010, S.142)

Während Videos eine standardisierte Framerate besitzen, besteht bei professionellen, digitalen Videoformaten ein direkter Zusammenhang zwischen der Framerate und der Audioabtastrate. Sollte dies nicht der Fall sein, so bieten die meisten Bearbeitungsprogramme eine Funktion zum Synchronisieren von Audio- und Videospur durch internes Konvertieren der Abtastrate (Holman, 2010, S.132-134).

Das Gegenteil zur Bild-Ton-Synchronität bildet die Asynchronität. Sie entsteht durch einen zeitlichen Versatz zwischen dem visuellen und dem auditiven Signal. Dieser Versatz kann sich durch eine Verzögerung oder einem verfrühten Einsatz der Tonspur im Verhältnis zur Bildspur äußern. In einem gewissen Rahmen ist das Vorhandensein einer zeitlichen Abweichung für den Menschen jedoch nicht bemerkbar. Dieses Intervall wird Synchronitätsfenster (synchrony window) genannt. Obwohl hier ein Versatz zwischen Bild- und Tonspur vorliegt, ist er dem Zuschauer nicht immer ersichtlich. Eine Verzögerung des Tons um -40ms (1 Frame im PAL-Format) wird von vielen Menschen sogar eher als synchron aufgefasst, als die tatsächlich physikalisch synchrone Fassung (Heide, 2009).

Eine Sicherstellung der synchronen Aufnahme garantiert nicht, dass Bild und Ton auch gleichzeitig am Hörerort ankommen. Ein zeitlicher Versatz kann ebenfalls bei der Umwandlung von analog zu digital, durch Komprimierungsverfahren, bei der Bearbeitung an Mischpulten, durch das Puffern von Daten an den Schnittstellen und bei der Encodierung entstehen. Abweichungen durch die Kopplung von Mischpulten können in den meisten Fällen durch ein Taktsignal, welches alle Geräte im Studio synchronisiert, entgegengewirkt werden (Heide, 2009, S.7). Eine Verzögerung bei der Encodierung hängt jedoch direkt von der Datenrate, Taktfrequenz und Kanalkonfiguration ab und liegt bei Dolby Digital etwa zwischen 179ms und 450ms (Slavik et al., 2008, S.629).

Codec	Bandbreite	Datenrate	Mono/Stereo	fs	Latenz
MPEG 1 Layer 2	8-20kHz	64-384kBit/s	m/s	48kHz	100ms
MPEG 1 Layer 3	15-20kHz	64-320kBit/s	m/s	48kHz	>300ms
MPEG 2 Layer 2	11kHz	64kBit/s	m	24kHz	>150ms
MPEG 2 AAC	8-20kHz	32-128kBit/s	m/s	diverse	>400ms

Tabelle 2: MPEG Audiocodecs im Überblick (Friedrich, 2008, S.196)

Eine detaillierte Beschreibung und Übersicht zur Übertragungstechnik im Allgemeinen und Audiocodecs im Besonderen können in den Werken von Slavik et. al. (2008), Kapitel 10 oder Friedrich (2008), Kapitel 11 gefunden werden.

Die Wahrnehmung von Asynchronität ist zusätzlich von der Informationsverarbeitung im Gehirn abhängig. So ist zu beachten, dass für die Verarbeitung von akustischen Signalen wie Sprache, Musik oder Geräuschen unterschiedlich viel Zeit benötigt wird, und auch das Einbeziehen der visuellen Information berücksichtigt werden muss. Auf die genauen Grundlagen der Sprachverarbeitung soll im Kapitel 2 "Grundlagen" ausführlich eingegangen werden. An dieser Stelle soll erwähnt werden, dass empirische Untersuchungen der Wahrnehmung von audiovisueller Synchronität das Vorhandensein eines Synchronitätsfensters bereits nachgewiesen haben (Rudloff, 1997; Heide, 2009). Dieses Synchronitätsfenster ist der Grund dafür, warum der Mensch audiovisuelle Informationen mit einem zeitlichen Versatz als synchron wahrnimmt, solange dieser Versatz in dem durch Versuche grob definierten Intervall von 200ms liegt. Ob und wie sich dieses Intervall in Abhängigkeit von den Sprachkenntnissen des Hörers verändert, ist Thema dieser Arbeit und soll im nächsten Unterpunkt in Hypothesen gefasst werden.

1.2 Hypothesen

Die bisherige Einleitung soll einen Überblick über die Fragestellung der Masterarbeit geben, welche im folgenden Abschnitt durch Grundlagen in der Psycholinguistik und der audiovisuellen Signalverarbeitung ausgebaut wird. Weiterhin dienen die allgemeinen Informationen als Basis für die hier aufgestellten Hypothesen und den im Kapitel 3 beschriebenen Versuch.

1. H0 Hypothese:

   Sprachkenntnisse haben keinen Einfluss auf die Breite des Synchronitätsfensters.

1. H1 Hypothese:

   Sprachkenntnisse haben einen Einfluss auf die Breite des Synchronitätsfensters.

In dieser Arbeit wird unter Verwendung von dreisprachigem Videomaterial das Erkennen von Synchronität verglichen. Das Videomaterial liegt dafür in den Sprachen Deutsch, Englisch und Persisch vor, wobei Deutsch die Muttersprache der Versuchsteilnehmer, Englisch die erste Fremdsprache und Persisch eine unvertraute Fremdsprache ist. Ähnliche Untersuchungen zum Erkennen von Synchronität wurden bereits durchgeführt (Exner, 1875; Dixon et al., 1980; Rudloff, 1997; Heide, 2009), doch bestand hier das Audiomaterial aus Geräusch, Musik und/oder der Muttersprache der Teilnehmer. Ein Vergleich von vertrauter und unvertrauter Sprache wurde von Navarra et. al 2010 durchgeführt und soll in einem späteren Abschnitt näher beschrieben werden. Die erste H1-Hypothese kann dann verworfen werden, wenn die Versuchsteilnehmer das synchrone Video unabhängig von der Videosprache bei unterschiedlichen Zeitversätzen, besonders aber in den Randbereichen von +120ms und -120ms, gleichermaßen gut erkennen. Im Gegensatz dazu wird die H1-Hypothese angenommen, wenn die Teilnehmer die synchrone Version zu bestimmten Zeitversätzen in einer Fremdsprache besser oder schlechter erkennen als in ihrer Muttersprache.

2. H0 Hypothese:

   Sprachkenntnisse haben keinen Einfluss auf die Lage des PSS.

2. H1 Hypothese:

Sprachkenntnisse haben einen Einfluss auf die Lage des PSS.

Untersuchungen zur Lage des Punkts subjektiver Synchronität (PSS = Point of subjective synchrony) wurden in den vergangenen Jahren mehrfach durchgeführt, unter anderem auch von Heide (2009) mit unterschiedlichen Versuchsdesigns. In der vorliegenden Arbeit soll ebenfalls die Lage des PSS untersucht und auf Unterschiede zwischen den Sprachen Deutsch, Englisch und Persisch überprüft werden. Die Lage des PSS zeichnet sich durch eine schlechtere Bewertung der physikalisch synchronen Fassung im Vergleich zu einer zeitlich versetzten Fassung aus und liegt nach dem Herstellungsversuch von Frau Heide für die Sequenz Sprache bei ca. -40ms (Heide, 2009, S.41). Die zweite H0-Hypothese kann dann verworfen werden, wenn sich die Lage des PSS in einer Fremdsprache signifikant von der in der Muttersprache unterscheidet.

3. H0 Hypothese:

Die Wahrnehmung von Synchronität ist für eine unvertraute Fremdsprache und Musik ist unterschiedlich.

3. H1 Hypothese:

Die Wahrnehmung von Synchronität ist für eine unvertraute Fremdsprache und Musik identisch.

Die dritte und letzte Hypothese basiert auf der Annahme, dass für eine unvertraute Fremdsprache kein mentales Lexikon zur Verfügung steht und eine Übersetzung daher nicht möglich ist. Die Informationsverarbeitung würde keinen lexikalischen Zugriff erfordern, wodurch die Vermutung entsteht, dass eine unvertraute Fremdsprache statt als Sprache als Musik wahrgenommen wird. In diesem Fall würde die Lage des Synchronitätsfensters für eine unvertraute Fremdsprache mit der für Musik übereinstimmen. Da im Hörversuch für diese Arbeit nur Sprachbeispiele und keine Musik genutzt wurden, stützt sich der Vergleich auf die Ergebnisse aus der Magisterarbeit von Frau Berta Luise Heide aus dem Jahr 2009. Die dritte H0-Hypothese kann dann angenommen werden, wenn sich die Bewertungen für Persisch signifikant von den Ergebnissen aus dem Versuch von Frau Heide, 2009, unterscheiden.

2. Grundlagen

Kapitel 2 dieser Masterarbeit gibt eine Einführung in die audiovisuelle Informationsverarbeitung. Dabei werden sowohl die Grundlagen der audiovisuellen Signalverarbeitung beschrieben, als auch wichtige Aspekte der Psycholinguistik aufgezeigt. Da sich diese Arbeit auf die Wahrnehmung von Bild-Ton-Synchronität konzentriert, werden im weiteren Verlauf nur zwei der fünf Sinne berücksichtigt. Um ein Verständnis für die Forschungsfrage zu entwickeln, bietet dieses Kapitel neben einem Überblick über die Anatomie des menschlichen Auges und Ohres auch eine Zusammenfassung über den aktuellen Stand internationaler Forschungen. Alle hier aufgeführten Informationen können in einschlägigen Standardwerken der Biologie, Anatomie und kognitiver Neurophysiologie gefunden werden.

2.1 Grundlagen der audiovisuellen Signalverarbeitung

In den folgenden Abschnitten 2.1.1 und 2.1.2 wird auf die Verarbeitung von visuellen und auditiven Informationen eingegangen. Dabei soll sowohl ein Überblick über die anatomischen Grundlagen als auch die physiologischen Aspekte gegeben werden. Der Unterpunkt 2.1.3 vergleicht abschließend den zeitlichen Ablauf der visuellen und auditiven Verarbeitungsprozesse.

2.1.1 visuelle Signalverarbeitung

Der Mensch sieht nicht mit den Augen, sondern mit dem Gehirn. Es sind jedoch die Augen, die dem Gehirn die Informationen liefern, aus denen eine dreidimensionale Abbildung der Umgebung erstellt werden.

Grob kann das Sehen als Umsetzung von Licht in elektrischen Signalen in der Netzhaut beschrieben werden, welche an die zentrale Sehbahn weitergeleitet werden. Dabei liegt das sichtbare Licht im Bereich von 400nm bis 750nm, welches die Farben Blau bis Rot abdeckt. Kürzere (ultraviolette) und längere (infrarote) Wellenlängen sind für den Menschen unsichtbar. Über mehrere brechende Medien erreicht ein Lichtstrahl die Netzhaut und bildet hier ein umgedrehtes und verkleinertes Bild der Umgebung ab.

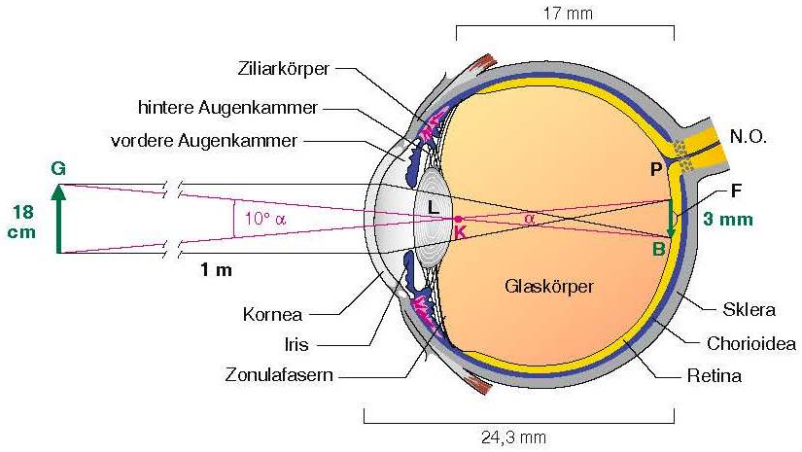

Abbildung 1: Aufbau des menschlichen Auges (Schmidt et al., 2006, S.244)

Um eine gute Sehleistung zu erhalten, ist die genaue Fokussierung der Lichtstrahlen auf die Netzhaut nötig. Das Scharfsehen erfolgt beim menschlichen Auge durch Änderung der Linsenbrechkraft. Bevor das Licht jedoch die Linse erreicht, trifft es auf die von der Iris umschlossenen Pupille. Die Pupille hilft mit ihrer Größenanpassung den Lichtdurchlass zu regulieren. Eine Leuchtdichtezunahme führt daher zu einer Verkleinerung der Pupille, um Blenden zu verhindern. Dieser Reflex ist der schnellste Mechanismus zur Helligkeitsanpassung, da er nur 0,5s-1s dauert und den Lichteinfall in diesem Zeitraum rund 25fach verringern kann (Schmidt et al., 2006, S.243-251).

Für die Bewegung des Auges sind sechs Muskeln verantwortlich, welche horizontale, vertikale und rollende Bewegungen ermöglichen. Ebenfalls gibt es verschiedene Augenbewegungsarten, zu denen unter anderem Sakkaden und Folgebewegungen gehören, welche Objekte im Zentrum des schärfsten Sehens halten und jeweils spezifischen neuronalen Kontrollsystemen unterliegen. Als Sakkaden werden sprungartige, bewusste oder unbewusste Augenbewegungen zur Fixation bezeichnet. Sakkaden sind zum Beispiel typisch beim Lesen. Vom visuellen Auslöser bis zum Beginn einer Sakkade beträgt die Latenz ca.200ms für normale Sakkaden und 70ms für Expressakkaden mit zielgerichteter Aufmerksamkeit. Sakkaden verfügen jedoch über nur geringe Sehschärfe. Folgebewegungen im Gegensatz sind bewusste, konjugierte Augenbewegungen, welche kleine bewegte Ziele im Fokus halten und mit 100ms-150ms eine geringere Latenz haben als Sakkaden. Wechseln sich langsame Folgebewegungen und Sakkaden ab, wie zum Beispiel bei einem Blick aus einem fahrenden Zug, so wird diese Sequenz optokinetischer Nystagmus genannt. Ausführlichere Informationen zu den genannten und weiteren Augenbewegungen

können im Werk von Robert F. Schmidt und Hans-Georg Schaible, 2006, S.253-255 gefunden werden.

Auf dem hinteren Teil der Retina befinden sich ungleichmäßig verteilt die Sinnesrezeptoren des Auges, Stäbchen und Zapfen. Der Mensch hat im Durchschnitt etwa 120 Millionen Stäbchen und 6 Millionen Zapfen. Die Fotorezeptoren geben ihre Informationen über Bipolarzellen an etwa eine Million Ganglienzellen weiter, welche den Übergang zur zentralen Sehbahn bilden. Die Stäbchen sind für das Sehen in der Dämmerung zuständig, während die Zapfen beim Farbsehen und im Tageslicht zum Einsatz kommen und entsprechend ihrer Empfindlichkeitsmaxima in drei Typen (Blau-Zapfen bei 440nm, Grünzapfen bei 535nm und Rotzapfen bei 565nm) eingeteilt sind.
Im Dunkeln sind die Fotorezeptoren depolarisiert, doch bei Lichteinfall werden eingelagerte Fotopigmente umgewandelt und es kommt zu einer Hyperpolarisation. Die Depolarisation wird erst durch die Weitergabe an das nächste Neuron ausgelöst. Im Vergleich zu den Stäbchen reagieren die Zapfen schneller, aber mit geringerer Empfindlichkeit. Da die Stäbchen ab den Zapfenbipolarzellen den gleichen Signalweg verwenden, führt die Erregung des Zapfensystems bei Tageslicht zur Hemmung der Stäbchenantwort, welche im Gegenzug in der Dämmerung die Zapfenantwort unterdrücken.

Die Signalverarbeitung funktioniert in der Netzhaut durch postsynaptische Potenziale, wie die durch die Ganglienzellen codierte Erregung und Hemmung. Diese äußern sich durch Zu- und Abnahme von Aktionspotenzialfrequenzen. Nachdem über die Ganglienzellen die Informationen gesammelt und codiert wurden, laufen die Nervenfasern im optischen Nerv (N.opticus) bis zur Sehnervenkreuzung (Chiasma opticum), wo sich die nasalen Fasern kreuzen. So wird die rechte Gesichtshälfte in der linken Hemisphäre repräsentiert und die linke Gesichtshälfte in der rechten Hemisphäre. Im Gegensatz dazu wird der innere Bereich des zentralen Gesichtsfeldes beidseitig weitergeleitet. Hinter der genannten Sehnervenkreuzung verlaufen gekreuzte und ungekreuzte Fasern als Tractus opticus bis zum Corpus geniculatum laterale im Thalamus. Dieser stellt die Schaltstation der Sehbahn zwischen der Retina und dem visuellen Cortex da. Durch eine monosynaptische Übertragung gelangen die Informationen der Sehnervenfasern auf die Schaltzellen und in die primäre Sehrinde. In der neueren Literatur wird dieser Bereich mit V1 gekennzeichnet und auch die weiteren visuellen Hirnareale in hierarchischer Form mit V2-V5, wobei V5 auch als MT bezeichnet wird.

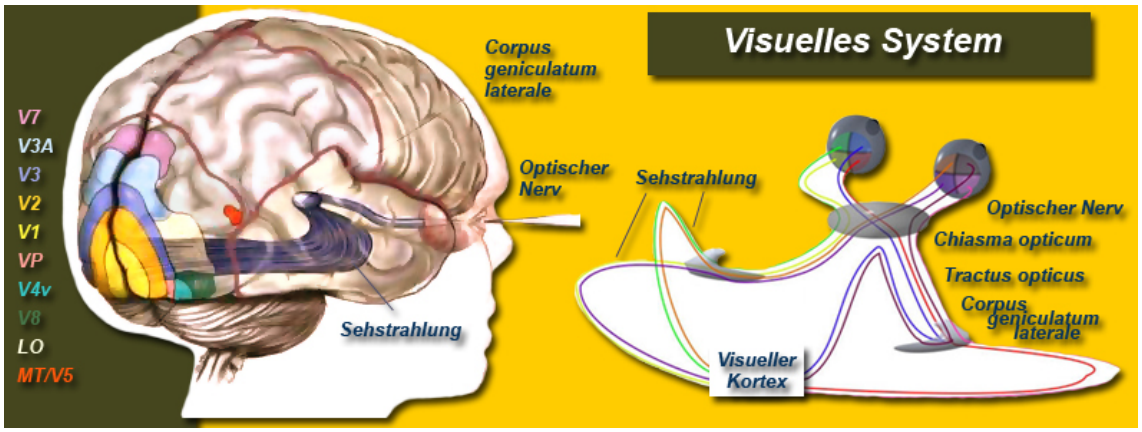

Abbildung 2: visuelles Netzwerk (Lehmann, 2009)

Die Projektion in die Sehrinde unterliegt dabei bestimmten Abbildungsregeln. Dazu gehört, dass benachbarte Orte auf der Netzhaut auch im Gehirn benachbart abgebildet werden und dass die Größe der Projektion proportional zur retinalen Ganglienzellendichte auf der Netzhaut ist. Somit wird der Bereich um die Fovea vergrößert und der Bereich der Netzhautperipherie verkleinert abgebildet. Die Nervenzellen der primären Sehrinde reagieren allerdings nicht wie die rezeptiven Felder der Retina auf Lichtreize, sondern auf komplexe Muster oder bestimmte Anordnung und aktivieren jeweils orientierungsspezifische Cortexneuronen. Diese Neuronen ermöglichen die folgenden drei Fähigkeiten (Ward, 2010, S. 106-107).

1. *Orientierung eines Reizes im Raum*
2. *Bestimmung der Richtung einer Reizbewegung*
3. *Erkennen von Ecken und Kanten*

Beginnend in der Retina werden durch zwei Hauptganglienzellklassen Informationen parallel verarbeitet. Während das großzellige M-System für Kontrast- und Bewegungssehen zuständig ist und nur 10 % der Ganglienzellenanzahl umfasst, nimmt das kleinzellige P-System 80 % ein und ist für die hohe räumliche Auflösung, Formanalyse und das Farbsehen verantwortlich. Beide Systeme projizieren getrennt über den Thalamus zur primären Sehrinde. Eine dritte heterogene Gruppe enthält die Ursprungszellen des K-Systems (10 %) und schickt Informationen sowohl über den Thalamus als auch Fasern ins Mittelhirn. In der bereits genannten Schaltstation werden sechs verschiedene Schichten abhängig vom System angeregt. So reagieren Schichten 1-2 auf das M-System, während die Schichten 3-6 auf Informationen vom P-System der Netzhaut reagieren. Dabei werden die typischen Eigenschaften aus der Retina beibehalten.

Die Parallelverarbeitung des M- und des P-Systems wird im primären visuellen Cortex (V1) fortgesetzt. Hier werden Farb-, Form- und Bewegungsreize in folgenden Schichten und Cortexarealen in getrennten Substrukturen repräsentiert. Das P-System ist neben der räumlichen Auflösung auch für die Wahrnehmung von Form und Farbe zuständig. Im V1 wird die Verarbeitung von Form und Farbe getrennt, wodurch im V2 Farbe getrennt von Bewegung und Form abgebildet wird und im V4 fortgesetzt wird. V4 wird allgemein als Farbzentrum des menschlichen Gehirns angesehen, da Verletzungen in diesem Bereich zur Behinderung des Farbsehens führen und eine Wahrnehmung nur noch in Graustufen vollzogen wird (Ward, 2010, S.111). Die Formanalyse lässt sich im V2 finden und anschließend in Einzelzellen im inferotemporalen Cortex (IT). Hier werden Formdetails wie Gesichter oder Hände verarbeitet. Im Gegensatz zum P-System wird das M-System von V1 in V2 nach V3, dem mediotemporalen Cortex (MT) und den medial superior temporal (MST) weitergeleitet. Diese Zellen reagieren auf Bewegung und räumliche Tiefe und steuern damit die visuelle Aufmerksamkeit. Weitere Informationen zum Farbsehen sind in den Werken von Schmidt et al., 2006, S. 282-285 und Ward, 2010, S. 111 zu finden und im Anhang ebenfalls kurz zusammengefasst.

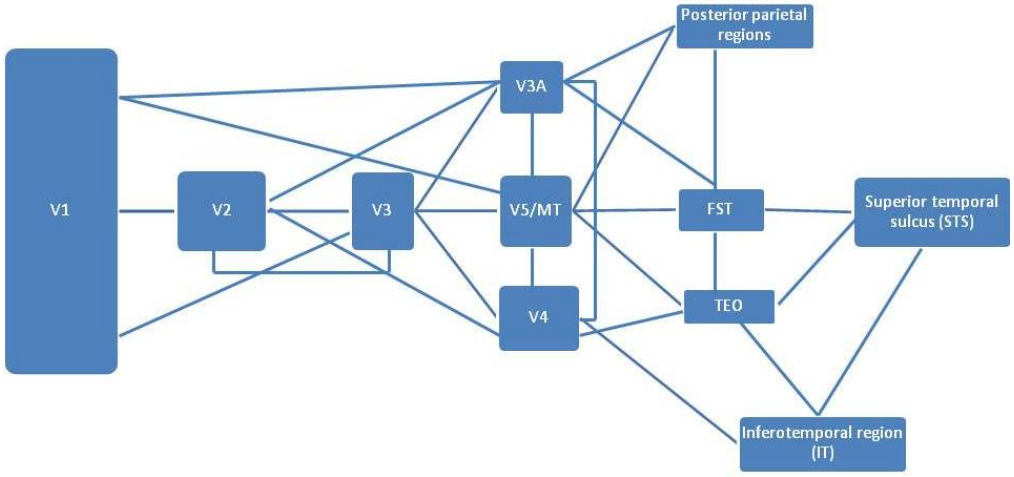

Abbildung 3: visuelle Informationsverarbeitung (Ward, 2010, S.110)

An dieser Stelle ist festzuhalten, dass die gewonnenen Informationen von der Netzhaut des Auges bis zum V1 gemeinsam verarbeitet und weitergeleitet werden und erst anschließend je nach Informationsart (Farbe, Bewegung, Identität etc.) in speziellen Arealen getrennt analysiert werden. Um aus visuellen Informationen eine sinnvolle Bedeutung zu erschließen, ist ab diesem Punkt die Integration von vorhandenem Wissen über Orte, Personen oder Objekte nötig. David Marr stellte 1976 die „Theory of vision“ auf, der zufolge die visuelle Informationsverarbeitung ähnlich der

computergestützten Datenverarbeitung funktioniert und in vier Stufen unterteilt ist: das Erkennen von Ecken, Längen, Kontrasten und Orientierung, dem Hinzufügen von Oberflächenstrukturen und räumlicher Tiefe, einem Vergleich mit einem dreidimensionalen Abbild aus dem Gedächtnis und dem Hinzufügen von semantischen Informationen wie Namen und Eigenschaften des erkannten Objektes. Zur zweiten Phase zählt auch das Gruppieren von Objekten, welches durch den Neuropsychologen Gestalt in den 30'er Jahren des 20. Jahrhunderts in den fünf Gestalt-Prinzipien beschrieben wurde:

1. *Das Gesetz der Nähe*
2. *Das Gesetz der Ähnlichkeit*
3. *Das Gesetz des guten Verlaufs*
4. *Das Gesetz der Geschlossenheit*
5. *Das Gesetz des gemeinsamen Schicksals*

Wie aus den Grafiken (Abbildung 2 und 3) zu erkennen ist, werden die visuellen Informationen gemeinsam bis zum Sehnerv geleitet und dort in unterschiedlichen Arealen in Abhängigkeit von der Information weiterverarbeitet. Nachdem im folgenden die auditive Signalverarbeitung betrachtet wird, soll abschließend der jeweilige zeitliche Ablauf gegenübergestellt werden.

2.1.2 Auditive Signalverarbeitung

Nachdem die visuelle Signalübertragung im vorherigen Abschnitt beschrieben wurde, soll nun die auditive Verarbeitung betrachtet werden. Vibrationen von Objekten erzeugen den für den Menschen wahrnehmbaren auditiven Reiz in Form einer Schallwelle. Die Schallübertragung zum Ohr erfolgt durch Druckschwankungen in der Luft, welche durch das Ausdehnen und Zusammenpressen von Luftmolekülen erzeugt werden. Der Mensch kann dabei Druckänderungen von $2*10^{-5}$ Pascal bis über 100 Pascal wahrnehmen. Die Frequenz dieser Druckschwankungen nehmen wir als Geräusche war. Während ein Ton eine Sinusschwingung mit nur einer Frequenz beschreibt und musikalische Klänge aus einem Grundton und mehreren Obertönen besteht, decken Geräusche fast allen Frequenzen des menschlichen Hörbereichs ab.

Abhängig vom Schalldruck wird ein bestimmter Ton mit einer subjektiven Lautstärke wahrgenommen. Die empfundene Lautstärke steigt dabei mit Erhöhung des Schalldrucks. Durch eine Veränderung der Frequenz ändert sich bei gleichem Schalldruck demzufolge auch die Wahrnehmung der Lautstärke. Das menschliche

Gehör ist in dem Frequenzbereich von 2 kHz bis 5 kHz empfindlicher als bei höher und niederfrequenteren Tönen, wodurch Schallereignisse lauter wahrgenommen werden und die Hörschwelle in diesem Bereich am niedrigsten ist. Um alle Frequenzen gleich laut zu hören, ist eine Anpassung des Schalldrucks nötig. Wird der Schalldruck über die Frequenz aufgetragen, erhält man die Kurve gleicher Lautstärke (Schmidt et al., 2006, S. 289-290). Der Mensch hört in einem Bereich von 20 Hz bis etwa 16 kHz und einem Lautstärkepegel von 4 bis 130phon. Der Hörsprachbereich deckt dabei Frequenzen und Lautstärken im gesamten Hörbereich ab. Eine Hörstörung in diesem Bereich würde dazu führen, dass Gesprochenes leiser oder nur mit eingeschränktem Sprachverständnis wahrgenommen wird.

Der Schall erreicht vom äußeren Gehörgang das Trommelfell, wo die Gehörknöchelchen Hammer, Ambos und Steigbügel die niedrigere Luftimpedanz an die höhere Impedanz der Innenohrflüssigkeit anpassen. Dies ist durch eine kleinere Fläche der Steigbügelfußplatte am ovalen Fenster im Bezug zum Trommelfell und der Hebelwirkung der Gehörknöchelchen möglich. Etwa 60 % der Schallenergie gelangen ins Innenohr, welches aus dem Vestibularapparat für den Gleichgewichtssinn und der einem Schneckenhaus gleichenden Cochlea für die Hörwahrnehmung besteht. Entlang der Cochlea verläuft die cochleare Trennwand. Trifft ein Schallsignal auf das Trommelfell, so schwingt die Steigbügelfußplatte im Ringband des ovalen Fensters abwechselnd ins Innenohr und zurück ins Mittelohr. Die Lymphflüssigkeit im oberen Teil der Cochlea kann nur durch das Eindrücken der Trennwand ausweichen, wodurch die Bewegung in die untere Hälfte übertragen wird, wo wiederum das Ausweichen durch die Elastizität des runden Fensters möglich ist. Schallwellen sorgen demnach für eine ständige Auf- und Abwärtsbewegung der Trennwand (Schmidt et al., 2006, S.294).

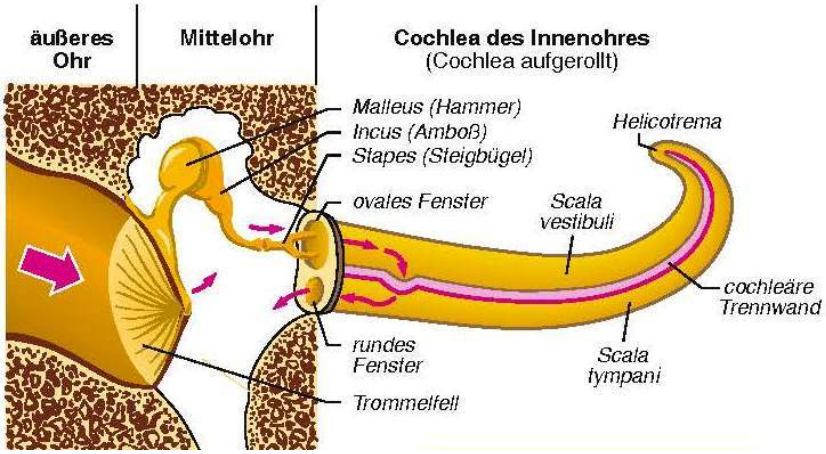

Abbildung 4: Schematische Darstellung des Ohres (Schmidt et al., 2006, S. 292)

An der Unterseite der Trennwand befindet sich die Basilarmembran mit Stützzellen, welche Hörsinnzellen mit Sinneszellen, sogenannte Haarzellen, besitzen. Die Bewegungen der Trennwand bewirkt eine Auslenkung der Sinneszellen, welche den Transduktionsprozess einleitet. Durch Schallwellen entstehende Schwingungen beginnen direkt hinter dem Steigbügel und setzten sich als Wanderwelle bis zur Cochleaspitze fort. Dabei wird ihre Amplitude im Verlauf verstärkt, nur um danach wieder stark abzusinken. Je höher der Ton ist, desto kleiner ist die dazugehörige Wellenlänge und umso näher liegt das Maximum am Steigbügel. Jede Frequenz hat einen bestimmten Ort der Maximalauslenkung und wird aufgrund der Aktivierung einzelner Haarzellen an diesen Ort codiert.

Ein Schallsignal und seine codierte Information erreichen über mehrere hintereinandergeschaltete Neuronen den auditiven Cortex im Temporallappen. In den ersten Neuronen der Hörnervenfasern werden die Entladungsrate (Schalldruckpegel), die Zeitdauer der Aktivierung und durch das Ortsprinzip die Frequenz codiert. Im Gegensatz dazu sind die höheren Neuronen der Hörbahn für komplexere Schallmuster wie die Phoneme, Informationsfilterung und räumliches Hören durch Laufzeitunterschiede ausgerichtet (Schmidt et al., 2006, S.304-305).

Wie in der Grafik (Abbildung 5) zu erkennen ist, teilen sich die aufsteigenden Fasern des Hörnervs im Hirnstamm und ziehen zum Nucleus Cochlearis ventralis und zum Nucleus Cochlearis dorsalis und übertragen das Signal bis zum zentralen Nervensystem. Da einige Neuronen zur Gegenseite kreuzen, entsteht eine Verbindung zwischen beiden Innenohren mit jeweils der linken und der rechten Hörrinde, wodurch es möglich ist, binaurale Schallsignale zu vergleichen. Im Anschluss an Umschaltungen zum Colliculus inferior und Corpus geniculatum mediale ziehen die aufsteigenden Hörstrahlen zur primären Hörrinde.

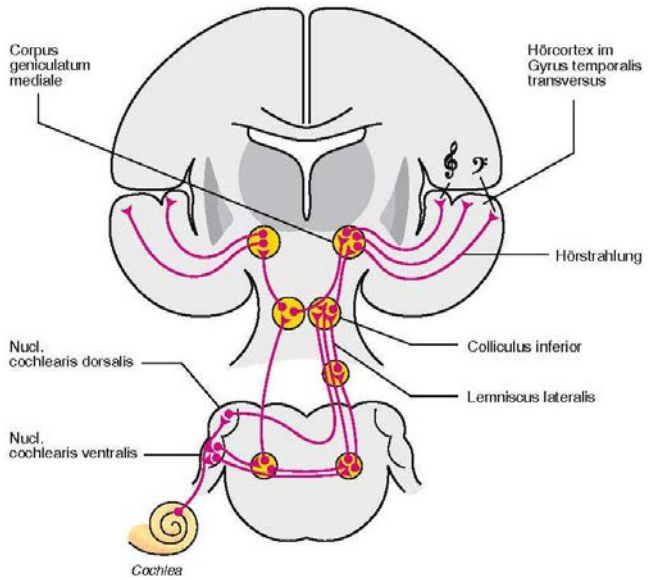

Abbildung 5: Schematische Darstellung der zentralen Hörbahn (Schmidt et al., 2006, S.304)

Anfangs werden die Schallinformationen nur weitergegeben, erst ab den höheren Neuronen werden die enthaltenden Informationen verarbeitet. Dabei wird nur der Nutzschall, welcher die relevanten akustischen Informationen enthält, bis zur Hörrinde transportiert, während nebensächliche Informationen wie Hintergrundgeräusche eliminiert werden. Die höheren Neuronen besitzen dabei die Eigenschaft auf Schallmuster, statt auf Sinustöne zu reagieren (Schmidt et al., 2006, S.308). Die genaue Verarbeitung der Schallinformation ist bisher nur in Ansätzen erforscht. So ist zwar bekannt, dass bestimmte Neuronen im auditiven Cortex auf Beginn, Ende, Mindestzeitdauer oder Frequenzmodulationen spezialisiert sind, aber das Erkennen von Sprachsignalen ist noch nicht geklärt. Der Mensch ist in der Lage Nutzschall von Störschall zu unterscheiden, was zu einem Großteil auf Lernprozesse und zu einem geringeren Anteil auf die Konzentration und den Willen des Hörenden zurückzuführen ist. Die Auswahl eines bestimmten Sprachsignals und das Ignorieren aller störenden Nebengeräusche wird auch als „Cocktail Party Effect“ bezeichnet.

Abhängig vom Schallereignis (Sprache, Musik, Geräusch) und der verbundenen Tätigkeit, können verschiedene Wege zur auditiven Signalverarbeitung genommen werden. So gibt es Anzeichen für eine Tonhöhenverarbeitung in den höheren kortikalen Regionen, welche die psychologische Wahrnehmung von Tonhöhen im Gegensatz zu den physikalischen Eigenschaften ermöglicht (Ward, 2010, S.215; nach Bendor und Wang, 2005).

Es wurde ebenfalls festgestellt, dass es zu einer Unterscheidung zwischen Inhalt und Quelle des akustischen Signals kommt und entsprechend unterschiedliche Neuronen auf das Codieren dieser Informationen spezialisiert sind. Rauschecker und Tian postulierten 2000, dass zwei Wege zur auditiven Signalverarbeitung existieren: ein dorsaler "Wo"-Weg über den Scheitellappen für die Lokalisation und ein ventraler "Was"-Weg entlang des Temporallappens zur Identifikation des Geräusches. Für die menschliche Sprache wird zusätzlich noch ein dorsaler "Wie"-Weg angenommen (Ward, 2010, S.215-216). So wie das Gesicht viele Informationen bei der visuellen Verarbeitung liefert, übermittelt eine Stimme nicht nur inhaltliche Informationen, sondern auch soziale, wie das Geschlecht, die Größe, das Alter und die Stimmung des Sprechers (Scherer, Banse und Wallbott, 2001). Eine genauere Betrachtung der Sprachsignalwahrnehmung folgt im Unterabschnitt 2.2.1.

2.1.3 Zeitlicher Ablauf

Nachdem in den Abschnitten 2.1.1 und 2.1.2 die visuelle und auditive Signalverarbeitung detailliert beschrieben wurde, soll hier noch einmal kurz der zeitliche Ablauf für die Verarbeitung verglichen werden. Während aus physikalischer Sicht das Licht mit ca. $c_L = 3 * 10^8 m/s$ schneller ist als der Schall in der Luft bei ca. 20 °C mit $c_s = 343 m/s$, wird aus physiologischer Sicht das auditive Signal schneller verarbeitet als das äquivalente visuelle Signal (Beck et al., 2004). Bereits 1963 konnten Spreng und Keidel in einem Versuch zeigen, dass auditive Signale zwischen 40-60ms früher als visuelle Reize zu einer Reizantwort führen (Guski, 2006, S.94-95; nach Spreng und Keidel, 1963). Dies kann auf die "lange" Wegstrecke des visuellen Signals von der Retina zum visuellen Cortex und dem vergleichbaren kürzeren Weg des auditiven Reizes vom Trommelfell zum auditiven Cortex zurückgeführt werden. Der visuelle Reiz durchläuft mehrere neurale Schichten und benötigt 50ms-80ms, um bis zur primären Sehrinde zu gelangen und ca. 110ms-150ms, bis es zu einem Erkennen der visuellen Information kommt (Gegenfurtner et al., 2002; Guski, 2006, S.94-95). Für eine synchrone Darbietung von audiovisuellem Material, wie es bei Filmen der Fall ist, muss daher eine Anpassung an die aufgezeigte Verarbeitungsgeschwindigkeit stattfinden.

Untersuchungen von van Wassenhove et al.(2005) konnten nachweisen, dass visuelle Informationen wie Lippenbewegung zu eine Beschleunigung der Verarbeitung führen. Dies beruht auf der Theorie der Vorausahnung und einer top-down Eliminierung möglicher Lautsegmente des Audiosignals durch die visuelle Information. fMRT und PET-Verfahren (siehe Kapitel 2.2.4) zeigten zusätzlich, dass bestimmte Gehirnregionen besonders auf synchrone Bild-Ton-Darbietung reagieren. Neuronenaktivitäten ließ weiterhin den Schluss zu, dass eine zeitliche Latenz zwischen visuellem und auditivem Signal, um eine starke Aktivierung der Neuronen auszulösen. Der Mensch besitzt durch dieses zeitliche Intervall ein Synchronitätsfenster, in dem die Integration der visuellen und auditiven Signale in einer akzeptablen, d. h. für den Menschen synchronen Form, stattfinden kann. Außerhalb dieses Fensters sind die Abweichungen zu groß und die Signale werden getrennt und als nicht zusammengehörend interpretiert. Auf weite Forschungen und Untersuchungsergebnisse wird im Kapitel 2.3 "Aktueller Forschungsstand" eingegangen.

2.2 Psycholinguistische Grundlagen

Um ein Verständnis für die Fragestellung dieser Masterarbeit zu gewinnen, sollen im weiteren Verlauf die Grundlagen der Psycholinguistik und ihr Beitrag zum Sprachverständnis dargelegt werden. Dem genauen Vorgang des Sprachverstehens und dem Aufbau des kognitiven Systems wird von der Psycholinguistik in experimentellen Versuchen auf den Grund gegangen. Informationsfluss und Verarbeitung bei Sprachproduktion und Sprachverstehen sind dabei Hauptaugenmerk der Untersuchungen und machen die Zusammenarbeit mit verschiedenen Disziplinen erforderlich. Da sowohl linguistische Faktoren in Form von Semantik, Syntax und Phonetik und psychologische Aspekte der kognitiven Verarbeitungsprozesse wichtige Rollen spielen, bildet die Psycholinguistik ein übergreifendes Forschungsgebiet.

2.2.1 Sprachsignalwahrnehmung

Das Wahrnehmen von Sprachsignalen bildet einen Sonderfall der auditiven Signalverarbeitung, da neben der Verarbeitung des Reizes auch das Gedächtnis und der Grad der Aufmerksamkeit wichtige Faktoren sind. Die erste Frage, die sich stellt, ist ab welchem Punkt ein Sprachsignal anders verarbeitet wird als andere akustische Reize der Umgebung und welche Eigenschaften ein Sprachsignal haben muss, damit es von anderen Signalarten getrennt betrachtet wird. Untersuchungen zeigen, dass der primäre auditive Cortex der rechten und linken Hemisphäre auf Sprache in gleicher Weise reagiert, wie auf andere auditive Signale. Daraus lässt sich ableiten, dass eine Unterscheidung erst danach erfolgt. fMRT-Studien (siehe 2.2.4) zeigen, dass, während die rechte Hemisphäre hinter dem auditiven Cortex auf Tonhöhenschwankungen reagiert, die Aktivierung der linken Seite bei Sprachsignalen viel höher ist als auf der rechten Seite. Scott und Wise postulierten 2004, dass die ventrale Region der linken Hemisphäre auf Sprache spezialisiert ist und einer Spur folgt, die über mehrere Stationen die Spracherkennung ermöglicht. Zu diesen Stationen gehören semantische Erinnerungen, das Erkennen vom Sprachinhalt, welches als Ende des Erkennungsprozesses und Anfang des Produktionsprozesses angenommen wird, und die Projektion vom semantischen System zur frontalen Region, welche für die Artikulation und den Wortabruf zuständig ist (Ward, 2010, S.228).

Die am weitesten verbreitete Theorie der Sprachwahrnehmung ist die von Liberman und Mattingly 1985 postulierte Motortheorie der Sprachwahrnehmung. Sie beschreibt den Abgleich unendlich verschiedener akustischer Signale auf eine endliche Anzahl

gespeicherter Repräsentationen im Gehirn. Dabei findet die Abbildung des akustischen Signals auf die motorische Bewegung der Sprachproduktion für denselben Laut statt. Um also ein Phonem zu erkennen, werden motorische Bewegungen der Sprachproduktion in Betracht gezogen, die nötig wären, um den wahrgenommenen Reiz nachzubilden. Unterstützt wird diese Theorie von den Studienergebnissen von Best und Avery aus dem Jahr 1999, der zufolge Erfahrung bei der Sprachproduktion von Phonemen für die Spracherkennung nötig ist (Ward, 2010, S.226). Heutige Theorien basieren auf einer Mischung von verschiedenen Modellen, die in den meisten Fällen die Motortheorie mit der auditiven Verarbeitung kombiniert. Eine kurze Einführung in weitere Modelle kann im Werk von Barbara Höhle, 2012, S.50-52 gefunden werden.

Um die Motortheorie mit der auditiven Verarbeitung zu verbinden, gehen aktuelle Theorien von zwei Wegen zur Wahrnehmung und Produktion von Sprache aus: auf lexikalisch-semantischer Basis über den ventralen "Was"-Weg und auf der Basis eines auditiv-motorischen Abgleich über den dorsalen "Wie"-Weg. Da der dorsale Weg ebenfalls an der Ortung eines Sprachsignals beteiligt ist, wird hier auch vom dorsalen "Wo"-Weg gesprochen.

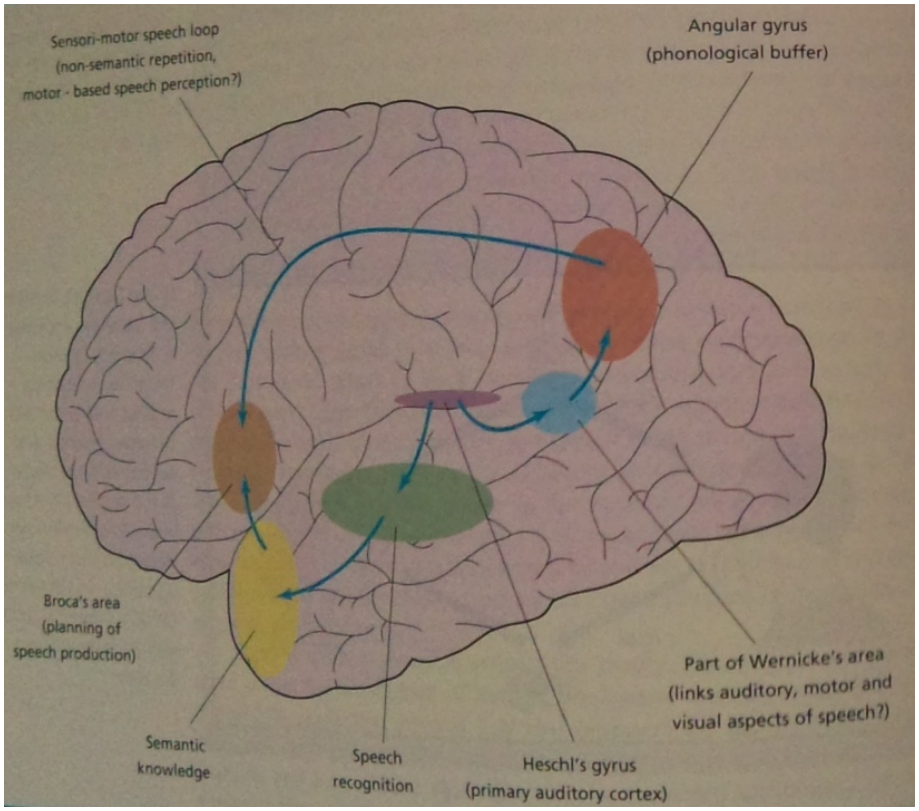

Abbildung 6: 2-Wege-Theorie zur Sprachwahrnehmung und -produktion (Ward, 2010, S.228)

Während der "Wie"-Weg mit der Motortheorie zu vergleichen ist und, sofern die Phoneme im Gedächtnisspeicher des Hörers vorhanden sind, sowohl für bekannte als auch für neue Wörter genutzt werden kann, funktioniert der "Was"-Weg nur für existierende Wörter. Um ein Wort zu erkennen, werden die auditiven Signale mit gespeicherten Mustern aus dem Langzeitgedächtnis vergleichen (pattern recognition). Dazu muss die Information allerdings nicht vollständig sein, da der Mensch in der Lage ist durch syntaktisches und semantisches Wissen vorhandene Lücken zu füllen. Die Verarbeitung wird dabei durch das akustische Gedächtnis (Echoic memory) unterstützt, welches die Speicherung des Reizes für ca.250ms erlaubt und somit eine parallele Weiterverarbeitung ermöglicht, auch wenn der physikalische Reiz bereits verklungen ist.

Die Wahrnehmung von Wort und Schrift ist sich sehr ähnlich, da der Mensch das gleiche Verständnis vom Inhalt erhält. Doch während man einen Text als konstantes Ganzes betrachten kann und einzelne Worte durch Leerzeichen getrennt sind, breitet sich Sprache über die Zeit aus, unterliegt starken Variationen und Wortanfänge und -enden sind nicht genauso klar erkennbar wie in schriftlicher Form. Barbara Höhle fasst bisherige Untersuchungen zur Erkennung von Sprachsignalen wie folgt zusammen "bislang ist nicht geklärt, welche speziellen akustischen Hinweise im Sprachsignal dem Hörer die Identifizierung von bestimmten Sprachlauten ermöglichen." (Höhle, 2012, S.43). Doch auch, wenn die Hinweise im Sprachsignal unklar sind, so sind doch die Probleme, die während der Verarbeitung auftreten, ersichtlich. Zum einen spielt die Sprechgeschwindigkeit eine große Rolle. Der Mensch versteht etwa 10 Phoneme pro Sekunde, was etwa 2 bis 3 Wörtern pro Sekunde entspricht und wofür eine sehr schnelle Verarbeitung nötig ist, und kann Sprache selbst bei einer künstlichen Beschleunigung auf 50-60 Phoneme pro Sekunde noch folgen (Eysenck et al., 2005, S.343; nach Werker und Tees, 1992). Außerdem besteht das Segmentierungsproblem, welches die Wortunterteilung auf Basis von Wissen erforderlich macht, da Wörter nicht bei der Schrift durch Lücken getrennt werden. Die Segmentierung erfolgt durch das Wissen von Buchstabenabfolgen, die in Silben einer Sprache nicht vorkommen (Eysenck et al., 2005, S.343-344; nach Dumay, Frauenfelder und Content, 2002).

Mattys erkannte 2004, dass Betonung ein stärkerer Indikator für Segmentierung ist als die Co-Artikulation (Eysenck et al., 2005. S.344; nach Mattys, 2004). Die Co-Artikulation beschreibt dabei die Abhängigkeit der Phonembildung von vorhergehenden und nachfolgenden Phonemen, welche generell innerhalb von Wörtern stärker ist als

zwischen Wörtern. An dieser Stelle soll kurz erwähnt werden, dass es sich bei Phonemen um die kleinste bedeutungsdifferenzierende Einheit einer Sprache handelt. So unterscheiden sich die Worte "Rat" und "Tat" nur durch ein Phonem. Grob umrissen hat eine menschliche Sprache zwischen 10 und 100 Phoneme. Die deutsche Sprache besitzt nach dem Institut für Sprache und Kommunikation der TU Berlin (2007-2008) ca. 40 Phoneme während im Englischen 44-50 Phoneme existieren. Vokalphoneme und auch konsonantische Phoneme entstehen dabei durch die Form und Stellung der Lippen bzw. des Kiefers und der Zunge. Neben den genannten Faktoren können auch individuelle Unterschiede der Sprecher die Verarbeitungsgeschwindigkeit beeinflussen. Dazu zählen zum Beispiel die individuelle Sprechgeschwindigkeit, mögliche Akzente und die Tonhöhe des Sprechers (Eysenck, 2005, S.343-345; Brown et al., 2000, S.125-130; Ward, 2010, S.225).

Eine wichtige Unterstützung zur Sprachsignalverarbeitung erhält das Gehirn durch gesammelte visuelle Information, denn Lippenlesen kann dabei helfen unzuverlässigen auditiven Input zu interpretieren (Ward, 2010, S.227; nach Sumby und Pollack, 1954). Ein von McGurk und MacDonald durchgeführter Versuch im Jahr 1976 zeigte jedoch, dass es zu Fehlinterpretationen bei abweichenden Informationen kommen kann. Die Darbietung der Lippenbewegung für /ga/ und der separate auditive Reiz /ba/ wurden von den Versuchsteilnehmern als neuer Laut /da/ wahrgenommen (McGurk und MacDonald, 1976).

2.2.2 Mentales Lexikon

Wie bereits erwähnt funktioniert die Spracherkennung über einen Abgleich der auditiven Form mit gespeicherten Mustern im Langzeitgedächtnis. Dieses Set an bekannten Wörtern stellt das phonologische Lexikon dar. Nach dem Sprachwissenschaftler Ferdinand de Saussure (1857-1913) kann ein Wort auf zwei Ebenen unterschieden werden: der lautlichen Gestalt und die Wortbedeutung. Modelle des mentalen Lexikons bauen darauf auf und nehmen mehrere Ebenen der Wissensrepräsentation im Gehirn an. Während zum Beispiel eine Ebene die Bedeutung eines Wortes speichert, codiert eine andere Ebene die lexikalisch-syntaktischen Eigenschaften (Höhle, 2012, S.56).

Für den Abgleich und Zugriff auf das mentale Lexikon wurden in den vergangenen Jahren unterschiedliche Modelle entwickelt. Das erste Modell von Oldfield und

Wingfield (1965) nahm eine nach Häufigkeit der Nutzung sortierte Wortliste an. Heutige Modelle gehen eher von einem Netzwerk als von einer Liste aus, wobei die am stärksten diskutierten Zugriffsarten Silben und Lautdehnungen bzw. Betonungen sind und somit sprachabhängig (Höhle, 2012, S56-58).

Auch wenn die Zugriffsmethode auf das mentale Lexikon bisher ungeklärt bleibt, sind sich Forscher über den Vorgang der Worterkennung relativ sicher. Sobald ein ausreichend langes Sprachsignal registriert wurde (1-2 Phoneme), findet eine Art Konkurrenzkampf ähnlich klingender Lauter statt. Das erste Modell zur Worterkennung wurde von Marslen-Wilson und Tyler 1980 als Kohorten-Modell vorgestellt. Dieses Modell geht von einer Aktivierung aller passenden Wörter im mentalen Lexikon aus und einer bottom-up Hemmung, sobald ein Wort nicht mehr mit dem Sprachsignal übereinstimmt.

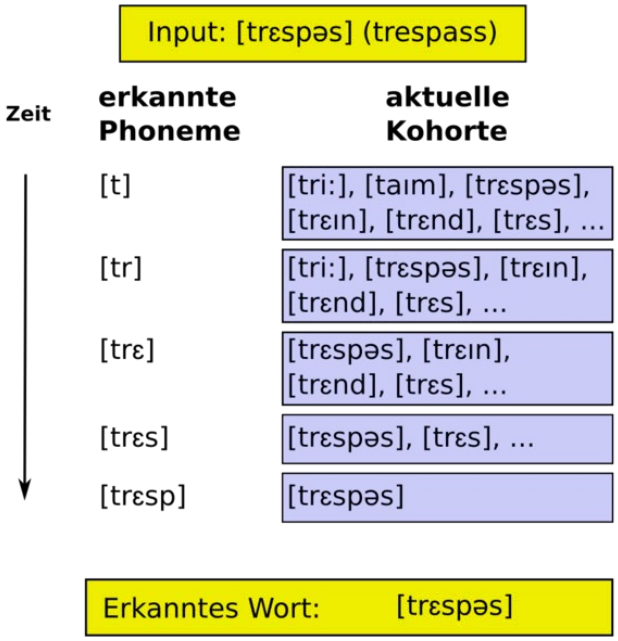

Abbildung 7: Kohorten-Modell (Bildquelle: [8])

Der Punkt, ab dem nur noch ein Wort mit dem auditiven Signal übereinstimmen kann, wird als Diskriminationspunkt (uniqueness point) bezeichnet. Eine absteigende Aktivierung kann durch Kontext Kandidaten eliminieren, was schließen lässt, dass nach dem Kohorten-Modell viele Quellen wie lexikalisches, syntaktisches und semantisches Wissen gleichzeitig verwendet und integriert werden, um das richtige Wort zu finden. In den 90'er Jahren wurde das Modell von Marslen-Wilson überarbeitet, da ein zu starker Fokus auf den Wortanfang gelegt wurde und unpassende Wörter eliminiert wurden, obwohl Untersuchungen zeigen, dass der Mensch auch dem Kontext nach unpassende Wörter in Betracht zieht (Höhle, 2012, S.77-78). Nach dem überarbeiteten Modell werden Wortkandidaten in unterschiedlichem Maße aktiviert. Durch das Aktivierungslevel ergibt sich ein variabler Anteil an der Zugehörigkeit zu einer Kohorte. Und auch der Einfluss des Kontexts auf das Erkennen von Wörtern wurde soweit reduziert, dass er erst bei der Integration in den entstehenden Satz zum Tragen kommt (Eysenck et al., 2005, S.349).

Weitere Modelle wie das TRACE-Modell von James McClelland und Jeffrey Elman aus dem Jahr 1986 und das SHORTLIST-Modell von Dennis Norris (1994) werden von Katharina Spalek im Werk von Barbara Höhle kurz beschrieben und diskutiert (Höhle, 2010, S.77-80).

2.2.3 Fremdsprache

Die genannten und darauf verwiesenen Modelle zur Worterkennung gelten für alle Sprachen, doch sollte eine Unterscheidung für den Hörer hinsichtlich seiner Muttersprache und erlernte Fremdsprachen getroffen werden. Alle Sprachen besitzen Silben, doch die Einteilung des Sprachsignals in silbengroße Einheiten ist je nach Sprache unterschiedlich. So würde das Wort "Panda" von einem Deutschen in die zwei Silben /pan/ und /da/ zerlegt werden, während ein Japaner /pa/ + /n/ + /da/ verstehen würde (Pisoni et al., 2006, S.553, nach Otake et al. 1993). Es wurde vorgeschlagen Sprachen nach ihrem Rhythmus zu unterscheiden, da dieser einen Einfluss auf die Wahrnehmung einer Sprache hat. Pike schlug diese Unterscheidung erstmals 1945 vor und teilte die menschlichen Sprachen in drei Kategorien ein:

1. "syllable-timing"
Der Sprachrhythmus teilt die Sprache in etwa gleichlange Silben ein (z. B. Französisch oder Spanisch).
2. "stress-timing"
Die Silben besitzen unterschiedliche Länge doch es existieren etwa gleichlange Abstände zwischen den Silben (z. B. Englisch oder Deutsch).
3. "mora-timing"
Dieser Rhythmus ist "syllable-timing" sehr ähnlich, doch wird die Sprache hier nicht in Silben, sondern in Mora von gleicher Länge unterteilt (z. B. Japanisch oder Tamil) (Pisoni et al., 2006, S.554, nach Pike, 1954).

An dieser Stelle soll erwähnt werden, dass die im Versuch verwendeten Sprachen unterschiedliche Sprachrhythmen besitzen. Während Deutsch und Englisch "stress-timing" Sprachen sind, gehört Persisch dem " The Accent Lab " (2011) zufolge zu den "syllable-timing" Sprachen.

Während der Sprachrhythmus die Segmentierung erleichtert, ist es das sprachliche Wissen, was es dem Hörer ermöglicht zu entscheiden, ob eine Segmentabfolge ein

Wort der eigenen Sprache ergibt oder nicht. Der Zugriff auf Wörter einer Zweitsprache eröffnet die Frage nach dem Aufbau des mentalen Lexikons. Im Folgenden ist das Revised-Hierarchical-Model von Kroll und Steward aus dem Jahr 1994 dargestellt. Es geht von separaten Lexika für Erst- und Zweitsprache aus, wobei die lexikalische Repräsentation der Erstsprache eine direkte Verbindung zum Kontext hat, während die Zweitsprache nur über einen Zugriff auf die Erstsprache erschlossen wird. In Abhängigkeit vom Sprachniveau entwickelt sich mit der Zeit eine direkte Verbindung auch für die Zweitsprache zum kontextuellen System.

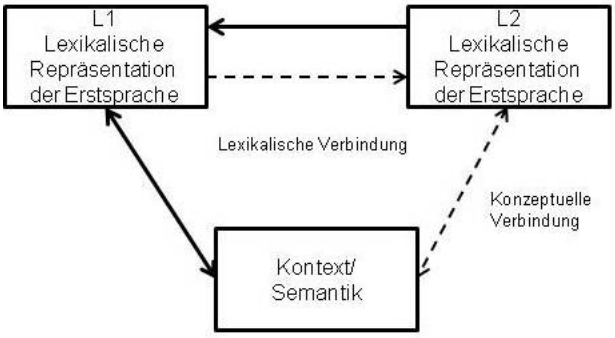

Abbildung 8: Revised-Hierarchical-Model (Höhle, 2012, S.180; nach Kroll und Stewart 1994)

Andere Modelle, wie das Bilingual-Interactive-Activation-Modell (BIA, BIA+) von Dijkstra und van Heuven gehen hingegen von einem multilingualen Netzwerk und einem integrierten Lexikon aus, wonach ähnlich klingende Wörter oder ähnlich aussehende Buchstaben aller Sprachen gleichzeitig aktiviert werden. Sprachkompetenz, Häufigkeit der Wörter und auch phonetische Informationen entscheiden über den Aktivierungsgrad der Kandidaten (Höhle, 2012, S.179-181).

Mit Hilfe von bildgebenden Messmethoden (siehe Abschnitt 2.2.4) wurde in neuropsychologischen Studien versucht, einen Eindruck von den Zugriffszeiten und Verarbeitungsgeschwindigkeiten für Erst- und Zweitsprache zu gewinnen. In Verhaltensstudien wurde nachgewiesen, dass Zweitsprachler Probleme haben Phoneme oder Vokallängen zu unterscheiden, die in der Muttersprache nicht auftreten bzw. keinen Bedeutungsunterschied besitzen. Für die Verarbeitung syntaktischer Informationen scheinen auf ineffiziente Art die gleichen Areale wie bei Muttersprachlern aktiviert zu werden. So werden für die Verarbeitung zwar keine anderen Hirnareale verwendet, die Verarbeitungs- und Zugriffszeit erhöht sich jedoch, je seltener die Strukturen verwendet werden (Mueller et al., 2006, S.183). Dies stimmt mit dem Activation-Threshold-Modell von Michel Paradis von 2004 überein. "Es geht davon aus, dass jede Sprache einen gewissen Schwellenwert hat, der bestimmt, wie gut auf die Sprache zugegriffen werden kann. [...] Zum einen ist er vom Leistungsniveau abhängig [...] Zum anderen hat auch die Frequenz des Zugriffs auf die Sprache einen Einfluss auf den Schwellenwert [...]" (Höhle, 2012, S.185-186; nach Paradis, 2004). Je mehr mit

einer Fremdsprache umgegangen wird, umso schneller werden Informationen verarbeitet. Dabei hängt der Grad der Sprachfähigkeit von Faktoren wie Erwerbsalter, Lerngeschichte und Motivation ab. Während ein einheitliches Zugriffsmodell für das multilinguale mentale Lexikon noch nicht entwickelt wurde, konnte nachgewiesen werden, dass sich mit steigender Sprachkompetenz die Verarbeitungszeit an die der Muttersprache annähert.

2.2.4 Messmethoden

Untersuchungen basieren auf einem von zwei Verfahren: verhaltensorientierte Methoden oder neurowissenschaftliche Methoden. Bei verhaltensorientierten Versuchen haben Probanden die Aufgabe eine lexikalische, syntaktische oder semantische Entscheidung zu treffen. So kann die Aufgabenstellung zum Beispiel darin bestehen zu entscheiden, ob das dargebotene Wort ein existierendes Wort oder ein Nicht-Wort ist. Dabei können die Stimuli visuell über einen Monitor oder auditiv über Kopfhörer oder Boxen übertragen werden. Gemessen wird in den meisten Fällen die Reaktionszeit, die der Proband braucht, um eine Entscheidung zur jeweiligen Fragestellung zu fällen. Zu diesen neurowissenschaftlichen Methoden zählen unter anderem die als fMRT abgekürzte funktionelle Magnetresonanztomografie, die Positronen-Emissions-Tomografie (PET) sowie die Elektroenzephalografie (EEG) bzw. die ereigniskorrelierten Potentiale (EKP) (Höhle, 2012, S.33).

Das EEG misst die elektrische Aktivität an der Schädeloberfläche und ist nicht invasiv. Allerdings müssen für Messungen eine hohe Anzahl von parallel ausgerichteten Neuronen feuern, damit sie sich summieren anstatt gegenseitig aufzuheben. Als Referenz wird ein unbeeinflusstes Hirnareal gewählt und mit den gemessenen Spannungswerten im aktivierten Bereich verglichen. Dafür werden Elektroden in einer Standardanordnung an der Schädeloberfläche angebracht. Die ermittelten Ergebnisse werden in Wellenform dargestellt und beschreiben nach einer zeitlichen Einteilung bezogen auf ein bestimmtes Ereignis die ereigniskorrelierten Potentiale. Durch ihre sehr hohe Zeitauflösung sind EKP-Messungen sehr gut für die Darstellung der Sprachverarbeitung geeignet. Eine detaillierte Beschreibung dieser EEG- und EKP-Messung kann im Werk von Höhle, 2012 S. 113-123 gefunden werden. Dabei geht der Autor dieses Kapitels, Heiner Drenhaus, auf die von Kutas und Hillyard erstmals 1980 festgehaltene N400 Komponente ein. N400 beschreib die Negativierung nach einem semantischen Fehler und tritt als maximale Abweichung zum Referenzwert etwa

400ms nach der Repräsentation auf. In ähnlicher Weise sind weitere Anhaltspunkte für sprachliche Verarbeitung benannt. So wird zum Beispiel der positive Ausschlag nach etwa 600ms als P600 bezeichnet und stellt die "Reanalyse und Reparatur einer anfänglich aufgebauten Struktur, die im weiteren Verlauf der Verarbeitung als inkorrekt oder nicht möglich identifiziert wurde", dar (Höhle, 2012, S.121).

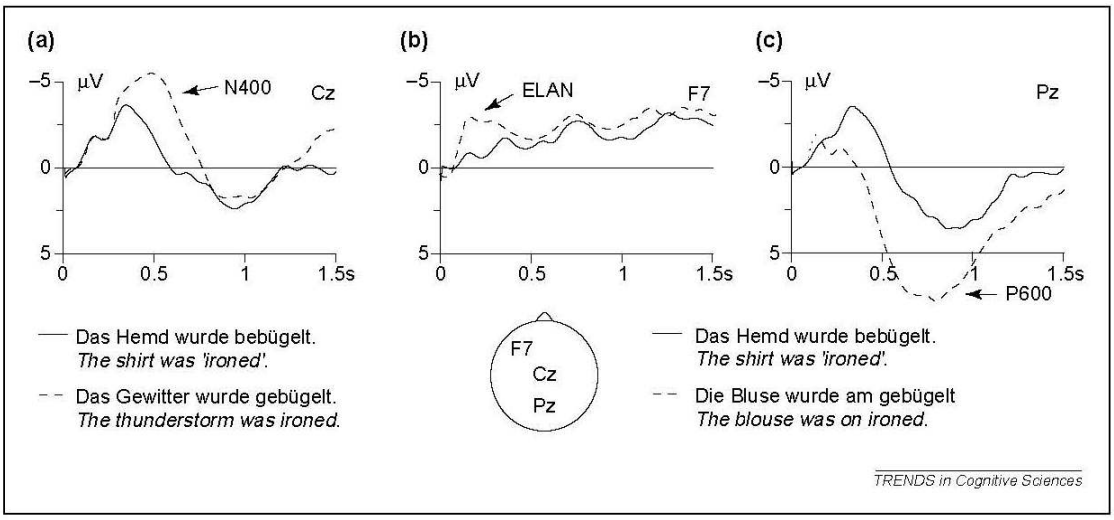

Abbildung 9: Sprachbezogene EKP-Werte: (a) N400, (b) very early left-anterior negativity (ELAN), (c) P600 (Friederici, 2002)

Während die bisher genannten Methoden zu den funktionellen bildgebenden Verfahren gehören, zählen die Computer Tomografie (CT) und Magnetresonanztomografie (MRT) zu den strukturellen bildgebenden Verfahren. Sie basieren auf den unterschiedlichen physikalischen Eigenschaften der Gehirnschichten (Schädel, graue Masse etc.) und der Annahme, dass neurale Aktivität eine physiologische Änderung in den Gebieten bewirkt. Im Fall von CT-Scan wird Röntgenstrahlung in Abhängigkeit von der Gewebedichte unterschiedlich stark absorbiert. Dieser Vorgang kann allerdings nicht für die funktionellen bildgebenden Methoden angepasst werden, sondern liefert ein statisches Bild, welches zur Diagnose von Krankheiten verwendet werden kann (Ward, 2010, S.49). Im Gegensatz zu einem CT-Scan wird bei MRT keine radioaktive Strahlung benötigt und bietet gleichzeitig eine bessere räumliche Auflösung. Der Vorgang kann soweit angepasst werden, dass das BOLD-Signal gemessen werden kann. BOLD steht für Blood-Oxygen-Level-Dependent und wird in diesem Zusammenhang als fMRT bezeichnet. Die Basis für diese Messung ist die Sauerstofferhöhung in bestimmten Hirnarealen bei der Aktivierung entsprechender Nervenzellen. In einem MRT-Scanner werden Probanden Aufgaben präsentiert, deren Lösung zu einer Aktivierung der Nervenzellen und einer Erhöhung des Sauerstoffbedarfs führt. Die Änderung des Sauerstofflevels kann durch eine fMRT-

Messung festgehalten werden und liefert somit ein Bild dynamischer Änderungen im Gehirn. Dabei unterliegt der Vorgang allerdings einer zeitlichen Latenz von etwa zwei Sekunden. Signaländerungen werden erst zwei Sekunden nach Reizpräsentation von den Geräten registriert und dauern ca. 12-15 Sekunden an (Höhle, 2012, S.34). Es kann festgehalten werden, dass fMRT-Messungen in einer verhältnismäßig stark verzögert Form die Konsequenzen kognitiver Aktivität, die Änderung im Sauerstofflevel, misst. MRT-Scanner sind zusätzlich sehr laut und schon kleinste Bewegungen können sich störend auf die Aufzeichnung auswirken (Ward, 2010, S.53-55)

In ähnlicher Form wie die fMRT ermittelt die PET eine Änderung im Blutstrom mithilfe Isotopenindikatoren. Da das Gehirn immer aktiv ist und alle Regionen durchgehend Blut und Sauerstoff erhalten, kann ein kognitiver Prozess nur durch die Beobachtung von relativen Änderungen zu einem Referenzwert gemessen werden. Die zeitliche und räumliche Auflösung von PET-Messungen ist im Vergleich zu fMRT-Messungen wesentlich schlechter, da der Indikator erst nach einer Verzögerung von etwa 30 Sekunden seinen maximalen Ausschlag erreicht (Ward, 2010, S.53). Trotz der zeitlichen Verzögerung kann PET für die Nachverfolgung kognitiver Prozesse genutzt werden. Der Indikator sammelt sich je nach Aufgabenstellung in bestimmten Hirnarealen an und kann über die Messung der Gammastrahlung farbcodiert dargestellt werden. Im Gegensatz zur fMRT ist der Teilnehmer einem invasiven Verfahren und einer geringen Menge Radioaktivität ausgesetzt. Generell liefern die bildgebenden Messverfahren keinen genauen Ablauf kognitiver Prozesse. Die farbcodierten Bereiche sind die Ergebnisse von statistischen Auswertungen und geben Auskunft darüber, welche Hirnareale wahrscheinlich bei der Bearbeitung bestimmter Aufgaben relevant sind. Die Auswertung der gewonnenen Daten unterliegt nicht nur den allgemeinen statistischen Fehlern, sondern auch der individuellen Interpretation der Ergebnisse. Ward formuliert es wie folgt (Ward, 2012, S.65):

> *The most common way of dealing with individual differences is effectively to assume that they don't exist. Or, more properly put, individual differences needn't get in the way of making claims about general brain functions.*

Im Folgenden sollen auf ein paar ausgewählte Studien der letzten zehn Jahre eingegangen werden und ihre Ergebnisse im weiteren Verlauf diskutiert und mit eigenen Ergebnissen verglichen werden.

2.3 aktueller Forschungsstand

Nachfolgend werden einzelne Studien zur audiovisuellen Integration und Sprachverarbeitung vorgestellt. Es wurden jeweils unterschiedliche Versuchsdesigns, -materialien und Aufgaben verwendet, sodass die Ergebnisse nur schwer untereinander zu vergleichen sind. Sie sollen daher nur einen Überblick über bisher durchgeführte Untersuchungen und die Vielfalt der Herangehensweisen geben.

Der von Kroll et al. 2002 durchgeführte Hörversuch zur Zugriffszeit auf das mentale Lexikon der Zweitsprache wurde bereits im Abschnitt 2.2.3 kurz angesprochen. In zwei Experimenten wurden Reaktionszeiten für den Zugriff auf die jeweilige Zweitsprache und Übersetzungsgeschwindigkeiten untersucht. Basis für die Experimente war das Revised-Hierarchie-Modell von Kroll und Stewart (1994) und das BIA-Modell von Dijkstra und van Heuven (1998). Im ersten Versuch wurden die Leistungen von englischen Muttersprachlern mit unterschiedlicher Sprachkompetenz des Französischen verglichen. Die erste Aufgabe bestand dabei in einem reinen Benennen eines visuell dargebotenen Wortes, während bei der zweiten Aufgabe die Übersetzung eines visualisierten Wortes in die jeweils andere Sprache (L1 zu L2, oder L2 zu L1) gefragt war. An diesem Versuch haben 59 Studentinnen des Mout Holyoke Colleges teilgenommen. Ihnen wurden für jeden der vier Aufgabentypen eine Liste von 30 französischen Wörtern, insgesamt 120 Wörter, präsentiert. Ein entsprechender Vergleich der Listen im Bezug auf Wortlänge kann im Bericht gefunden werden (Kroll et al., 2002, S.145). Die 59 Teilnehmer wurden für den Versuch in Abhängigkeit von ihrem Sprachlevel in zwei Gruppen aufgeteilt: fließende (mehr als 5 Jahre) und nicht fließende Sprecher (weniger als 5 Jahre). Die Ergebnisse zeigen einen signifikanten Unterschied für die Zugriffszeiten auf die Zweitsprache. So war die Latenz bei den fließenden Sprechern geringer als bei den Teilnehmern mit geringerer Sprachkompetenz. Es wurde daher geschlussfolgert, dass sich die Zugriffszeit auf das Zweitsprachenlexikon mit dem Level der Sprachfähigkeit ändert (Kroll et al., 2002, S.147).

Im Vergleich zur Verhaltensforschung nutzen die meisten Studien zur audiovisuellen Verarbeitung bildgebende Verfahren, wie in der PET-Studie von Macaluso et al. von 2003. Macaluso und Kollegen untersuchten auditive und visuelle Interaktion und ihre Beeinflussung auf die Aktivierung von bestimmten Hirnarealen. An dem Versuch nahmen 8 männliche Probanden teil, deren Muttersprache Englisch war. Ihnen wurde in einem PET-Scanner-Raum ein Video vorgespielt, auf dem Kopf und

Lippenbewegung eines Sprechers zu sehen waren. Aus einer Liste von 200 Wörtern (160 Nontarget, 40 Target) wurden 50 Wörter (40 Non-Target, 10 Target) ausgewählt und präsentiert. Im Fall der asynchronen Darbietung wurde das Audiosignal 240ms früher als das Bildsignal abgespielt. Während der Monitor fixiert war, wurde das Audiosignal über eine von zwei Boxen wiedergegeben, wobei sich eine Box direkt unter dem Monitor befand und die andere 16,6° vom visuellen Stimulus entfernt (Macaluso et al., 2003, 727). Somit konnten Messungen mit synchroner bzw. asynchroner Darbietung und räumlicher Übereinstimmung bzw. Abweichung durchgeführt werden. Die Aufgabe der Probanden während des Versuchs bestand darin einen Kopf zu drücken, sobald sie eines der Target-Wörter hörten. Ein Ergebnis dieser Studie war der Befund, dass die Aktivität im superior temporal sulcus (STS) bei synchroner Darbietung höher ist, unabhängig von der räumlichen Übereinstimmung. Weiterhin wurden Hirnareale in lateralen und dorsalen Hinterkopf-Regionen festgestellt, die nur auf zeitliche und räumliche Synchronität reagieren und eine Region im rechten inferior parietal Cortex, welche auf audiovisuelle Synchronität und räumliche Abweichung, den Bauchrednereffekt, reagiert (Macaluso et al., 2003, S. 730).

Navarra et al. führten 2005 Untersuchungen zur Ausweitung des Synchronitätsfensters durch die Präsentation von asynchronem audiovisuellem Material durch. Sie untersuchten wie flexibel der Bereich ist, in dem eine Integration möglich ist. Es nahmen 12 Probanden teil. Bei dem audiovisuellen Material handelte es sich um eine Aufnahme eines männlichen Sprechers, welcher 1000 Wörter von einer Liste vorlas. Das Sprachsignal wurde über Boxen, welche sich in einem Abstand von 5 cm rechts und links vom Monitor befanden, wiedergegeben. In 8 Blöcken von jeweils 5 min wurde den Probanden das Video präsentiert, mit der Aufgabe aus der Liste Target-Wörter mitzuzählen. Gleichzeitig sollte eine TOJ-Aufgabe (temporal order judgment) durchgeführt werden. Hierzu wurde eine rote LED-Lampe 2 cm unter dem Monitor angebracht und White-Noise übertragen. Über zwei Tasten sollten die Probanden entscheiden, welches der beides Ereignisse zuerst erschien. Die zeitliche Abweichung des auditiven und visuellen Stimulus betrug dabei ±500ms, ±350ms ±250ms, ±150ms, ±50ms und 750ms, wobei eine positive Angabe eine auditive Verzögerung bedeutete. 4 der 8 Blöcke beinhalteten eine zeitliche Verzögerung des Audiosignals des Videos um 300ms. Die Ergebnisse deuteten an, dass das Beobachten von asynchronem audiovisuellem Material zu einer Änderung des Synchronitätsfensters führt und eine Synchronitätsentscheidung über Nicht-Sprachsignale beeinflusst (Navarra et al., 2005, S.503). Die Komplexität des Sprachsignals führt bei asynchroner Darbietung zu einer

Ausweitung des Synchronitätsfensters, doch durch einen weiteren Versuch mit einer auditiven Verzögerung von 1000ms statt 300ms zeigte, dass sich das Synchronitätsfenster nur bedingt anpassen lässt. Eine korrekte Integration auditiver und visueller Informationen ist bei einer Asynchronität von mehr als 300ms vom Menschen nicht mehr umsetzbar (Navarra et al., 2005, S.505).

Zwei Experimente von van Wassenhove et al. im Jahr 2005 nutzen den McGurk-Effekt als Grundlage, um die Toleranzgrenzen für den McGurk-Effekt bei verschiedener audiovisueller Asynchronität zu untersuchen. Frühere Forschungsergebnisse legten nahe, dass sich dieses Synchronitätsfenster in einem Bereich von +40ms bis -240ms bewegt. Es nahmen 43 Studenten der University of Maryland in zwei Gruppen teil. Eine Gruppe bestand aus 21 Teilnehmern, die zweite aus 22 und sie bekamen jeweils gruppenabhängig unterschiedliches audiovisuelles Material gezeigt. Als Videobasis dienten die Stimuli aus einer Untersuchung von Grant et al. (1998), welche von analog auf digital übertragen, und zusätzlich so synchronisiert wurde, dass der McGurk-Effekt entstand. Es ergaben sich 29 Stimuli in einem Bereich von ±467ms um den Synchronitätspunkt herum, inklusive eines synchronen Stimulus. Im ersten Experiment wurde mit einem 3AFC-Verfahren und jeweils 10-facher Wiederholung der 29 Stimuli die Integration audiovisueller Informationen überprüft, indem Probanden auswählen konnten, ob sie /ba/, /ga/ oder /da/ bzw. in der zweiten Gruppe /pa/, /ka/ oder /ta/ verstanden haben. Im zweiten Experiment wurden für jeden Zeitversatz 6 Wiederholungen angeboten und ein 2AFC verwendet. Die Probanden sollten entscheiden, ob das dargebotene audiovisuelle Material "gleichzeitig" oder "nacheinander" erschien. Ergebnisse des ersten Versuchs zeigen, dass die Probanden bei visuellem /ka/ und auditivem /pa/ im Bereich von 167ms bis -267ms Audioverfrühung bzw. -verzögerung ein kombiniertes /ta/ wahrgenommen haben. Weiterhin konnte gezeigt werden, dass bei abweichender audiovisueller Information (z. B. Audio: /ba/, Video: /ga/) die synchrone Darbietung signifikant schlechter erkannt wurde, als bei audiovisueller Informationsübereinstimmung. Dies unterstützt die aufgestellte Theorie, dass an der erfolgreichen audiovisuellen Integration von Sprache auch passende Gesichtsbewegungen beteiligt sein müssen (van Wassenhove et al., 2005, S.603-604). Zusätzlich konnte gezeigt werden, dass "Leading auditory information decreases integration, while leading visual information tends to enhance it." (van Wassenhove et al., 2005, S.605). Dies wird besonders im Bezug auf Sprachsignale auf unterschiedliche Verarbeitungsgeschwindigkeit für auditive und visuelle Informationen zurückgeführt (siehe Abschnitt 2.1.3).

Den Ansatzpunkt für die vorliegende Masterarbeit bildet die Magisterarbeit von Berta Luise Heide von 2009 zur perzeptiven Synchronität von Bild und Ton. Mit 105 Versuchsteilnehmern untersuchte sie in drei Versuchdesigns die audiovisuelle Integration und den Punkt subjektiver Synchronität. Sie ließ die Probanden in einem Erkennungsversuch, Herstellungsversuch und Beurteilungsversuch drei unterschiedliche Audiosignale - Sprache, Musik, Geräusch - hinsichtlich Synchronität bewerten. Im Erkennungsversuch sollte innerhalb eines Paarvergleichs entschieden werden, welches Video physikalisch synchron war. Die Ergebnisse für diesen Versuch zeigen, dass zwischen -40ms und -80ms die häufigste Fehlbewertung abgegeben worden ist. Der Herstellungsversuch erlaubte es den Teilnehmern den optimalen Bild-Ton-Versatz einzustellen und somit den Punkt subjektiver Gleichzeitigkeit herzustellen, welche zwischen -32ms und -40ms lag. Als Letztes führte Frau Heide einen Beurteilungsversuch durch, bei dem der Grad der Synchronität durch Schulnoten bewertet werden sollte. Auch hier zeigten die Ergebnisse, dass eine Tonverzögerung um -40ms oder -80ms besser bewertet wurden, als die physikalisch synchrone Fassung.

2009 nutzten Stevenson et al. bildgebende Messverfahren, um einen Eindruck von den Hirnregionen zu bekommen, welche an der audiovisuellen Integration und dem Synchronitätsfenster beteiligt sind. Dies geschah mithilfe eines fMRT-Scanners und gemessenen BOLD-Werten. An diesem zwei-phasigen Experiment nahmen 8 englische Muttersprachler teil. Die Stimuli für diesen Versuch stammten aus „The Hoosier Audiovisual Mulit-Talker Database“ und beinhalteten eine Sprecherin, welche zehn bekannte Substantive vortrug. In der ersten Phase wurden den Probanden einzelne Wörter mit audiovisuellen Versätzen im Bereich von +400ms bis -400ms (inklusive 0ms) in 100ms Intervallen präsentiert, wobei jeder Versatz sieben Mal vorkam. Im Bezug auf die asymmetrische Verteilung der wahrgenommenen Synchronität des audiovisuellen Materials konnten Ergebnisse früherer Studien erneut belegt werden und zeigten, dass eine vorauseilende Tonspur mit höherer Wahrscheinlichkeit als asynchron erkannt wird, als eine verzögerte. Zusätzlich wurden zwei sich nicht überlappende Sub-Bereiche im multisensory superior temporal cortex (mSTC) ermittelt, von denen S-mSTC eine BOLD-Aktivierung nur für synchrones audiovisuelles Material unter einer zeitlichen Latenz von 100ms zeigte. B-mSTC hingegen wies unterschiedliche Aktivierungsgrade bei allen Sprachsignalen in Abhängigkeit von der Asynchronität auf. Nach der Auswertung der Daten wurde von Stevenson et al. (2007) die Hypothese aufgestellt, dass S-mSTC als Filter zur ersten

Überprüfung von Synchronität dient, während B-mSTC asynchrone Signale bzw. mehrere Sinnesreize wahrnimmt und mit steigendem zeitlichen Versatz eine höhere Verarbeitungsleistung zur Integration nötig ist. Aufgrund des Versuchsaufbaus konnte diese Theorie während der Studie jedoch nicht überprüft werden (Stevenson et al., 2009, S.3313-3316).

Navarra et al. führten 2010 eine Untersuchung zum PSS für Muttersprachler und Nicht-Muttersprachler durch. Es nahmen 11 englische Muttersprachler ohne Spanisch-Kenntnisse, 9 spanische Muttersprachler mit sehr geringen Englischfähigkeiten und 7 spanische Muttersprachler mit sehr hoher Sprachkompetenz in Englisch. Die Einschätzung wurde jeweils von den Probanden mithilfe eines Fragebogens durchgeführt. Das Material bestand dabei aus einem Video mit einem bilingualen Sprecher des Englischen und Spanischen, welcher Sätze in beiden Sprachen vortrug. Die Latenz betrug dabei zwischen -440ms, -320ms, -280ms, -200ms, -80ms, +40ms, +80ms, +120ms und +280ms. Das Audiosignal wurde für jeden Videoclip mit einem Fade-in und Fade-out versehen und über zwei Lautsprecher an den Seiten des Monitors wiedergegeben. In 4 Blöcken á 54 Trials wurden den Probanden spanische und englische Sätze in zufälliger Reihenfolge dargeboten. Die Aufgabe während des Versuchs bestand aus einer SJ-Entscheidung. Die Probanden sollten angeben, ob auditive und visuelle Information synchron dargeboten wurden oder nicht. Die Ergebnisse zeigen, dass in der Muttersprache der jeweiligen Teilnehmer das auditive Signal stärker verzögert werden muss als in einer unvertrauten Fremdsprache. Im Vergleich dazu konnte gezeigt werden, dass sich der PSS für eine sehr vertraute Fremdsprache nicht von dem der Muttersprache unterscheidet (Navarra et al., 2010, S.86).

Die hier genannten Studien sollten einen Überblick über die Forschungsvielfalt im Bereich der audiovisuellen Informationsverarbeitung und -integration geben und den Ansatzpunkt für den folgenden Hörversuch bilden. Eine detaillierte Beschreibung der jeweiligen Untersuchungen und weiterführende Literatur befinden sich im Literaturverzeichnis der beigelegten CD.

3. Experimentelle Versuchsdurchführung

Nachdem im vorherigen Kapitel auf die Grundlagen der audiovisuellen Informationsverarbeitung eingegangen wurde, wird in diesem Kapitel der im Rahmen der Masterarbeit durchgeführte Hörversuch dargelegt. In Unterpunkt wird auf die Erstellung des Videomaterials, den Versuchsaufbau in die Versuchsdurchführung eingegangen.

3.1 Material

Mit dem durchgeführten Hörversuch soll das Erkennen von physikalischer Synchronität in Abhängigkeit von Sprachkenntnissen untersucht werden. Hierfür wurden in Vorüberlegungen benötigte Materialien zusammengestellt, um eine Vergleichbarkeit der Ergebnisse untereinander zu ermöglichen. Neben vergleichbaren Texten und Videoaufzeichnungen wurden mindestens 30 Probanden, eine akustisch trockene Versuchsumgebung und eine Software zum Erheben empirischer Daten benötigt. Im Folgenden soll kurz auf die Erstellung der Materialien eingegangen werden.

3.1.1 Sprachmaterial

Die Auswahl des passenden Sprachmaterials für die Videoaufzeichnungen orientierte sich an der Phonetik der Laute und der damit verbundenen Positionen von Lippen und Kiefer. Plosive wie /p/ und /b/ werden in derselben Position mit gepressten Lippen für den Druckaufbau erzeugt. Sie können in einem Video nicht auseinandergehalten werden, doch liefern die zusammengepressten Lippen einen guten Hinweis für den Zuschauer, dass es sich um einen der beiden Plosive handeln muss. Das Gleiche gilt für die Frikative /f/ und /v/. Und auch der nasale Laut /m/ ist an den Bewegungen der Lippen gut zu erkennen.

Diese fünf Laute sind der Grundstein für die Arbeit in der Filmsynchronisation, da Synchrontexter darauf achten müssen, entsprechende Dialoge so zu übersetzen und anzupassen, dass in der synchronisierten Fassung die gleichen Lippenbewegungen an den gleichen Stellen vorkommen. Die genannten Laute sind durch die eindeutige Positionierung der Lippen nicht nur ein erster Anhaltspunkt, um gute Synchronisation zu erkennen, sondern auch ein Merkmal für audiovisuelle Synchronität.

Damit die Versuchsergebnisse nicht von einer erhöhten Anzahl an oben genannten Lauten in einer Sprache beeinflusst werden, wurde in Vorüberlegungen auf eine Gleichverteilung in den jeweiligen Sprachen geachtet. Als Textgrundlage wurde der in der Linguistik oft verwendete Text der Aesop-Fabel genutzt: Der Nordwind und die Sonne. Dieser Text wurde in viele Sprachen übersetzt und bildet die Referenz für das "Internationales Phonetic Alphabet", da eine Transkription eine Darstellung der phonetischen Laute einer Sprache liefert. Während die Transkription der deutschen Fabel und ihrer englischen Übersetzung dem „Handbook of the International Phonetic Association“ entnommen wurden, wurde die phonemische Transkription des Persischen von der Muttersprachlerin Mina Fallahi erstellt.

Der deutsche Text, welcher die Basis für den phonetischen Vergleich und der nachfolgenden Satzauswahl für die Videoaufzeichnung bildete, lautet wie folgt:

Einst stritten sich Nordwind und Sonne, wer von ihnen beiden wohl der Stärkere wäre, als ein Wanderer, der in einen warmen Mantel gehüllt war, des Weges daherkam. Sie wurden einig, daß derjenige für den Stärkeren gelten sollte, der den Wanderer zwingen würde, seinen Mantel abzunehmen. Der Nordwind blies mit aller Macht, aber je mehr er blies, desto fester hüllte sich der Wanderer in seinen Mantel ein. Endlich gab der Nordwind den Kampf auf. Nun erwärmte die Sonne die Luft mit ihren freundlichen Strahlen, und schon nach wenigen Augenblicken zog der Wanderer seinen Mantel aus. Da mußte der Nordwind zugeben, daß die Sonne von ihnen beiden der Stärkere war.

Abbildung 10: Transkription der Aesop-Fabel auf Deutsch (Handbook of the IPA, 1999, S.89)

Eine phonemische Transkription des deutschen Textes sieht dann wie folgt aus:

/aɪnst ˈʃtrɪtɛn zɪç ˈnɔrtvɪnt ʊnt ˈzɔnɛ | veːr fɔn ˈiːnɛn ˈbaɪdɛn voːl deːr ˈʃtɛrkɛrɛ ˈvɛːrɛ | als aɪn ˈvandɛrɛr | deːr ɪn ˈaɪnɛn ˈvarmɛn ˈmantɛl gɛˈhʏlt vaːr | dɛs ˈveːgɛs daˈheːrˌkaːm ziː ˈvʊrdɛn ˈaɪnɪç | das ˈdeːrjeːnɪgɛ fyːr deːn ˈʃtɛrkɛrɛn ˈgɛltɛn ˈzɔltɛ | deːr deːn ˈvandɛrɛr ˈtsvɪŋɛn ˈvʏrdɛ | ˈzaɪnɛn ˈmantɛl ˈaptsuːˌneːmɛn deːr ˈnɔrtvɪnt bliːs mɪt ˈalɛr maxt | ˈaːbɛr jeː ˈmeːr eːr bliːs | ˈdɛstoː ˈfɛstɛr ˈhʏltɛ zɪç deːr ˈvandɛrɛr ɪn ˈzaɪnɛn ˈmantɛl aɪn ˈɛntlɪç gaːp deːr ˈnɔrtvɪnt deːn kampf aʊf nuːn ɛrˈvɛrmtɛ diː ˈzɔnɛ diː lʊft mɪt ˈiːrɛn ˈfrɔɪntlɪçɛn ˈʃtraːlɛn | ʊnt ʃoːn naːx ˈveːnɪgɛn aʊgɛnˈblɪkɛn tsoːk deːr ˈvandɛrɛr ˈzaɪnɛn ˈmantɛl aʊs daː ˈmʊstɛ deːr ˈnɔrtvɪnt ˈtsuːgeːbɛn | das diː ˈzɔnɛ fɔn ˈiːnɛn ˈbaɪdɛn deːr ˈʃtɛrkɛrɛ vaːr/

Abbildung 11: Phonemische Transkription der Aesop-Fabel auf Deutsch (Handbook of the IPA, 1999, S.88)

Für die Videoaufzeichnung wurde der letzte Satz der Fabel gewählt, um zum Einen eine Gleichverteilung der leicht erkennbaren Laute /p/, /b/, /f/, /v/ und /m/ zu gewährleisten und zum Anderen in jeder der drei Sprachen, die für den Versuch ausgewählt wurden, einen unterschiedlichen Satz zu verwenden.

Da musste der Nordwind zugeben, dass die Sonne von ihnen beiden der Stärkere war.
Plosiv: **b(2), p(0)**, t(5), d(6), k(1), g(1)
Frikativ: **f(1), v(2)**, s(3), z(1), ʃ(1), ç(0), j(0), h(0)
Affrikate: pf(0)
Nasal: **m(1)**, n(8), ŋ(0)
Vibrant: r(6)
Lateral: l(0)

Die Aesop-Fabel ist in verschiedenen Sprachen verfügbar und wurde in der Übersetzung jeweils so angepasst, dass weiterhin die phonetischen Merkmale einer Sprache abgebildet werden. Die englische Version lautet wie folgt:

The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shined out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.

Abbildung 12: Transkription der Aesop-Fabel auf Englisch (Handbook of the IPA, 1999, S.44)

Da das Englische andere Laute als die deutsche Sprache besitzt, sieht die entsprechende phonemische Transkription wie folgt aus:

ðə ˈnoɹθ ˌwɪnd ən (ð)ə ˈsʌn wɚ dɪsˈpjutɪŋ ˈwɪtʃ wəz ðə ˈstɹɑŋɚ, wɛn ə ˈtɹæveləɚ ˌkem əˈlɑŋ ˈɹæpt ɪn ə ˈwoɹm ˈkloʊk. ðe əˈɡɹid ðət ðə ˈwʌn hu ˈfɚst səkˈsidəd ɪn ˈmekɪŋ ðə ˈtɹæveləɚ ˈtek ɪz ˈkloʊk ˌɑf ʃʊd bi kənˈsɪdɚd ˈstɹɑŋɚ ðən ðɪ ˈəðɚ. ðɛn ðə ˈnoɹθ ˌwɪnd ˈblu əz ˈhɑɹd əz i ˈkʊd, bət ðə ˈmoɹ hi ˈblu ðə ˈmoɹ ˈkloʊsli dɪd ðə ˈtɹæveləɚ ˈfold hɪz ˈkloʊk əˈɹaʊnd ɪm; ˌæn ət ˈlæst ðə ˈnoɹθ ˌwɪnd ˌɡev ˈʌp ði əˈtɛmpt. ˈðɛn ðə ˈsʌn ˈʃaɪnd ˌaʊt ˈwoɹmli ənd ɪˈmidiətli ðə ˈtɹæveləɚ ˈtʊk ˌɑf ɪz kloʊk. ən ˈsoʊ ðə ˈnoɹθ ˌwɪnd wəz əˈblaɪdʒd tə kənˈfɛs ðət ðə ˈsʌn wəz ðə ˈstɹɑŋɚ əv ðə ˈtu.

Abbildung 13: Phonemische Transkription der Aesop-Fabel auf Englisch (Handbook of the IPA, 1999, S.44)

Wie bei der deutschen Version der Fabel wurde auch im Englischen jeder der Sätze untereinander verglichen, um einen Überblick über das Auftreten bestimmter Laute zu bekommen. Die Gleichverteilung der Indikatoren für Synchronität /p/, /b/, /f/, /v/ und /m/ konnte am ehesten mit dem zweiten Satz der Fabel gewährleistet werden.

They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other.

Plosiv: **b(1), p(0)**, t(5), d(6), k(6), g(2), θ(0), ð(7)

Frikativ: **f(2), v(1)**, s(4), z(1), ʃ(1), j(0), h(1), (ʒ)

Affrikate: pf(0), w(1), dʒ(0)

Nasal: **m(1)**, n(3), ŋ(2)

Vibrant: ɹ(3)

Lateral: l(2)

Als dritte und für die Teilnehmer des Versuchs unbekannte Sprache wurde Persisch gewählt. Der Text der Fabel und die phonemische Umschrift sind im Folgenden dargestellt.

Bod-e shomal va horshid davo mi-kardend ke kodomjek xavi-tar ast. dar-on vaxt mosafer-i resid ke shenel-e garm-i be doure-hodesh-pidshideh-bud. onho tavofox-kard-and kas-i-ke ovval mo-afax-shav-ad mosafer-ra madzbur-be daravardan-e shenel-ash konad bojad xavi-tar az digari darnazar gerefteh shavad. sepas bod-e shomal to haddi ke mitavanest vazid. ammo hartshe bishtar vazid mosafer shenel-ash-ra tangtar be-doure-hodash pitshid. va saeandzom ba-de shomal dast az talash bardasht. bazd horshid tshonan-garm tobid ke mosafer shenel-ash-ra fouri daravard. va banobanin bod-e shomol madzbur shod eteraf-konad ke horshid az beine an do xavi-tar ast.

bɒːde ʃomɒːl væ xorʃiːd dæʔˈvɒː mikærdænd ke kodɒːmjek Gævitær æst dær ɒːn væGt mosɒːferiː resiːd ke ʃenele gærmiː be dore xodæʃ piːtʃiːde bud ɒːnhɒː tævæfoG kærdænd kæsiː ke ævæl moæfæG ʃævæd mosɒːfer rɒː mædʒbur be dærɒːværdæne ʃenelæʃ konæd bɒːjæd Gæviː tær æz diːgæriː dær næzær gerefte ʃævæd sepæs bɒːde ʃomɒːl tɒː hædiː ke miːtævɒːnest væziːd æmɒː hærtʃe biːʃtær væziːd mosɒːfer ʃenelæʃ rɒː tængtær be dore xodæʃ piːtʃiːd væ særændʒɒːm bɒːde ʃomɒːl dæst æz tælɒːʃ bærdɒːʃt bæʔˈd xorʃiːd tʃenɒːn gærm tɒːbiːd ke mosɒːfer ʃenelæʃ rɒː foriː dærɒːværd væ bænɒːbæriːn bɒːde ʃomɒːl mædʒbur ʃod eterɒːf konæd ke xorʃiːd æz beijne ɒːndo Gævitær æst

Abbildung 14: (phonemische) Transkription der Aesop-Fabel auf Persisch ([1], Mina Fallahi)

Um auch weiterhin eine vergleichbare Anzahl an Indikatoren im Video darzubieten, wurde der vierte Satz der Fabel gewählt:

bazd horshid tshonan-garm tobid ke mosafer shenel-ash-ra fouri daravard.
Plosiv: **b(2), p(0)**, t(), d(), k(), g(),ʔ()
Frikativ: **f(2), v(1)**, s (), z (), (), j (), h (), (),x(),ɣ(),ɢ()
Affrikate: d- (), t- ()
Nasal: **m(1)**, n (), ()
Vibrant:- ()
Lateral: l ()

Die nachfolgende Tabelle (Tabelle 3) zeigt eine Übersicht über die Anzahl der Laute /p/, /b/, /f/, /v/ und /m/ in den drei Sprachen des Hörversuchs. Es sind nur /b/ und /m/ in allen Sprachen in gleicher Anzahl vertreten, während die Laute /p/, /f/ und /v/ unterschiedlich verteilt sind. Im Ganzen ergibt sich eine vergleichbare Anzahl an Indikatoren mit 6 im Deutschen, 5 im Englischen und 6 im Persischen, sodass die Probanden für ihre Entscheidungen die gleiche Basis an Indikatoren zur Verfügung haben.

	Deutsch	Englisch	Persisch
Plosiv p	2	1	2
Plosiv b	0	0	0
Frikativ f	1	2	2
Frikativ v	2	1	1
Nasal m	1	1	1

Tabelle 3: Vergleich der Anzahl an Indikatoren für Deutsch, Englisch und Persisch

Nachdem die Aufnahmen abgeschlossen waren, wurden die Audiodaten hinsichtlich Energie, durchschnittliche Sprecherfrequenz, Jitter und Shimmer verglichen, um eine Darbietung vergleichbarer auditiver Signale zu gewährleisten. Diese Faktoren wurden bereits vor der Versuchsdurchführung erfasst und ihr Einfluss auf die Versuchsteilnehmer durch eine vergleichbare Präsenz in allen Materialen kontrolliert. Die Analyse wurde mit dem Programm Praat [a] und den Audiodaten im wave-Format durchgeführt.

	Deutsch	Englisch	Persisch
Energie	$3{,}435*10^{-5}$ Joule/m^2	$5{,}742*10^{-5}$ Joule/m^2	$8{,}949*10^{-6}$ Joule/m^2
Sprecherfrequenz	226,004 Hz	223,016 Hz	229,371 Hz
Jitter	2,285 %	2,395 %	2,443 %
Shimmer	15,190 %	15,287 %	15,308 %

Tabelle 4: Vergleich für Energie, Frequenz, Jitter und Shimmer für Deutsch, Englisch und Persisch

Eine grafische Gegenüberstellung der Aufnahmen kann im Anhang 7.2 gefunden werden.

3.1.2 Videomaterial

Das Bildmaterial für den Versuch wurde im Tonstudio der Technischen Universität Berlin, EN324, mit einer Canon XF100 Kamera und einem Sennheiser Nieren-Mikrofon aufgezeichnet. An der Aufnahmen beteiligten sich drei Muttersprachlerinnen, die jeweils entweder den deutschen, englischen oder persischen Satz in normalem Sprechtempo vortrugen. Ein Vergleich der drei ausgewählten Sätze ergab, dass für das Vortragen des deutschen Satzes ca. 6 Sekunden, des englischen ca. 7 Sekunden und des persischen ca. 5 Sekunden benötigt werden. Die zeitlichen Unterschiede von bis zu zwei Sekunden wurden als insignifikant für die Entscheidungsfindung angesehen. Um später eine eindeutige Entscheidung über die Synchronität treffen zu können, wurde nur der Kopf-Schulter-Bereich aufgezeichnet, sodass Lippenbewegungen gut erkennbar und in einer zweckentsprechenden Größe abgebildet waren.

Für die Aufnahmen wurde vom Fachbereich Audiokommunikation eine Starterklappe zum Luftanregen zur Verfügung gestellt, sodass die Synchronität von Bild und Ton gewährleistet bzw. in späteren Bearbeitungen sicher gestellt werden konnte. Nach Beendigung der Aufzeichnung lag das Bildmaterial im MXF-Format, einem Material-eXchange-Format, unter Verwendung des MPEG-1/2 Video Codecs vor. Die Auflösung betrug 1440x1080 Pixel und die Abtastrate lag bei 25 Frames/Sekunde, was dem PAL-Standard entspricht. Mit dem Sennheiser Nieren-Mikrofon wurde nur einkanalig mit einer Abtastrate von 48000 Hz, 16 Bits/Sample und einer Bitrate von 768kb/s aufgezeichnet. Als Codec wurde PCM S16 LE (signed 16-bit little-endian) verwendet, welches ein kontinuierliches analoges Signal in ein diskretes digitales Signal umwandet.

Im Anschluss wurden alle drei Videos mithilfe der Videobearbeitungssoftware Sony Vegas Pro 10.0 [b] bearbeitet. Hier wurde zuerst die physikalische Synchronität durch Anpassen der Tonspur zur Klappenposition sicher gestellt, die Lautstärke normalisiert, anschließend die Klappe am Anfang herausgeschnitten und abschließend das Ergebnis gerendert. Dabei wurde darauf geachtet, dass die Video-Einstellungen beibehalten wurden. Diese Videos dienten in der jeweiligen Sprache als Original-Vorlage mit physikalisch synchroner Bild- und Tonspur. Im Anschluss wurden in jeder Sprache jeweils Videos mit einem Tonvorlauf von +120ms, +80ms und +40ms sowie einer Tonverzögerung von -40ms, -80ms, -120ms erstellt. Das Intervall von 40ms wurde gewählt, da dieser im PAL-Format genau einem Frame entspricht.

Für die Nutzung des Materials im erstellten Programm "Eiusdem Temporis" (siehe Kapitel 3.1.4), musste dieses allerdings im mkv-Format vorliegen. Beim mkv-Format handelt es sich um ein Containerformat des Matroska Projekts, Matroska Video. Die verlustlose Konvertierung wurde mit der freien Software "MediaCoder" [c] durchgeführt. Das vollständige Videomaterial bestand am Ende aus 21 Clips, jeweils 3 Videos in den 6 Zeitversätzen -120ms, -80ms, -40ms, +40ms, +80ms und +120ms sowie 3 physikalisch synchrone Fassungen.

3.1.3 Probanden

An dem Hörversuch haben insgesamt 30 Studenten der Technischen Universität Berlin teilgenommen. Davon waren 16 der Teilnehmer männlich und 14 weiblich. Das Alter der Teilnehmer lag zwischen 20 und 32 Jahren mit einem Mittelwert von 26,5 Jahren. Alle Probanden hatten normale oder korrigierte Sehstärke und es lagen keine Informationen über eventuelle Hörschäden vor.

Die Informanten kamen aus den Fachbereichen „Audiokommunikation und -technologie“, „Sprache und Kommunikation“ oder „Informatik“ und verfügten über kein spezielles Training oder Vorwissen für diesen Hörversuch. Voraussetzung für die Teilnahme war lediglich die deutsche Muttersprache, gutes Englisch und keine Kenntnisse der persischen Sprache.

3.1.4 Eiusdem Temporis

Als Benutzeroberfläche wurde das für diese Testreihe entwickelte Programm „Eiusdem Temporis“ [griech. „gleichzeitig] genutzt. Das Programm wurde vom TU Berlin Informatikstudenten Stefan Schütz in der Programmiersprache C# geschrieben und dient zur Wiedergabe des Videomaterials und der Speicherung der Probandenantworten. „Eiusdem Temporis“ ermöglicht es dem Versuchsleiter das Originalvideo und die dazugehörigen Testreize in eine Liste zu laden und die Anzahl der Trials selbst festzulegen.

Zu Beginn jeder Programmdurchführung wird eine zufällige Liste mit Original-Testreiz-Kombinationen erstellt und nach jeder Antwort des Teilnehmers ein mitlaufender Zähler für die entsprechende Kombination erhöht. Dies stellt sicher, dass nach Erreichen der vorgegebenen Trial-Anzahl die entsprechende Kombination nicht noch einmal dargeboten wird. Im Folgenden ist eine mögliche Darstellung des internen Programmstatus für die Kombinationen der deutschen Fassung bei 8 Trials dargestellt.

Deutsch_+120ms (1/8)
Deutsch_+80ms (0/8)
Deutsch_+40ms (3/8)

Deutsch_Original +

Deutsch_-40ms (1/8)
Deutsch_-80ms (4/8)
Deutsch_-120ms (0/8)

Abbildung 15: möglicher Progammstatus - "Eiusdem Temporis"

Nachdem vom Probanden allgemeine Daten wie Geschlecht, Alter und ein Kürzel zur Anonymisierung eingegeben wurden, kann der Hörversuch beginnen.

Um den Versuchsteilnehmer nicht abzulenken, werden die Videos nacheinander auf der gesamten Programmfläche statt nebeneinander wiedergegeben. Dies nutzt auch die Größe und Auflösung besser aus, sodass Teilnehmer einen guten Blick auf die Lippenbewegungen haben.

Nach dem Abspielen von zwei Videos, einem physikalisch synchronen Video und einem sprachlich zugehörigen Testreiz, wird der Teilnehmer aufgefordert eine Entscheidung darüber zu treffen, welches der beiden Videos physikalisch synchron war. Dies geschieht über ein Pop-up-Fenster. Ein Fortsetzen des Versuchs ist erst nach einer Entscheidung möglich. Die Antworten werden in Excel-Tabellen abgespeichert und mit dem vom Probanden ausgewählten Kürzel versehen.

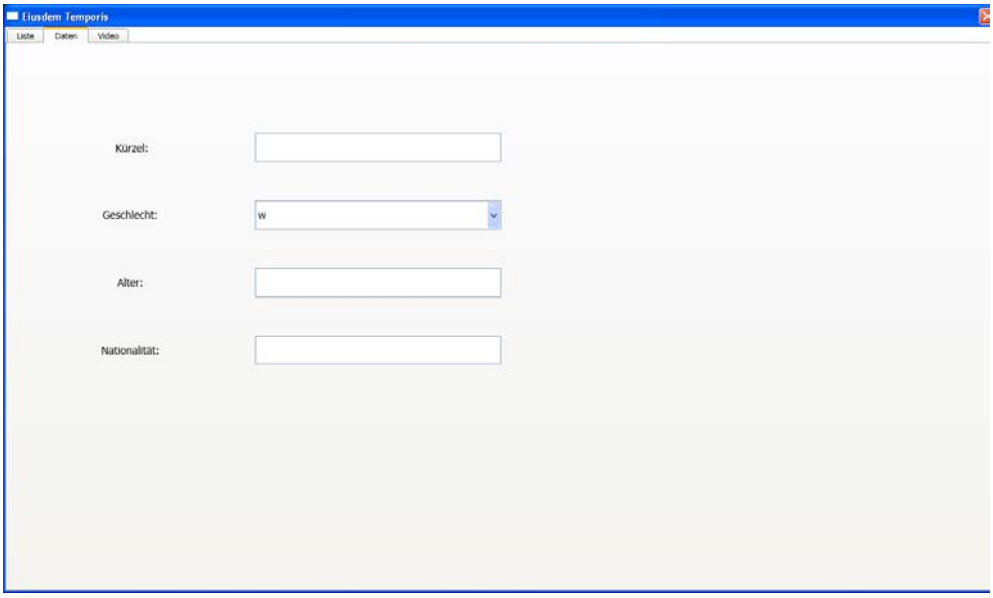

Abbildung 16: Screenshot des Programms "Eiusdem Temporis"

Das Programm wurde mehreren Testphasen unterzogen, um sicher zu stellen, dass sowohl die Darbietung vom deutschen, englischen und persischen Material also auch von Original und Testreiz in zufälliger Reihenfolge erfolgt, Einstellungen schnell und einfach erfolgen können und die Antworten korrekt abgespeichert werden.

Das Endergebnis benötigt zum erfolgreichen Einsatz eine Windows-Benutzeroberfläche, mindestens den Media Player 10.0 sowie mindestens das Microsoft .Net Framework 3.5. Das lauffähige Programm kann auf der beigelegten CD als zip-Datei gefunden werden.

3.2 Versuchsaufbau

Ein wichtiger Punkt für die Reliabilität der Daten, die während des Versuchs gesammelt werden sollten, war die Frage der physikalischen Bild-Ton-Synchronität am Hörerort. Um diese zu gewährleisten, wurde in einem Vorversuch die Latenz des Eintreffens von Bild und Ton am Sitzplatz des Versuchsteilnehmers mithilfe eines Oszilloskops gemessen.

Es handelte sich dabei um das Digital Oszilloskop HP Agilent 54624A, welches von Fachbereich der TU Berlin für die Dauer der Messungen zur Verfügung gestellt wurde. Auf zwei Kanälen konnten jeweils das Eintreffen des akustischen Signals und die Intensitätsänderungen durch einen Schwarz-Weiß-Wechsel des visuellen Signals aufgezeichnet werden. Als visuelles Signal diente die nachbearbeitete deutsche Aufzeichnung inklusive der Klappe. Das Video wurde für diesen Zweck so geschnitten, dass das Bild synchron zum Klappenschlag vom schwarzen Hintergrund zum aufgezeichneten Video wechselte. Für die Messung wurde dieses Video als Original und als Testreiz in "Eiusdem Temporis" geladen.

Abbildung 17: Versuchsaufbau im MiniLab der TU Berlin

Nachdem die physikalische Synchronität durch die Bearbeitung mit Vegas Pro 10.0 auf dem verwendeten Samsung Q45 Tyreen-Notebook sichergestellt wurde, mussten durch die weitere Übertragungskette entstehenden Abweichungen berücksichtigt werden. Dazu zählen das Abspielen des Videos im Programm "Eiusdem Temporis", die Auffrischungsrate des Monitors und die akustische Signalübertragung über die angeschlossene Fostex 6301B Box. Der Einfluss der Luftübertragung ist in dem genutzten MiniLab der TU Berlin zwar gering, kann aber trotzdem zu Abweichungen der Bild-Ton-Synchronität führen und musste somit als Punkt in der Übertragungskette bedacht werden.

Um alle Faktoren einzuschließen und die physikalische Synchronität am Hörerort gewährleisten zu können, wurde mithilfe eines Kondensator-Mikrofons das akustische Signal und mithilfe eines Fototransistors die Helligkeitsschwankung beim Bildwechsel erfasst. Ein Schaltplan für die genutzte Konstruktion ist im Unterpunkt 7.4 im Anhang zu finden. Die zeitliche Abweichung der zwei Signale konnte anschließend auf dem Oszilloskop überprüft werden.

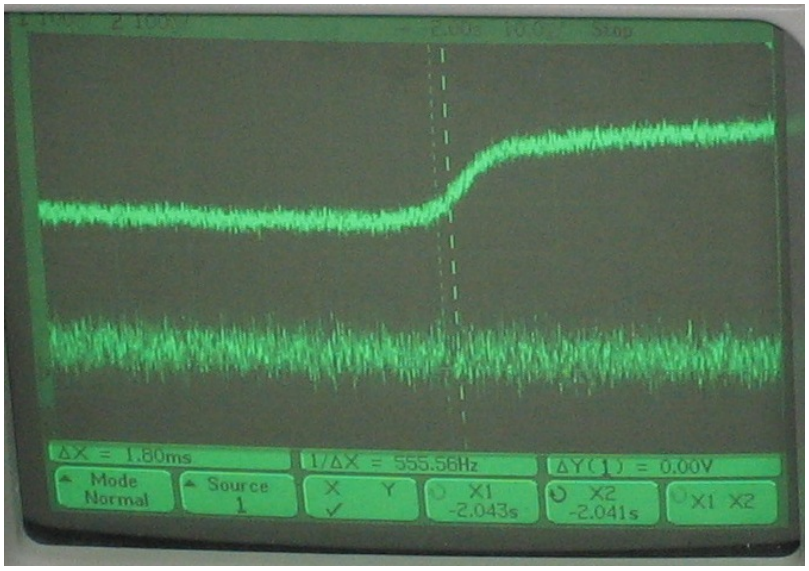

Abbildung 18: Oszilloskopbessung

Auf der obigen Grafik (Abbildung 18) sind beide Signaleingänge zu erkennen. Während im ersten Kanal die Helligkeitsschwankung von schwarz auf weiß durch eine Anhebung der Spannungsanzeige deutlich gemacht wurde, musste beim auditiven Signal auf eine durch die Klappe ausgelöste starke Amplitude geachtet werden. Die gestrichelten senkrechten Linien stellen Cursoreinstellungen des Oszilloskops dar, mit deren Hilfe das Ablesen der zeitlichen Latenz der jeweiligen Ausschläge in den zwei Kanälen möglich ist.

Die erste Messreihe mit 20 Messungen ergab eine durchschnittliche Abweichung von 77,38ms, die der Ton zu früh eintraf. Nach mehrfacher Anpassung des Testvideos und mehreren Messreihen konnte die Abweichung auf 75,5ms genau bestimmt werden. Eine abschließende Testreihe ergab eine Schwankung von +2ms<0ms<-2ms, die der Ton vor bzw. nach dem Bild am Hörerort eintrifft. Diese Abweichung wird für den Versuch als akzeptabel und vernachlässigbar eingestuft.

3.3 Versuchsdurchführung

Nach erfolgreicher Messung und Korrektur der durch die Übertragungskette entstandenen Latenz wurde in einem Vorversuch erneut die Videowiedergabe und die Datenerfassung durch das Programm "Eiusdem Temporis" überprüft. Letzte Fehler in der randomisierten Wiedergabe der verschiedenen Sprachen wurden dabei korrigiert, um sicher zu stellen, dass eine Sprache nicht zu häufig wiederholt wird. Dies würde beim Probanden zum Nachlassen der Konzentration führen und einen Gewöhnungseffekt bewirken.

Für die Versuchsdurchführung stand das MiniLab im Hauptgebäude der Technischen Universität Berlin zur Verfügung. Hierbei handelt es sich um einen akustisch trockenen Raum, welcher für solche und ähnliche Hörversuche genutzt wird. Im Versuchsraum wurde wie im Vorversuch das Bildmaterial mithilfe des Programms "Eiusdem Temporis" auf einem Samsung Q45 Aura Tyreen-12"- Notebook abgespielt und das Mono-Signal über eine Fostex 6301B Box wiedergegeben, welche sich leicht erhöht über dem Monitor befand.

Probanden konnten sich über einen doodle-Kalender für jeweils einstündige Times lots anmelden und erhielten nach Abschluss eine Unterschrift für das Seminar "Einführung in das empirische Arbeiten" für ihre Teilnahme.

Vor Beginn des Versuchs erhielten die Teilnehmer eine kurze Einführung in den Umgang mit dem Programm durch den Versuchsleiter, welcher sich während der Durchführung in einem Nebenraum befand und bei eventuellen Fragen oder Unklarheiten zur Verfügung stand. Die Probanden wurden über die Art jedoch nicht die Forschungsfrage des Versuches informiert, um ihre Aufmerksamkeit gezielt auf die Synchronität des dargebotenen Materials zu richten.

Der Versuchsablauf war für alle Probanden gleich und beinhaltete das Abspielen von jeweils zwei Videos - physikalisch synchrone Fassung und zeitlich verschobene Fassung - bevor über ein Pop-up-Fenster die Entscheidung getroffen werden musste, welches von den beiden Videos physikalisch synchron war. Erst nach dem Beantworten der Frage, konnte mit dem nächsten Videopaar fortgesetzt werden. Die Reihenfolge der Sprachen und Zeitversätze war dabei für jeden Probanden unterschiedlich.

Der Zeitaufwand für die Versuchsteilnehmer ergab sich aus der Abspielzeit der einzelnen Videopaare, multipliziert mit der Anzahl der Zeitversätze, multipliziert mit der Trial-Anzahl und plus der individuellen Dauer für jede Entscheidung. Im Durchschnitt wurden 45 min benötigt:

e = Englisch = 7 s
d = Deutsch = 6 s
p = Persisch = 5 s
x = Entscheidungsdauer in [s]

$$t(x) = (2 * e + 2 * d + 2 * p) * 6[Zeitversätze] * 8[Trials] + x$$
$$t(x) = 28min48s + x$$

Der Versuch fand an Tagen vom 25. 09. 2012 bis 30. 10. 2012 jeweils zwischen 10:00 bis 17:00 Uhr in dem genannten MiniLab der TU Berlin statt.

3.4 Analyse

Nach Beendigung des Versuchs lag von jedem Teilnehmer eine mit seinem anonymisierten Kürzel versehende Excel-Datei vor, welche die Entscheidungen für die einzelnen Videopaare enthielt. Ein kompletter Datensatz bestand aus 144 Einträgen "Richtig" oder "Falsch", welche im Folgenden nach Sprache und Versatz sortiert wurden, sodass eine Übersicht über die Anzahl korrekter Antworten für jeden Zeitversatz entstand. Die maximale Anzahl hing von der vom Versuchsleiter eingestellten Trialanzahl ab und lag für diesen Hörversuch bei 8 Wiederholungen.

Im weiteren Verlauf wurden die ermittelten prozentualen Werte an richtigen Antworten ins Statistikprogramm SPSS Version 15.0 [d] übertragen. Folgend ist ein Auszug der entsprechenden Tabellen zu sehen:

Original_Persisch_00.mkv	Persisch_-040.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-040.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-040.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-040.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-040.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-040.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-040.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-040.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-080.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-080.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-080.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-080.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-080.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-080.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-080.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-080.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-120.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-120.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-120.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-120.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-120.mkv	FALSCH
Original_Persisch_00.mkv	Persisch_-120.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-120.mkv	Richtig
Original_Persisch_00.mkv	Persisch_-120.mkv	Richtig

Sprache	Versatz	Anzahl_richtig	Prozent
p	-40	5	62,5
p	-80	3	37,5
p	-120	4	50

Abbildung 19: Datentransformation von Excel zu SPSS

Die folgende Auswertung konzentrierte sich vor allem darauf ob und wie sich die Fehlerurteile zwischen den Zeitversätzen und den drei Sprachen veränderten. Hierfür wurde mit Hilfe von SPSS Mittelwertvergleiche durch eine einfaktorielle ANOVA (Varianzanalyse) unter Berücksichtigung verschiedener Faktoren wie der Sprache, den Zeitversätzen und im weiteren Verlauf auch dem Geschlecht der Versuchsteilnehmer durchgeführt. Da im Versuch drei Sprachen zum Einsatz kamen, konnte kein t-Test für die Auswertung durchgeführt werden, welcher der ANOVA mit zwei Gruppen entspricht. Die hier verwendete univariate Varianzanalyse ermittelt den Einfluss einer unabhängigen Variable auf eine abhängige Variable durch einen Mittelwertvergleich (Eckstein, 2012, S.127). Für diesen Versuch wurde der Einfluss der Videosprache auf das Erkennen der physikalischen synchronen Fassung untersucht.

Um eine Aussage über die Lage des Punktes subjektiver Synchronität (PSS) und den gerade merklichen Unterschied (JND) treffen zu können, wurden weiterhin die prozentualen Fehlurteile für jeden Zeitversatz und jeden Probanden mithilfe des Programms Matlab R2010a [e] analysiert. Im folgenden Kapitel 4 werden die Ergebnisse der Untersuchung anhand von Grafiken dargestellt und im Vergleich mit anderen Studien diskutiert.

4. Ergebnisse

Nachdem im vorherigen Kapitel der Versuchsaufbau und die Durchführung beschrieben wurden, werden in diesem Kapitel die aus der Datenanalyse gewonnenen Ergebnisse grafisch dargestellt und Interpretations- und Erklärungsansätze gebildet, welche abschließend zu der folgenden Zusammenfassung der Arbeit im Kapitel 5 überleiten.

4.1 Darstellung der Ergebnisse

Die Fragestellung für die vorliegende Masterarbeit konzentrierte sich auf den Einfluss der Sprache auf die Wahrnehmung von Asynchronität. Um dieser Frage nachzugehen, wurde mit 30 Probanden ein Hörversuch an der TU Berlin durchgeführt. Es wurden in mehreren Trials Videopaare mit unterschiedlichen zeitlichen Latenzen in drei verschiedenen Sprachen dargeboten und Probanden entschieden jeweils, welches Video physikalisch synchron war.
Für die Datenanalyse und Darstellung der Ergebnisse wurden die Programme SPSS 15.0, Matlab R2010a und Microsoft Office Excel 2007 verwendet. Das Niveau des α-Fehlers wurde dabei auf $\alpha = 0{,}005$ gesetzt. "d", "e" und "p" stehen im weiteren Verlauf jeweils für die Sprachen Deutsch, Englisch und Persisch.

In einer ersten Auswertung wurde die Anzahl der richtig abgegebenen Synchronurteile für die unterschiedlichen zeitlichen Versätze und Sprachen zusammengefasst. Die maximal mögliche Anzahl von 240 Antworten pro Zeitversatz wäre eingetroffen, wenn jeder der 30 Versuchsteilnehmer diesen Versatz in allen 8 Trials richtig erkannt hätte. In der folgenden Grafik (Abbildung 20) ist zu erkennen, dass die Antworttendenz unterschiedlich ausgefallen ist, an den äußereren Grenzen (+120ms und -120ms) allerdings die synchrone Version der deutschen Sprache sicherer erkannt wurde als in Englisch und Persisch. Für alle Sprachen sind ein Anstieg der fehlerhaften Synchronurteile und ein Absinken der richtigen Urteile zwischen einer Tonvoranstellung von +40ms und einer Verzögerung von -80ms zu verzeichnen. Dies deutet darauf hin, dass das Erkennen der physikalisch synchronen Fassung in diesem Zeitbereich für alle Sprachen gleichermaßen schwer war. Die nachfolgende Tabelle (Tabelle 5) zeigt die Summe korrekter Synchronurteile für alle Sprachen und Zeitversätze und bildete die Basis für das folgende Balkendiagramm (Abbildung 20).

Statistiken

richtig_metrisch

versatz			Sprache_Video d	e	p
-120	N	Gültig	30	30	30
		Fehlend	0	0	0
	Summe		168,00	148,00	165,00
-80	N	Gültig	30	30	30
		Fehlend	0	0	0
	Summe		138,00	126,00	140,00
-40	N	Gültig	30	30	30
		Fehlend	0	0	0
	Summe		122,00	129,00	126,00
40	N	Gültig	30	30	30
		Fehlend	0	0	0
	Summe		133,00	129,00	142,00
80	N	Gültig	30	30	30
		Fehlend	0	0	0
	Summe		166,00	172,00	156,00
120	N	Gültig	30	30	30
		Fehlend	0	0	0
	Summe		206,00	186,00	186,00

Tabelle 5: Summe korrekter Synchronurteile für Deutsch (d), Englisch (e) und Persisch (p)

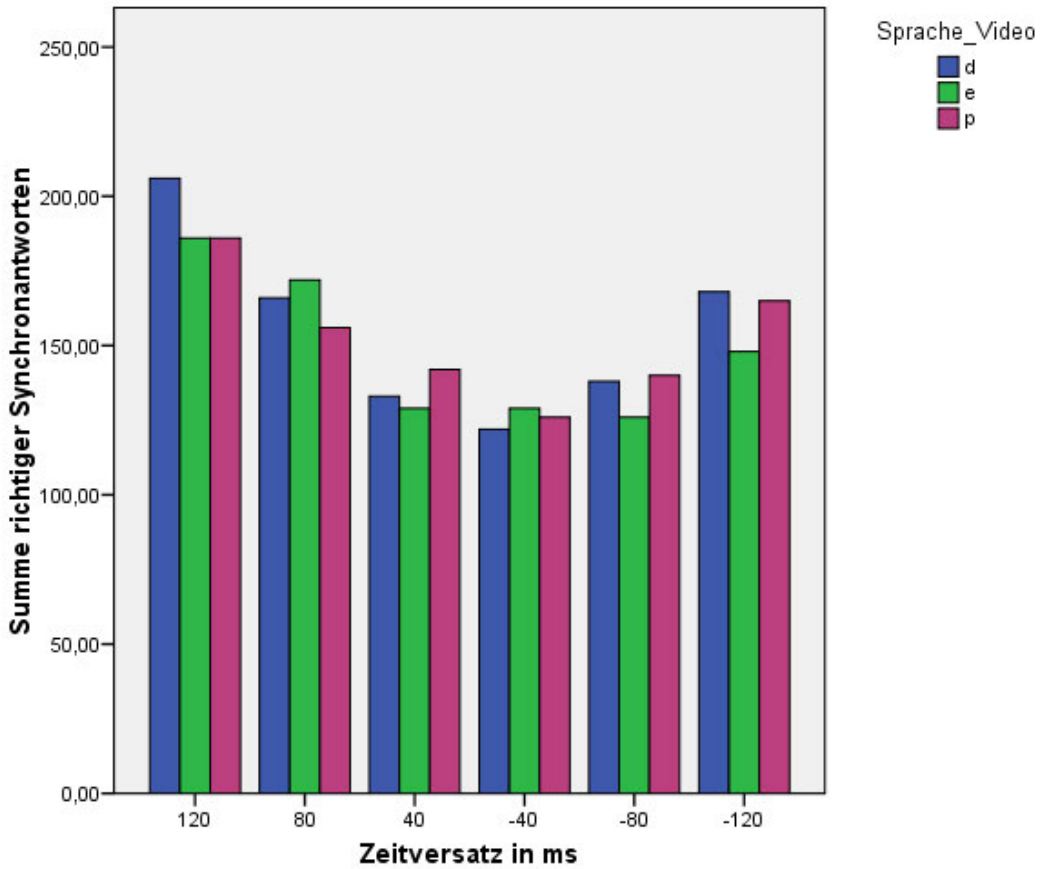

Abbildung 20: Balkendiagramm der summierten korrekten Synchronurteile für Deutsch (d), Englisch (e) und Persisch (p)

Eine genauere Unterteilung der Ergebnisse erlaubt eine Analyse der Häufigkeitsverteilung im Bezug auf die Ratewahrscheinlichkeit. Für den Hörversuch wurde ein Forced-Choice-Verfahren verwendet, bei dem der Teilnehmer sich zwischen zwei Stimuli entscheiden muss. In diesem Fall liegt die Ratewahrscheinlichkeit bei 50 %. Daraus kann geschlussfolgert werden, dass das richtige Erkennen der synchronen Fassung in weniger oder gleich 4 von 8 Trials mit dem Raten durch den Probanden gleichgesetzt werden kann und ebenso die prozentualen Fehlurteile 50% nicht überschreiten können. In der untenstehenden Tabelle (Tabelle 6) sind die Häufigkeiten der richtigen Synchronurteile nach Trialanzahl sortiert aufgelistet und werden im Folgenden an zwei Beispielen grafisch (Abbildung 21 und 22) dargestellt. Alle Entsprechenden Grafiken können auf der beiliegenden CD gefunden werden.

Versatz	Sprache	Häufigkeiten richtiger Synchronurteile								
		1	2	3	4	5	6	7	8	Mittelwert
-120	d	0	0	1	4	8	11	5	1	5,6
	e	0	1	3	8	7	7	4	0	4,93
	p	0	1	3	5	3	10	5	3	5,5
-80	d	0	2	4	9	6	7	2	0	4,6
	e	0	4	8	7	4	3	4	0	4,3
	p	0	1	6	8	5	7	3	0	4,67
-40	d	0	4	7	6	10	2	1	0	4,06
	e	0	1	7	10	7	4	1	0	4,3
	p	0	3	6	9	6	6	0	0	4,2
40	d	0	3	3	9	10	3	2	0	4,43
	e	2	3	0	13	6	3	3	0	4,3
	p	0	2	3	8	9	5	2	1	4,73
80	d	0	1	2	6	4	7	8	2	5,53
	e	0	0	1	5	7	9	4	4	5,73
	p	0	1	1	6	12	5	3	2	5,2
120	d	0	1	1	1	2	2	9	14	6,86
	e	0	0	1	4	5	4	10	6	6,2
	p	0	2	0	0	4	10	10	4	6,2

Tabelle 6: Exceltabelle der Häufigkeiten korrekter Synchronurteile, inkl. der Mittelwerte

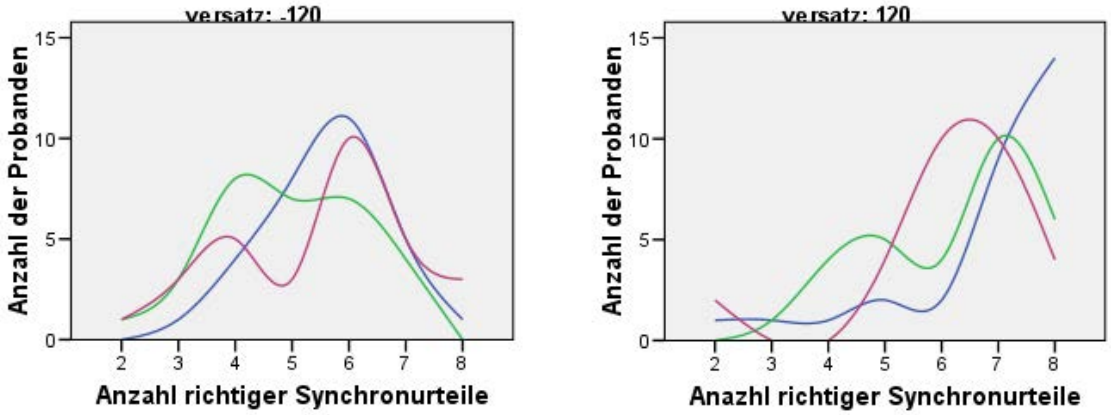

Abbildung 21: Häufigkeitsverteilung richtiger Synchronurteile für die Zeitversätze -120ms und +120ms; blau = Deutsch, grün = Englisch, rot = Persisch

Für die Interpretation der Grafiken 21 und 22 ist zu beachten, dass auf der x-Achse jeweils die korrekten Synchronurteile bei einer Maximalanzahl von 8 Trials pro Versatz aufgetragen sind, während die y-Achse die Anzahl der Probanden widerspiegelt, die eine entsprechende Anzahl im Paarvergleich als physikalisch synchron korrekt erkannt hat. Dabei stellt die blaue Linie die deutsche, die grüne Linie die englische und die rote Linie die persische Version dar.

In einem Vergleich der beiden maximalen Zeitversätze für diesen Hörversuch ist leicht ersichtlich, dass eine Tonvoranstellung von +120ms in allen Sprachen besser erkannt wurde als die Verzögerung um -120ms. In beiden Fällen gilt zusätzlich, dass die deutsche Version durchschnittlich mit größerer Sicherheit erkannt wurde als die englische oder persische. Im Gegensatz zu den maximalen Zeitversätzen wird eine zeitliche Verzögerung von -40ms in Englisch (Mittelwert: 4,3) und Persisch (Mittelwert: 4,2) durchschnittlich sicherer erkannt als im Deutschen (Mittelwert: 4,06).

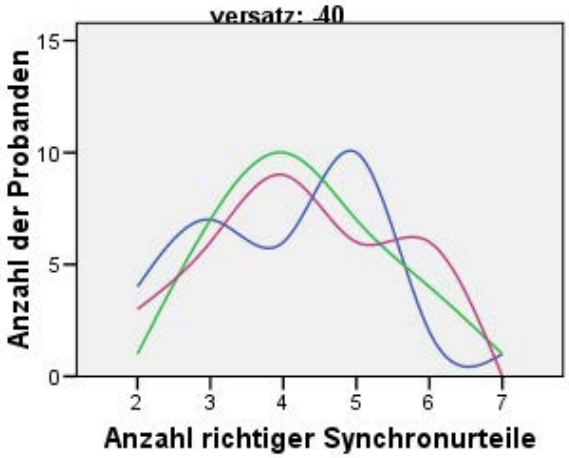

Abbildung 22: Häufigkeitsverteilung richtiger Synchronurteile für den Zeitversaz -40ms; blau = Deutsch, grün = Englisch, rot = Persisch

Im Anschluss an die Häufigkeitsanalyse wurde ein Vergleich der Mittelwerte für jede Sprache und die zeitlichen Latenzen durchgeführt. Die Mittelwerte beziehen sich dabei auf die prozentualen Fehlurteile während des Hörversuchs. Wie aus der Tabelle (Tabelle 7) abzulesen ist, liegen die maximalen inkorrekten Synchronurteile zwischen einer Tonverzögerung von -40ms und -80ms.

	d	e	p
-120	30	38,333	31,25
-80	42,5	47,5	41,666
-40	49,167	46,25	47,5
0	50	50	50
40	44,583	46,25	40,833
80	30,833	28,333	35
120	14,167	22,5	22,5

Tabelle 7: Exceltabelle der Mittelwerte fehlerhafter Synchronurteile in Prozent

Weiterhin kann erneut gezeigt werden, dass die Randbereiche des Hörversuchs (+120ms, -120ms) im Deutschen eine geringere Fehlerquote annehmen, als in den anderen beiden Sprachen. Im Vergleich zu Deutsch und Persisch kann zusätzlich eine Verschiebung des Maximums von der Tonverzögerung -40ms auf -80ms für die englische Version gezeigt werden.

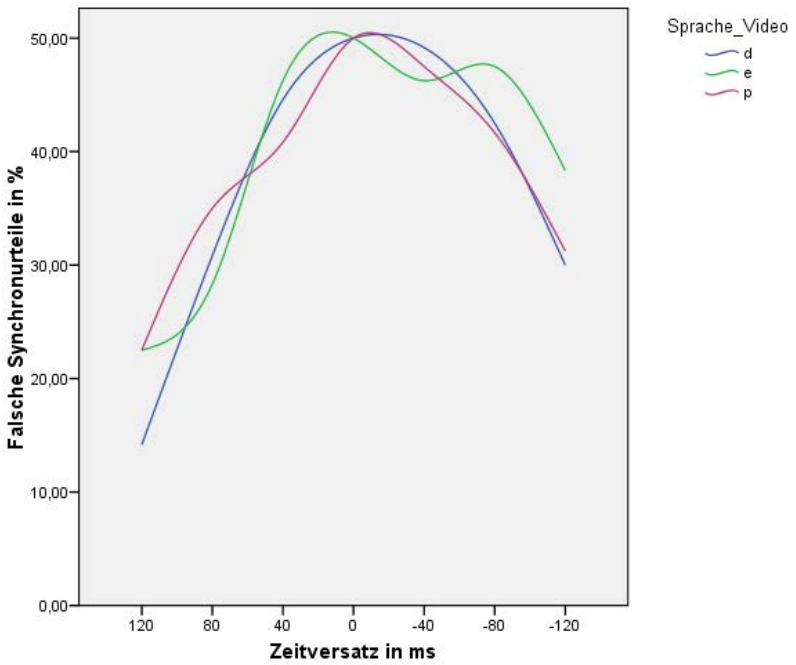

Abbildung 23: Liniendiagramm Mittelwerte fehlerhafter Synchronurteile in Prozent

Mithilfe des Statistik-Programms SPSS wurde mithilfe der aus Excel übertragenden Daten eine einfaktorielle ANOVA auf Grundlage der der prozentualen Fehlurteile durchgeführt, um die statistische Signifikanz der Urteile zu bestimmen. Bei der einfaktoriellen ANOVA handelt es sich um eine Varianzanalyse, welche den Einfluss einer Variable durch einen Mittelwertvergleich untersucht (Bortz, 2005, Kapitel 7). Überprüft wurde der Einfluss der Sprache auf die korrekten Synchronurteile für jeden Zeitversatz. Für die Analyse wurde das Signifikanzniveau auf 5 % festgelegt. Im tabellarischen und grafischen Vergleich der Mittelwerte waren unterschiedlich starke Abweichungen zwischen den Sprachen erkennbar dargestellt. In allen drei Sprachen ist ein deutlicher Rückgang der fehlerhaften Synchronurteile für den Bereich der Tonvoranstellung von +40ms, +80ms und +120ms zu bemerken. Die höchsten fehlerhaften Synchronurteile weisen die zeitlichen Abweichungen von +40ms, -40ms und -80ms mit über 40 % auf (Tabelle 7 und Abbildung 23).

Ein statistischer Vergleich konnte allerdings keine signifikanten Unterschiede zwischen den Sprachen bestätigen. Für die Tonverzögerung von -120ms konnte im Mittelwertvergleich ein Signifikanzwert von p = 0,135 (F = 2,050) ermittelt werden, welcher über dem Signifikanzniveau von p = 0,05 liegt. Ein signifikanter Unterschied konnte zwischen den Mittelwertgruppen der drei Sprachen nicht nachgewiesen werden.

ONEWAY ANOVA

prozent

versatz		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
-120	Zwischen den Gruppen	1211,806	2	605,903	2,050	,135
	Innerhalb der Gruppen	25713,542	87	295,558		
	Gesamt	26925,347	89			
-80	Zwischen den Gruppen	597,222	2	298,611	,900	,410
	Innerhalb der Gruppen	28854,167	87	331,657		
	Gesamt	29451,389	89			
-40	Zwischen den Gruppen	128,472	2	64,236	,261	,771
	Innerhalb der Gruppen	21401,042	87	245,989		
	Gesamt	21529,514	89			
40	Zwischen den Gruppen	461,806	2	230,903	,716	,492
	Innerhalb der Gruppen	28052,083	87	322,438		
	Gesamt	28513,889	89			
80	Zwischen den Gruppen	680,556	2	340,278	1,064	,350
	Innerhalb der Gruppen	27833,333	87	319,923		
	Gesamt	28513,889	89			
120	Zwischen den Gruppen	1388,889	2	694,444	1,962	,147
	Innerhalb der Gruppen	30791,667	87	353,927		
	Gesamt	32180,556	89			

Tabelle 8: Ergebnisse der einfaktoriellen ANOVA

In einem Post-Hoc-Test wurden zusätzlich paarweise Mittelwertvergleiche zwischen den Gruppen für jeden Zeitversatz durchgeführt, um einen Einfluss oder eine Tendenz zu ermitteln. SPSS bietet die Möglichkeit unterschiedliche Post-Hoc-Tests durchzuführen. Dazu zählen unter anderem der Tukey-HSD-Test, welcher einen paarweisen Vergleich aller berechneten Mittelwerte erlaubt, und der LSD-Test, welcher paarweise t-Tests durchführt. Der Tukey-HSD-Test wird nach Reiter (2005) dann signifikant, wenn der Unterschied zwischen zwei Mittelwerten x_i und y_i größer ist als der Honestly-Significant-Difference-Wert, welcher in diesem Fall mit $p < 0{,}05$ signifikant wäre. Die vorhergehende einfaktorielle ANOVA konnte keinen signifikanten Spracheinfluss nachweisen, wodurch auch in einem nachfolgenden Post-Hoc-Test nur Tendenzen und keine signifikanten Unterschiede für den Tukey-Test ermittelt werden konnten. Einzig der LSD-Test (least significant difference) liefert schwach signifikante Werte bei der Bewertung deutscher und englischer Synchronität für einen Zeitversatz von -120ms ($p = 0{,}064$) und für die Vergleichspaare Deutsch-Englisch und Deutsch-Persisch für einen Tonvorzug von +120ms ($p_{1,2} = 0{,}09$). Eine Auflistung der Signifikanzwerte kann in Verbindung mit weiteren Post-Hoc-Tests auf der beigelegten CD im Ordner "Hörversuch_Daten" gefunden werden.

Um die Ergebnisse abzusichern, wurden mit Hilfe des Kolmogorov-Smirnov-Anpassungstests die Daten auf Normalverteilung überprüft. Eine Normalverteilung liegt dann vor, wenn die Ergebnisse des Tests keine signifikanten Abweichungen aufweisen. Wie aus der nachflgenden Tabelle (Tabelle 9) zu entnehmen ist, weichen nur die Beurteilungen des englischen Materials bei einem Zeitversatz von +40ms und die deutsche Fassung bei +120ms von der Gaußschen Normalverteilung ab.

Kolmogorov-Smirnov-Anpassungstest

			Parameter der Normalverteilung[a,b]		Extremste Differenzen				
versatz		N	Mittelwert	Standardabweichung	Absolut	Positiv	Negativ	Kolmogorov-Smirnov-Z	Asymptotische Signifikanz (2-seitig)
-120	deutsch_proz	30	70,00	14,157	,205	,162	-,205	1,121	,162
	eng_proz	30	61,67	16,717	,157	,157	-,154	,862	,447
	pers_proz	30	68,750	20,1690	,222	,124	-,222	1,214	,105
-80	deutsch_proz	30	57,50	16,930	,171	,171	-,149	,937	,343
	eng_proz	30	52,50	20,075	,183	,183	-,102	1,002	,268
	pers_proz	30	58,333	17,4733	,183	,183	-,163	1,004	,266
-40	deutsch_proz	30	50,83	16,391	,195	,159	-,195	1,068	,204
	eng_proz	30	53,75	14,735	,200	,200	-,133	1,098	,179
	pers_proz	30	52,500	15,8794	,163	,163	-,137	,890	,406
40	deutsch_proz	30	55,42	16,309	,170	,165	-,170	,931	,352
	eng_proz	30	53,75	19,737	,258	,175	-,258	1,413	,037
	pers_proz	30	59,167	17,6573	,158	,158	-,142	,868	,438
80	deutsch_proz	30	69,17	19,623	,184	,136	-,184	1,005	,264
	eng_proz	30	71,67	17,036	,156	,156	-,144	,853	,460
	pers_proz	30	65,000	16,8666	,226	,226	-,174	1,236	,094
120	deutsch_proz	30	85,83	19,896	,300	,238	-,300	1,643	,009
	eng_proz	30	77,50	18,395	,240	,126	-,240	1,314	,063
	pers_proz	30	77,500	18,0993	,245	,157	-,245	1,342	,054

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Tabelle 9: Ergebnisse des Kolmogorov-Smirnov-Anpassungstest für Normalverteilung

Da die Varianzanalyse sehr robust gegenüber Nicht-Normalverteilung ist, kann vom korrekten Einsatz der einfaktoriellen ANOVA ausgegangen werden.

Eine starke Abweichung der Daten von der Normalverteilung würde den Einsatz des parameterfreien Friedman-Tests zum Bestätigen signifikanter Unterschiede erfordern, welcher zur Kontrolle der bisher gewonnen Ergebnisse zusätzlich durchgeführt wurde. Der Friedman-Test wird dann signifikant ($p < 0,05$), wenn sich der ermittelte Medianwert einer Stichprobe signifikant von denen der anderen Stichproben unterscheidet (Leard Statistics, 2013). Im Fall dieses Hörversuchs trifft dies auf den Zeitversatz von +120ms ($p = 0,017$) zu.

Statistik für Test[a]

versatz	N	Chi-Quadrat	df	Asymptotische Signifikanz
-120	30	4,102	2	,129
-80	30	3,822	2	,148
-40	30	,433	2	,805
40	30	,699	2	,705
80	30	2,116	2	,347
120	30	8,143	2	,017

a. Friedman-Test

Tabelle 10: Ergebnisse des parameterfreien Friedman-Tests

Um eine Aussage darüber treffen zu können, welche der drei Sprachen signifikante Unterschiede hinsichtlich der korrekten Synchronurteile aufweist, wurde anschließend in einem Post-Hoc-Test der parameterfreie Wilcoxon-Test mit Bonferroni-Anpassung durchgeführt. Dies bedeutet, dass durch die mehrfachen Variablenvergleiche das Signifikanzniveau entsprechend der Anzahl von Paarvergleichen angepasst werden muss. Für die Auswertung der Daten wurden jeweils die korrekten Synchronurteile der deutschen, englischen und persischen Fassung miteinander vergleichen, was einer Division des Signifikanzniveaus durch 3 entspricht (Leard Statistics, 2013). Bei der Auswertung der Testergebnisse muss daher ein Signifikanzniveau von $p<0,033$ für schwach signifikante Unterschiede und $p<0,017$ für signifikante Unterschiede beachtet werden. Ein schwach signifikanter Unterschied kann im Folgenden für die deutschen und englischen Synchronurteile bei einer Tonverzögerung von -120ms bestätigt werden ($Z = -2,134$, $p = 0,033$). Die Beurteilung der deutschen und englischen Fassung ($Z = -2,220$, $p = 0,026$) und der deutschen und persischen Fassung (Z = -

2,270, p = 0,023) weisen bei einem Zeitversatz von +120ms einen schwach signifikanten Unterschied auf. Diese Ergebnisse bestätigen die Tendenzen des LSD-Post-Hoc-Tests, welcher im Anschluss an die einfaktorielle ANOVA durchgeführt wurde.

Statistik für Test[c]

versatz		eng_proz - deutsch_proz	pers_proz - deutsch_proz	pers_proz - eng_proz
-120	Z	-2,134[a]	-,277[a]	-2,091[b]
	Asymptotische Signifikanz (2-seitig)	,033	,782	,037
-80	Z	-1,202[a]	-,391[b]	-1,305[b]
	Asymptotische Signifikanz (2-seitig)	,229	,696	,192
-40	Z	-,881[b]	-,598[b]	-,015[a]
	Asymptotische Signifikanz (2-seitig)	,378	,550	,988
40	Z	-,504[a]	-,856[b]	-,627[b]
	Asymptotische Signifikanz (2-seitig)	,615	,392	,531
80	Z	-,639[b]	-1,090[a]	-1,621[a]
	Asymptotische Signifikanz (2-seitig)	,523	,276	,105
120	Z	-2,220[a]	-2,270[a]	-,044[a]
	Asymptotische Signifikanz (2-seitig)	,026	,023	,965

a. Basiert auf positiven Rängen.

b. Basiert auf negativen Rängen.

c. Wilcoxon-Test

Tabelle 11: Ergebnisse des Post-Hoc Wilcoxon-Tests

Um einen Überblick über die Stärke des Zusammenhanges zwischen den Variablen zu gewinnen, wird im Folgenden kurz die Effektstärke für die Varianzanalyse betrachtet. Nach Bortz und Döring (2006, S.615) lässt sich die Effektgröße E mit Hilfe von Eta-Quadrat (η^2) nach folgender Formel berechnen:

$$E = \sqrt{\frac{\eta^2}{1-\eta^2}}$$

Für eine einfaktorielle Varianzanalyse wurden die Effektgrößen wie folgt festgelegt:

E = 0,10: kleine Effektgröße
E = 0,25: mittlere Effektgröße
E = 0,40: große Effektgröße

Weiterhin gilt für die Varianzaufklärung η^2:

η^2 = 0,01: kleiner Effekt
η^2 = 0,10: mittlerer Effekt
η^2 = 0,25: großer Effekt

Aus der nachfolgenden Tabelle kann abgeleitet werden, dass im Bezug auf den prozentualen Anteil richtiger Antworten und die Sprache bis auf eine zeitliche Verzögerung von -40ms (η^2 = 0,006) alle Zeitversätze einen kleinen bis mittleren Effekt hätten (0,01< η^2<0,10). Mit E_{-120} = 0,212 und E_{120} = 0,208 weisen die maximalen Zeitversätze +120ms und -120ms ebenfalls die größte Effektgröße in diesen Hörversuch auf, was in diesem Fall einer kleinen bis mittleren Effektgröße entspricht. Weitere Werte zur Effektstärke können aus der folgenden Tabelle (Tabelle 12) entnommen werden.

Zusammenhangsmaße

versatz		R	R-Quadrat	Eta	Eta-Quadrat
-120	prozent * sprache_ord	-,030	,001	,212	,045
-80	prozent * sprache_ord	,019	,000	,142	,020
-40	prozent * sprache_ord	,044	,002	,077	,006
40	prozent * sprache_ord	,086	,007	,127	,016
80	prozent * sprache_ord	-,096	,009	,154	,024
120	prozent * sprache_ord	-,180	,032	,208	,043

Tabelle 12: Tabelle der Zusammenhangsmaße von Sprache und prozentualen Synchronurteilen für die ANOVA

Nach Betrachtung der prozentualen Fehlurteile wurde mit Hilfe von Matlab die Position des Punktes subjektiver Synchronität (PSS) und des gerade wahrnehmbaren Unterschieds (JND) ermittelt. Hierzu stand das Matlab curve fitting tool zur Verfügung, welches als Gleichung für die Kurvenanpassung die folgende gaußsche Funktion annimmt:

$$f(x) = a1 * e^{(-((x-b1)/c1)^2)}$$

wobei der Punkt subjektiver Synchronität bei 50 % korrekter Antworten eintrifft und der gerade wahrnehmbare Unterschied über die Differenz zwischen dem oberen und unteren Schwellenwert nach folgender Formel berechnet werden kann:

$$JND = (p_{75} - p_{25})/2$$

Mithilfe des Matlab-Skriptes kann für die drei Sprachen die Unterschiedsschwelle berechnet werden. Dies ist möglich, da für den Hörversuch ein Konstanzverfahren benutzt wurde, welches gleichmäßig um einen Schwellenwert verteilte Reize in zufälliger Reihenfolge anwendet. Der PSS ist dabei genau der Wert, bei der in 50 % der Fälle vom Versuchsteilnehmer ein korrektes Synchronurteil getroffen wurde.

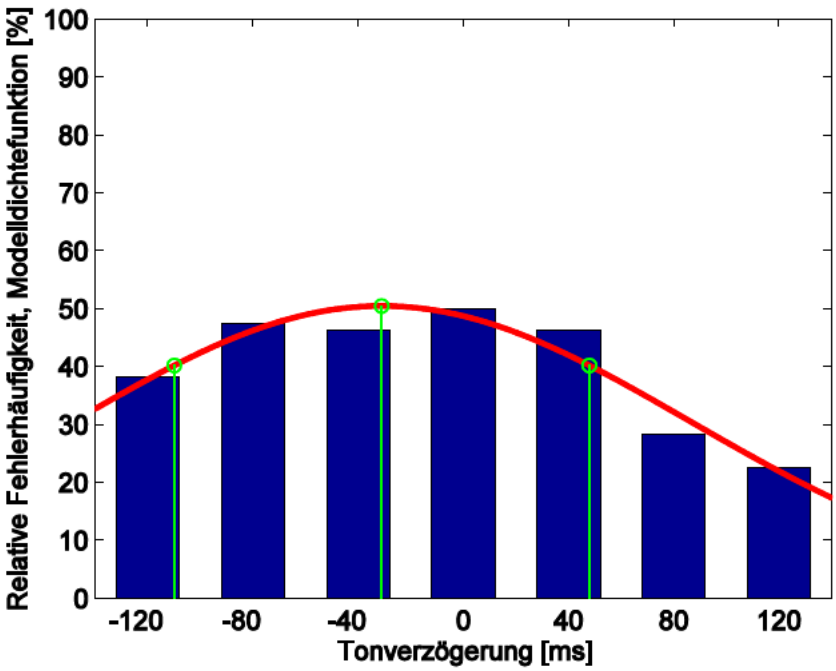

Abbildung 24: Kurvenanpassung am Beispiel der Bewertung des deutschen Materials

Für die Auswertung wurde das Skript für jeden Probanden und für jede Sprache einzeln angewendet, sodass für die anschließende Analyse in SPSS jeweils 90 Einträge für den PSS, den unteren und den oberen Schwellenwert zur Verfügung standen. Aus den ermittelten Daten wurden Mittelwerte berechnet und durch eine weitere einfaktorielle ANOVA ein Signifikanztest zur Bestimmung von Unterschieden zwischen den Werten in Abhängigkeit von der Sprache durchgeführt.

Nach den Synchronurteilen des deutschen, englischen und persischen Materials, ergeben sich die durchschnittlichen Werte für den PSS und JND wie folgt:

	Deutsch	Englisch	Persisch
PSS	-19,54ms	-31,02ms	-16,12ms
JND 1	-82,64ms	-109,91ms	-90,38ms
JND 2	43,56ms	47,87ms	58,14ms
JND	63,1ms	78,89ms	74,26ms

Tabelle 13: PSS und JND-Werte für Deutsch, Englisch und Persisch

Hinsichtlich des PSS und des JND ergeben sich nur sehr geringe Unterschiede, welche nach einer Varianzanalyse als insignifikant betrachtet werden können.

ANOVA-Tabelle

			Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
jnd1 * Sprache	Zwischen den Gruppen	(Kombiniert)	11281,181	2	5640,591	1,540	,220
	Innerhalb der Gruppen		318636,9	87	3662,493		
	Insgesamt		329918,1	89			
jnd2 * Sprache	Zwischen den Gruppen	(Kombiniert)	919,145	2	459,572	,157	,855
	Innerhalb der Gruppen		254577,1	87	2926,174		
	Insgesamt		255496,2	89			
pss * Sprache	Zwischen den Gruppen	(Kombiniert)	2507,716	2	1253,858	,600	,551
	Innerhalb der Gruppen		181775,9	87	2089,378		
	Insgesamt		184283,6	89			

Tabelle 14: SPSS-Tabelle der PSS- und JND-Varianzanalyse

Zusätzlich könnten bei signifikanten Ergebnissen nur kleine Effekte nachgewiesen werden. Die Werte für die Varianzaufklärung liegen mit η^2 = 0,004 unter dem Wert für einen kleinen Effekt und mit η^2 = 0,014 bzw. η^2 = 0,034 zwischen einem kleinen und einem mittleren Effekt. Ebenso ergeben sich die Eta-Werte zu einer kleinen Effektgröße.

Zusammenhangsmaße

	Eta	Eta-Quadrat
jnd1 * Sprache	,185	,034
jnd2 * Sprache	,060	,004
pss * Sprache	,117	,014

Tabelle 15: Zusammenhangsmaße für Sprache und PSS- bzw. JND-Werte

In einem abschließenden Test wurden die Beurteilungen von Männern und Frauen dieses Versuchs verglichen. Es nahmen insgesamt 30 Probanden teil, von denen 16 männlich waren. Durch die geringe Anzahl in den zwei Gruppen können Ergebnisse nur als Tendenz und nicht als statistisch repräsentativ interpretiert werden. Der durchgeführte Hörversuch war nicht für einen Vergleich der männlichen und weiblichen Bewertungsunterschiede ausgelegt. Die Auswertung der Ergebnisse soll daher an dieser Stelle auch nur eine kurze Erwähnung finden.

In der folgenden Abbildung (Abbildung 25) sind jeweils die fehlerhaften Synchronurteile in Prozent für die sechs unterschiedlichen Zeitversätze dargestellt. Die linke Grafik stellt die Urteile der Männer und die rechte die der Frauen dar. Dabei gilt weiterhin, dass die blaue Linie der Bewertung des deutschen, die grüne Linie der des englischen und die rote Linie der des persischen Materials entspricht.

Die Fehlurteile der Männer weisen im Deutschen bei +120ms den geringsten Wert auf im Vergleich zu Englisch und Persisch. Während bei einem zeitlichen Versatz von +40ms die fehlerhaften Synchronurteile das Maximum für die Sprachen Deutsch und Englisch mit knapp bzw. über 50 % erreichen, konnten im Persischen die meisten Probanden bei einem Zeitversatz von -40ms die physikalisch synchrone Version nicht von der asynchronen unterscheiden. Für alle drei Sprachvarianten und Sprachkompetenzen ergibt sich der gleiche prozentuale Anteil an Fehlurteilen für eine Tonverzögerung von -120ms.

Die Entscheidungen der weiblichen Teilnehmer weisen bei +120ms fast identische Werte und ähnliche Verläufe bis -40ms in den drei Sprachen auf. Während im Deutschen und Persischen die meisten Fehlentscheidungen bei einer Tonverzögerung von -40ms abgegeben wurden, erreicht Englisch das Maximum bei einem Versatz von -80ms. In allen drei Sprachen nimmt die Wahrnehmung von Asynchronität mit steigender Tonverzögerung wieder zu und die fehlerhaften Synchronurteile wieder ab.

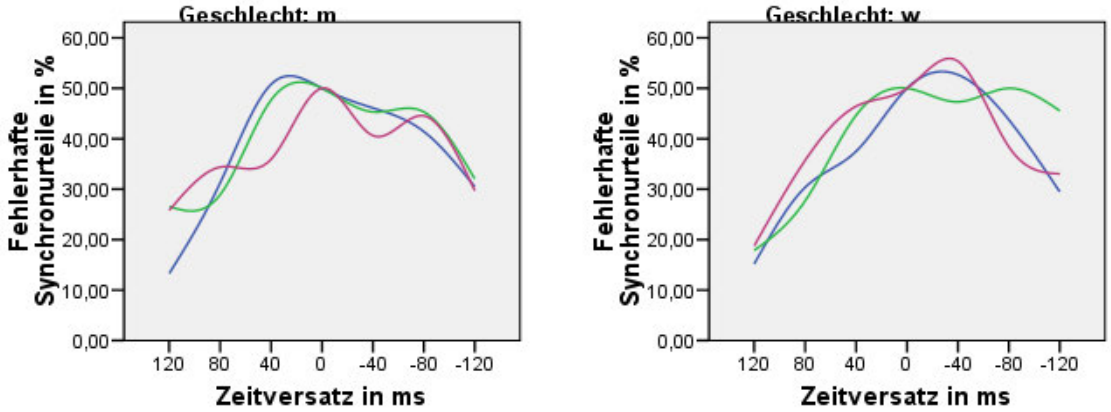

Abbildung 25: Gegenüberstellung der Fehlurteile von Männern und Frauen

In einem T-Test für unabhängige Stichproben wurden die Bewertungen der männlichen und weiblichen Teilnehmer auf signifikante Unterschiede überprüft. Für eine Tonverzögerung von -120ms konnte ein signifikant besseres Erkennen der physikalisch synchronen Version in der englischen Sprache durch die männlichen als durch die weiblichen Teilnehmer nachgewiesen werden ($F = 0{,}078$, $p = 0{,}024$). Auch im Persisch, bei einer Tonverzögerung von -40ms, erkennen die Männer mit signifikant weniger Fehlurteilen die synchrone Fassung sicherer ($F = 1{,}144$, $p = 0{,}009$). Im Gegensatz dazu weist der Tonvorzug von +40ms im Deutschen bei den Frauen einen geringeren prozentualen Anteil an Fehlurteilen auf, welcher mit $p<0{,}05$ signifikanten ist ($F = 0{,}000$, $p = 0{,}023$). Die komplette Tabelle des T-Tests kann auf der beigefügten CD gefunden werden.

Mit Hilfe des Matlab-Skripts wurden nach einer Kurvenanpassung abschließend die PSS- und die JND-Werte ermittelt und auf signifikate Unterschiede überprüft:

	Sprache	Männlich	Weiblich
jnd1	d	-77,9979	-87,054
	e	-93,4652	-130,0038
	p	-100,5415	-81,5872
jnd2	d	47,8179	38,094
	e	59,3452	33,2038
	p	66,4815	50,0372
pss	d	-15,09	-24,48
	e	-17,06	-48,4
	p	-17,03	-15,78

Tabelle 16: Tabelle der PSS- und JND-Werte für Männer und Frauen

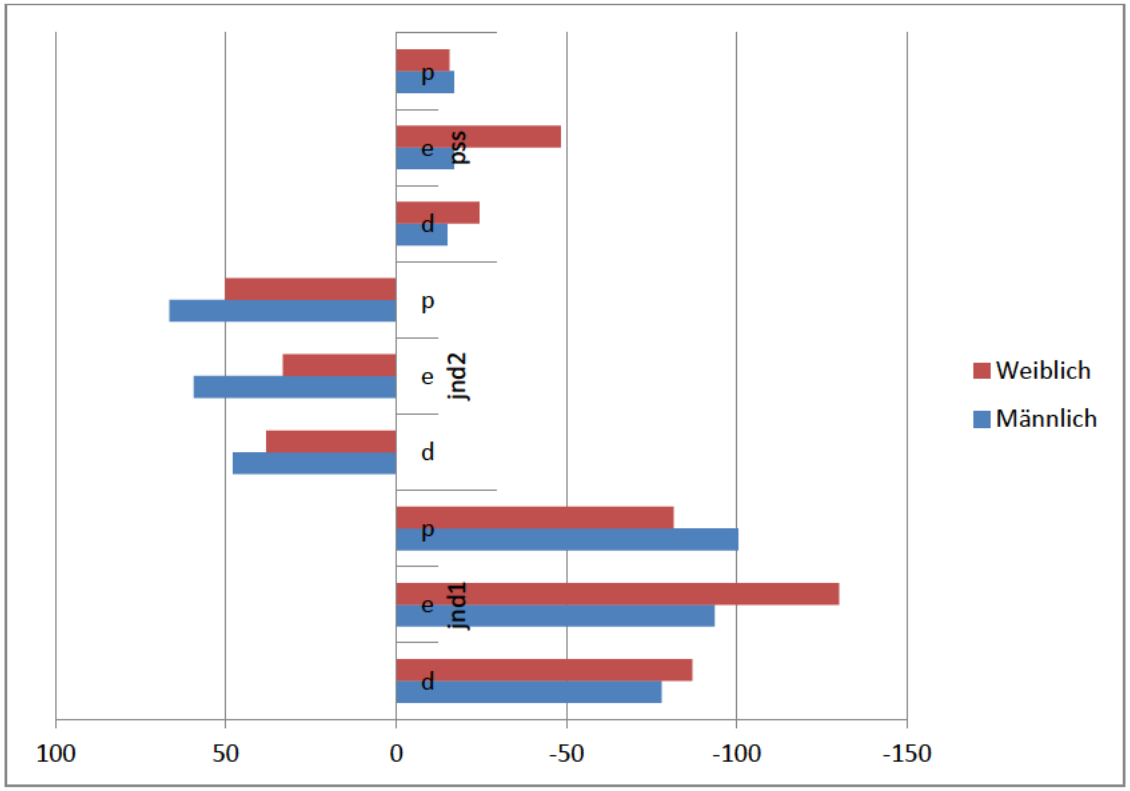

Abbildung 26: Excel-Grafik der PSS- und JND-Werte für Männern und Frauen

Ein mit SPSS durchgeführter T-Test konnte im Gegensatz zum Vergleich der prozentualen Fehlurteile nur schwach signifikante Unterschiede für den PSS im Englischen bestätigen ($F = 6{,}076$, $p = 0{,}051$), welcher sich für die männlichen Teilnehmern zu einer Tonverzögerung von -17,06ms und für die weiblichen Teilnehmern zu -48,4ms ergab. Für die weiteren Ergebnisse konnte kein signifikanter Einfluss der Sprache nachgewiesen werden.

Da dieser Hörversuch nicht für einen Vergleich der Bewertungen beider Geschlechter ausgelegt war, weisen die Stichproben einen zu geringen Umfang auf, um statistisch repräsentative Aussagen treffen zu können. Die gewonnen Ergebnisse zeigen nur eine Tendenz für eine unterschiedliche Breite des Synchronitätsfenster und für eine Verlagerung des PSS für Männer und Frauen im Bezug auf die dargebotenen Sprachen auf. Weitere Untersuchungen in diesem Bereich wären von Interesse, um aussagekräftige Ergebnisse liefern zu können.

Als letzter Punkt sei an dieser Stelle erwähnt, dass die Daten die Ergebnisse von früheren Studien (Heide, 2009; Rudloff, 1997) im Bereich Asynchronitätswahrnehmung bestätigen konnten. Zwischen den einzelnen Zeitversätzen konnten unabhängig von der Sprache signifikante Unterschiede nachgewiesen werden. Die nachfolgende Grafik zeigt einen Auszug der entsprechenden Post-Hoc-Tabelle aus SPSS. Die vollständigen Daten können der SPSS-Datei auf der beiliegenden CD entnommen werden.

			Tukey-HSD				
						95%-Konfidenzintervall	
(I) versatz	(J) versatz	Sprache_Video	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	Untergrenze	Obergrenze
120	-120	d	15,83333*	4,14341	,003	3,4937	28,1730
		e	15,83333*	4,27367	,005	3,1057	28,5609
		p	8,75000	4,24052	,379	-3,8789	21,3789
	-80	d	28,33333*	4,14341	,000	15,9937	40,6730
		e	25,00000*	4,27367	,000	12,2724	37,7276
		p	19,16667*	4,24052	,000	6,5378	31,7956
	-40	d	35,00000*	4,14341	,000	22,6603	47,3397
		e	23,75000*	4,27367	,000	11,0224	36,4776
		p	25,00000*	4,24052	,000	12,3711	37,6289
	0	d	35,83333*	4,14341	,000	23,4937	48,1730
		e	27,50000*	4,27367	,000	14,7724	40,2276
		p	27,50000*	4,24052	,000	14,8711	40,1289
	40	d	30,41667*	4,14341	,000	18,0770	42,7563
		e	23,75000*	4,27367	,000	11,0224	36,4776
		p	18,33333*	4,24052	,000	5,7044	30,9622
	80	d	16,66667*	4,14341	,002	4,3270	29,0063
		e	5,83333	4,27367	,820	-6,8943	18,5609
		p	12,50000	4,24052	,054	-,1289	25,1289

*. Die Differenz der Mittelwerte ist auf dem Niveau .05 signifikant.

Tabelle 17: Auszug aus der Post-Hoc-ANOVA-Tabelle der Vergleiche zeitlicher Versätze

Wie aus der Tabelle (Tabelle 17) abzulesen ist, wurde besonders die Tonvoranstellung von +120ms im Vergleich zu den anderen Zeitversätzen signifikant unterschiedlich bewertet. Aus der vollständigen Tabelle, welche auf der CD zu finden ist, können für Deutsch 13 signifikant unterschiedliche Bewertungen nachgewiesen werden. Im Englischen sind es 9 und im Persischen 7. Dabei ist anzumerken, dass fünf der Vergleiche zwischen den Zeitversätzen in allen drei Sprachen an den gleichen Stellen signifikant sind. Dies trifft auf die Bewertungen von +120ms/ -80ms, +120ms/ -40ms, +120ms/ 0ms, +120ms/ +40ms und 0ms/ 80ms zu. In allen Fällen liegen die Signifikanzwerte zwischen $0{,}000 \leq p \leq 0{,}018$ und sind damit signifikant bzw. höchst signifikant. Es wurde jeweils unabhängig von der Sprache die Tonvoranstellung von +120ms signifikant sicherer erkannt, als der Großteil der anderen dargebotenen Zeitversätze. Während für die jeweiligen Latenzen kein Unterschied zwischen den

Sprachen festgestellt werden konnte, existiert dieser für jede Sprache zwischen den unterschiedlichen Latenzen. Auf eine Diskussion dieses Ergebnisses soll an dieser Stelle verzichtet und auf die Arbeiten von Frau Berta Heide (2009) und Herrn Ingo Rudloff (1997) verwiesen werden.

Im letzten Unterpunkt dieses Kapitels werden die gewonnenen Ergebnisse diskutiert, analysiert und Interpretationsansätze präsentiert.

4.2 Diskussion

Im Folgenden werden die Ergebnisse des Hörversuchs interpretiert und Erklärungsansätze dargelegt. Dabei werden jeweils die Hypothesen der Arbeit erneut präsentiert und aufgrund der gewonnen Daten bestätigt oder widerlegt.

4.2.1 Diskussion der Hypothesen

Die erste Hypothese dieser Masterarbeit betrachtet die Lage des Synchronitätsfensters in Abhängigkeit von der Sprachkompetenz. Verglichen wurden dabei die Synchronurteile der Probanden für das audiovisuelle Material, welches jeweils in sechs Zeitversätzen in ihrer Muttersprache Deutsch, ihrer ersten Fremdsprache Englisch und der ihnen unvertrauten Fremdsprache Persisch dargeboten wurde.
Für alle drei Sprachen und Sprachkompetenzen konnte ein erhöhter Prozentwert an fehlerhaften Synchronurteilen für die Tonverzögerungen -40ms und -80ms nachgewiesen werden. Diese Ergebnisse decken sich mit der Untersuchung von Frau Heide (Heide, 2009, S. 52).

Um die Bewertungen in den jeweiligen Sprachen signifikant zu unterscheiden, wurde eine einfaktorielle ANOVA im Statistik-Programm SPSS durchgeführt. Diese Varianzanalyse erlaubt den Vergleich von Mittelwerten von mehr als zwei Gruppen. Im Fall dieser Masterarbeit ergaben sich die drei Gruppen zu Deutsch, Englisch und Persisch. So wurden unter anderem die Mittelwerte zwischen den drei Gruppen für den Zeitversatz +40ms untereinander auf signifikante Unterschiede hin vergleichen. Die Varianzanalyse lieferte in einem ersten Vergleich keine signifikanten Unterschiede. Nachfolgende Post-Hoc-Tests, welche für die jeweiligen Zeitversätze Paarvergleiche durchführten, konnten nur Tendenzen für die Randbereiche von +120ms und -120ms nachweisen. Die Ergebnisse zeigten bei +120ms einen Anstieg korrekter Synchronurteile für Deutsch im Vergleich zu Englisch und Persisch. Im Gegensatz dazu wurde die Tonverzögerung von -120ms im Englisch schlechter von der physikalisch synchronen Fassung unterschieden als die deutsche oder persische Fassung bei gleichem Versatz. Signifikante Unterschiede konnten hier nicht bestätigt werden, doch Tendenzen mit $p = 0{,}064$ bzw. $p= 0{,}09$ könnten eine Grundlage für weitere Untersuchungen bilden.

Ein Test auf Normalverteilung der Daten (Kolmogorov-Smirnov-Test) ergab in zwei Fällen signifikante Abweichungen von der geforderten Normalverteilung. Da die

Varianzanalyse auch bei kleinen Stichprobenumfängen robust gegen Nicht-Normalverteilung ist, kann weiterhin davon ausgegangen werden, dass die Ergebnisse korrekt sind. Zur Kontrolle wurde der Friedman-Test für abhängige Stichproben durchgeführt. Dieser Test ist parameterfrei und setzt keine Normalverteilung der erhobenen Daten voraus. Die Ergebnisse des Friedman-Tests liefern einen signifikanten Unterschied, welcher durch einen folgenden Wilcoxon-Test für gepaarte Stichproben bestätigt werden konnte. Es ergaben sich signifikante Werte für die Randbereiche des Hörversuchs bei einer Tonverzögerung von -120ms und einem Tonvorzug von +120ms.

Mit den gewonnenen Daten kann die erste H1-Hypothese bestätigt werden. Die Sprachkenntnisse haben einen Einfluss auf die Breite des Synchronitätsfensters. Das signifikant bessere Erkennen der physikalisch synchronen oder asynchronen Fassung bei +120ms im Deutschen lässt darauf schließen, dass diese Tonvoranstellung bereits außerhalb des Synchronitätsfensters liegt. Die Integration des auditiven und visuellen Signals erfolgte bei diesem Versatz nur in 34 von 240 Darbietungen und führte zu einem fehlerhaften Erkennen der synchronen Version. Im Gegensatz dazu konnte im Englischen und Persisch die physikalisch synchrone Version in 22,5 % der Fälle nicht von der asynchronen unterschieden werden. Dies deutet auf eine erfolgreiche Integration beider Signale hin. Während in der Muttersprache, in diesem Fall Deutschen, ein verfrühtes auditives Signal ab ≥+80ms dazu führt, dass es unabhängig vom visuellen Signal verarbeitet wird, benötigten die Teilnehmer für eine Fremdsprache eine größere Abweichung, damit eine Asynchronität erkannt wird.

Es konnte weiterhin nachgewiesen werden, dass die im Versuch gesetzte Obergrenze von -120ms im Englisch signifikant schlechter beurteilt wurde als im Deutschen oder Englischen. Die hohe Anzahl an Fehlurteilen weist erneut auf eine Integration des auditiven und visuellen Signals hin, wodurch die Unterscheidung des physikalisch synchronen von dem asynchronen Video nicht immer möglich war. Dies erlaubt die Schlussfolgerung, dass eine Tonverzögerung von -120ms noch im Synchronitätsfensters für die erste Fremdsprache liegt.

Die an den Hörversuch angepassten Ober- und Untergrenzen des Synchronitätsfensters können der folgenden Tabelle (Tabelle 18) entnommen werden. Die Werte für Deutsch, Englisch und Persisch können dabei jeweils für Werte der

Muttersprache, ersten Fremdsprache und unvertrauten Fremdsprache interpretiert werden.

	Deutsch	Englisch	Persisch
Untergrenze	+40ms	+40ms	+40ms
Obergrenze	-80ms	-120ms	-80ms

Tabelle 18: An die verwendeten Zeitversätze angepasste Ober- und Untergrenzen der Synchronitätsfenster

Die genauen Werte für den eben wahrnehmbaren Unterschied für die Tonvoranstellung und -verzögerung können der vorhergehenden Tabelle (Tabelle 13) entnommen werden. Die daraus gewonnenen Daten erlauben die Interpretation, dass das deutsche Synchronitätsfenster im Vergleich zu dem englischen und persisch am schmalsten ist. Im Gegensatz dazu legen die Ergebnisse die Vermutung nahe, dass sich das Synchronitätsfenster in einer vertrauten Fremdsprache ausdehnt und zu einer stärkeren Tonverzögerung verlagert, sodass die Integration von auditiven und visuellen Informationen auch bei größeren Latenzen erfolgt. Die Werte für Persisch zeigen ebenfalls eine Ausdehnung des Synchronitätsfensters, weisen allerdings eine leichte Verlagerung zu einer Tonvoranstellung im Vergleich zum deutschen Synchronitätsfenster auf. Eine Gegenüberstellung der JND-Werte der drei Sprachen ist im Folgenden grafisch dargestellt:

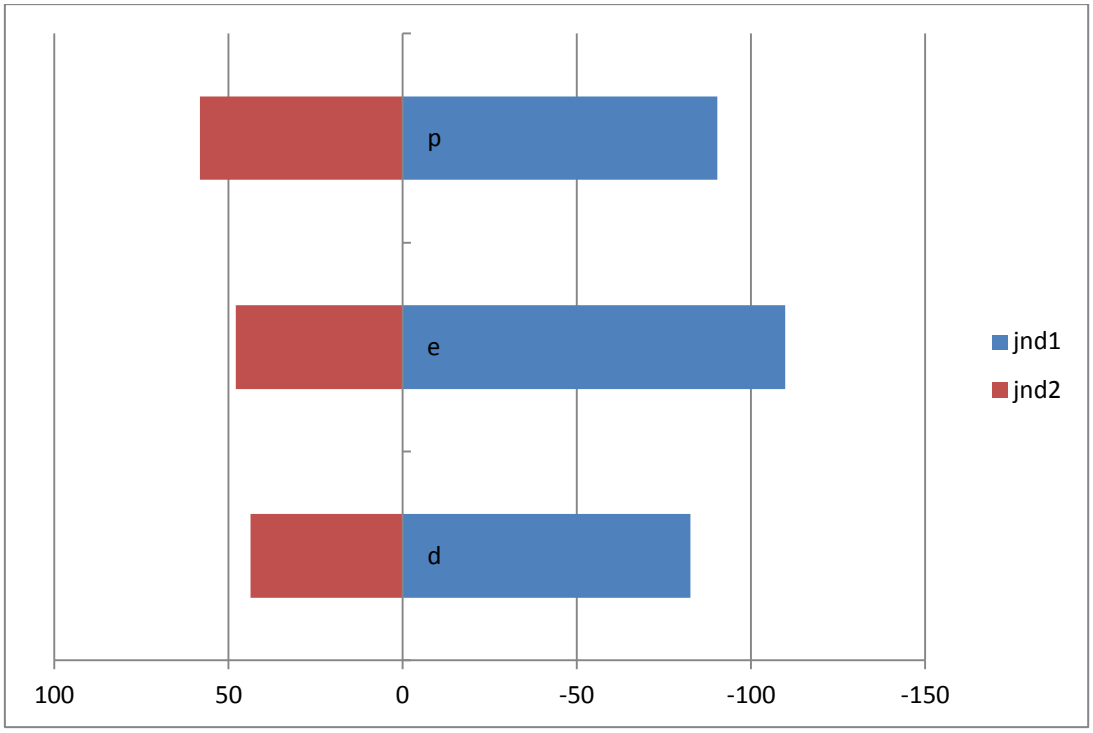

Abbildung 27: Excel-Grafik der JND-Werte für Deutsch (d), Englisch (e) und Persisch (p)

Die Breite des Synchronitätsfensters für die jeweiligen Sprachen ergibt sich dabei zu folgenden Werten:

	Deutsch	Englisch	Persisch
Breite	126,2ms	157,8ms	148,5ms

Tabelle 19: Breite des Synchronitätsfensters für Deutsch, Englisch und Persisch auf Basis der JND-Werte

Zusätzlich wurde für die Entscheidung der zweiten Hypothese dieser Masterarbeit der Punkt subjektiver Synchronität verglichen. Dieser Punkt beschreibt den zeitlichen Versatz, der von den Teilnehmern in der Hälfte der Fälle trotz asynchroner Darbietung als physikalisch synchron wahrgenommen wurde. Unter Angabe der gruppengesamten durchschnittlichen prozentualen Fehlurteile für jeden Versatz und jede Sprache erfolgte eine Kurvenanpassung mithilfe der Matlab curve fitting tool. Die ermittelten PSS-Werte können der vorhergehenden Tabelle (Tabelle 13) entnommen werden.

Da sowohl für die Lage des PSS als auch die JND-Werte keine signifikanten Unterschiede in Abhängigkeit von der Sprachkompetenz nachgewiesen werden könnten, kann die zweite H1-Hypothese verworfen werden. Die Lage des Punktes subjektiver Synchronität hängt nicht von der Sprachkompetenz ab. Aus den ermittelten Ergebnissen lassen sich jedoch zwei unterschiedliche Interpretationsansätze ableiten.

Navarra et al. (2010) konnten ihn einem ähnlichen Versuch einen signifikanten Unterschied der PSS-Lage nachweisen, wenn geringe Sprachkenntnisse in einer Fremdsprache vorlagen. Bei einer hohen Sprachkompetenz wurden ähnliche Werte wie für die Muttersprache ermittelt. Navarra et al. schlussfolgerten, dass die Erfahrung mit einer Fremdsprache einen Einfluss auf die Lage des Punktes subjektiver Gleichheit hat und mit zunehmender Erfahrung sich der PSS-Wert dem der Muttersprache angleicht. Für den vorliegenden Hörversuch wurde Englisch als erste Fremdsprache vorausgesetzt, die Sprachkompetenz jedoch nicht durch einen Einstufungstest oder eine Selbsteinschätzung genauer erfasst. Betrachtet man die Ergebnisse unter diesem Gesichtspunkt, kann das Fehlen von signifikanten Unterschieden auf eine hohe Sprachfähigkeit der Probanden zurückgeführt werden. Die Teilnehmer des Hörversuchs waren alle Studenten der Fachrichtungen „Audiokommunikation und -technologie", „Sprache und Kommunikation“ oder „Informatik“ in einem Alter zwischen 20 und 32 Jahren. Damit sind eine hohe Affinität zu audiovisuellen Medien und ein fast täglicher passiver oder aktiver Gebrauch der englischen Sprache anzunehmen.

Die zweite H1-Hypothese kann in diesem Fall nur für die Annahme einer hohen Sprachkompetenz verworfen werden.

Die gewonnenen Daten bestätigen die Ergebnisse früherer Untersuchungen (Navarra et al., 2010; Stekelenburg und Vroomen, 2011), dass die visuelle Information des Lippenlesens die kognitiven Reaktionen auf ein Sprachsignal beschleunigen kann. Nach Navarra et al. (2010) trifft dies besonders auf die Muttersprache zu.

> *"The processing of the visual information (probably most effective in one's own native language) may speed up the processing of the ensuing auditory information and, as a consequence, reduce the apparent temporal separation between the visual and auditory signals." (Navarra et al., 2010, S.88)*

Um frühzeitig Informationen aus dem Audiosignal extrahieren zu können, bindet das menschliche System visuelle Informationen ein, welche ein Voraussehen möglicher phonetischer und lexikalischer Segmente ermöglicht. Wie wirksam dieser Prozess genutzt werden kann, hängt von der bestehenden Erfahrung mit den Lauten und Lippenbewegungen der jeweiligen Sprache ab (Mueller et al., 2006). Dies zeigt sich besonders im Vergleich der Muttersprache zu den beiden anderen im Versuch genutzten Sprachen. Das Synchronitätsfenster ist im Deutschen, der Muttersprache der Teilnehmer, am schmalsten, was auf eine Integration von audiovisuellen Informationen in einem geringeren zeitlichen Rahmen als in anderen Sprachen hinweist.

Für Englisch und Persisch konnte sowohl ein breites Synchronitätsfenster als auch eine höhere Anzahl prozentualer Fehlurteile für den Tonvorzug von +120ms und Fall des Englischen ebenfalls für die Tonverzögerung von -120ms nachgewiesen werden. Diese Ergebnisse deuten auf ein größeres zeitliches Intervall hin, indem audiovisuelle Information integriert wird und ein Erkennen von Asynchronität nicht vollständig möglich ist.

Die Analyse der Lage des PSS und der JND-Werte wiesen keine signifikanten Unterschiede auf und auch die einfaktorielle ANOVA der prozentualen Fehlurteile blieb ohne signifikantes Ergebnis. Dies erlaubt eine zweite mögliche Interpretation. Wie oben bereits zusammengefasst, kann die zweite H1-Hypothese verworfen werden. Daraus folgt, dass die Sprache im verwendeten Videomaterial und die damit verbundene

Sprachkompetenz der Probanden keinen Einfluss auf die Lage des Punktes subjektiver Synchronität oder die Schwelle des eben wahrnehmbaren Unterschieds hat. Dieser Aussage schließt sich die Schlussfolgerung an, dass das Erkennen von Synchronität und Asynchronität unabhängig von der wahrgenommenen Sprache ist. Dies zeigt sich besonders durch die Übereinstimmung der Werte der deutschen und persischen Beurteilungen. Während Deutsch die Muttersprache der Versuchsteilnehmer war, wurde vorausgesetzt, dass das Persische für sie eine unvertraute Fremdsprache war. Die Erwartung, dass sich das Nichtwissen über diese Fremdsprache auf die Synchronurteile der Probanden auswirkt, konnte nicht bestätigt werden. Einzig eine Verlagerung hin zu einer Tonvoranstellung im Vergleich zu den deutschen Werten konnte nachgewiesen werden.

Diese Theorie baut auf den Untersuchungen von Macaluso et al. (2003) und Stevenson et al. (2007) auf. In PET- und fMRT-Studien konnten Gehirnregionen bestimmt werden, die besonders stark auf synchrone audiovisuelle Signale reagieren. Dazu gehören nach Macaluso et al. (2003) ventrale und dorsolaterale Okzipital-Areale und der linke STS (Superior temporal sulcus). Die Ergebnisse von Stevenson et al. (2007) bestätigen diese Funde und beschreiben einen erhöhten BOLD-Level im multisensorischen superior temporal cortex (mSTC).

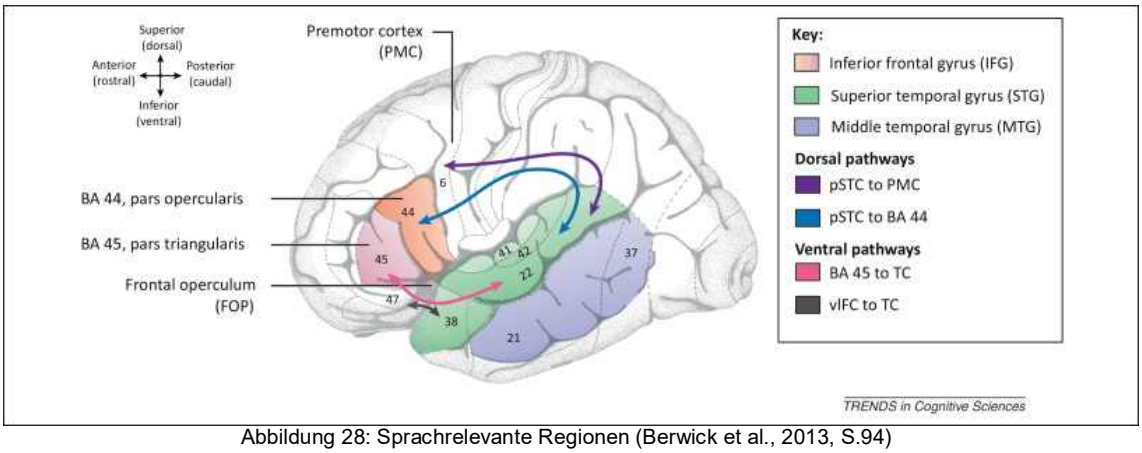

Abbildung 28: Sprachrelevante Regionen (Berwick et al., 2013, S.94)

Es zeigt sich, dass in diesen Bereichen Neuronen besonders stark auf synchrone multisensorische Signale reagieren. Die Neuronen reagieren zusätzlich auf die "temporal rule", welche für auditive und visuelle Reize eine zeitliche Latenz erwartet, um eine optimale Aktivierung der Neuronen zu ermöglichen. Weiter Funde in den Studien lassen die Vermutung zu, dass der mSTC als Filter zum Erkennen von Asynchronität agiert. Synchrone Stimulus-Paare werden für die Wahrnehmung und

Interpretation verbunden und andere Signale werden getrennt verarbeitet (Stevenson et al., 2007, S.3316).

Unter diesem Gesichtspunkt können die Ergebnisse als Bestätigung der Filter-Theorie interpretiert werden. Die fehlenden signifikanten Unterschiede für die dargebotenen Sprachen und Zeitversätze deuten auf eine erste Einstufung der Synchronität und anschließenden Zugriff aufs mentale Lexikon zur Interpretation des Signalinhalts hin. Aus dieser Theorie ist die Annahme abzuleiten, dass das Vorhandensein eines mentalen Lexikons und die Zugriffsgeschwindigkeit auf dieses Lexikon keinen Einfluss auf die Integration audiovisueller Information haben.

Dies würde ebenfalls mit der Theorie von Phillips et al (2000) übereinstimmen, wonach der auditive Cortex bereits in Ansätzen wichtige Funktionen für die weitere Verarbeitung tätigt. So wird angenommen, dass ihm phonologische Repräsentationen zur Verfügung stehen und sowohl für das Erkennen des Sprachsignals verantwortlich ist als auch eine Erwartungshaltung für nachfolgenden Stimuli einnimmt (Näätänen et al. 2001). In der nachfolgenden Grafik (Abbildung 29) würde demnach das Erkennen von audiovisueller Sprache, die Phonem-Antizipation und das Einschätzen der Synchronität der Phase 0 entsprechen. Hier ist ein Zugriff auf das mentale Lexikon, syntaktische und semantische Strukturen und das Einbeziehen von Erinnerungen noch nicht erfolgt. Dieses Modell der Sprachverarbeitung würde bestätigen, dass die Sprache keinen Einfluss auf das Erkennen von Synchronität hat.

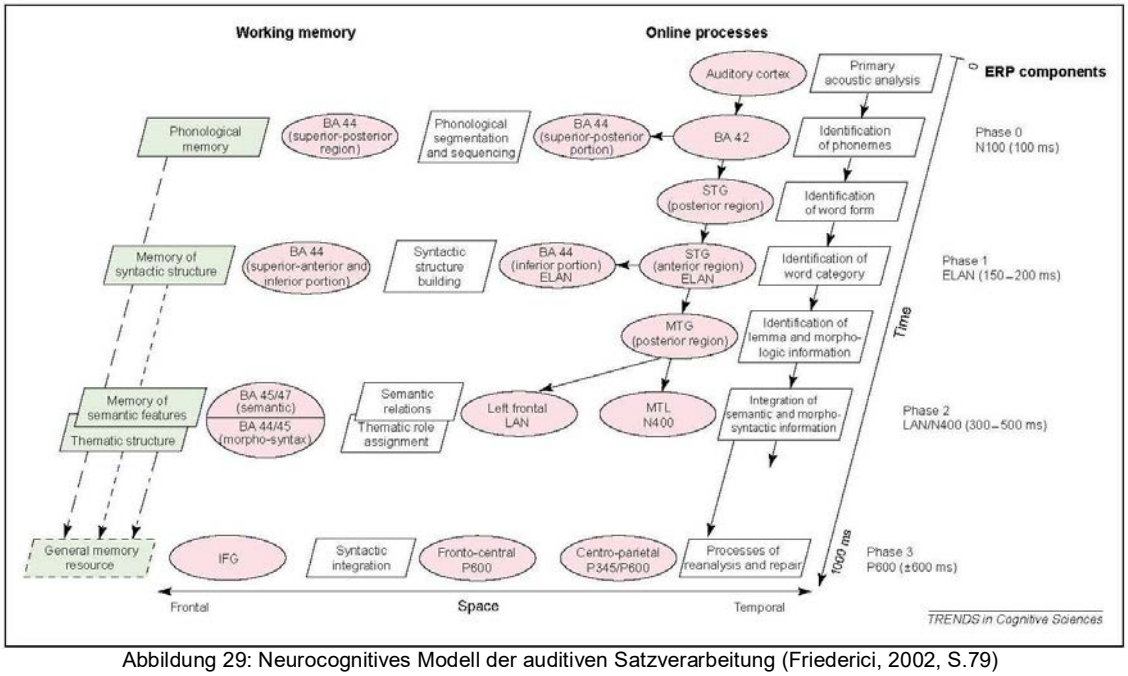

Abbildung 29: Neurocognitives Modell der auditiven Satzverarbeitung (Friederici, 2002, S.79)

Die dritte Hypothese stellt eine unvertraute Fremdsprache in den Vordergrund der Betrachtung. Im Fall dieser Masterarbeit war dies die persische Sprache. Im Folgenden wird jeweils noch einmal die erste und zweite Hypothese im Bezug auf das Persische untersucht. Die dritte H1-Hypothese basiert auf der Theorie, dass ohne ein mentales Lexikon für eine Fremdsprache das auditive Signal wie Musik verarbeitet wird. Diese Hypothese kann verworfen werden. Die Analyse der Daten durch eine einfaktorielle ANOVA und eine Kurvenanpassung mithilfe des Matlab-Skriptes ergab keine signifikante Unterschiede für die prozentualen Synchronurteile und auch nicht für die Lage des PSS und die Schwellenwerte. Es konnte nur eine leichte Verbreitung des Synchronitätsfensters im Vergleich zum deutschen Synchronitätsfensters und ein schwach signifikanter Unterschied der persischen und deutschen Bewertungen des Zeitversatzes +120ms bestätigt werden.

Heide (2009) wies in ihrer Magisterarbeit zur perzeptiven Synchronität von Bild und Ton einen signifikanten Unterschied zwischen Sprache und Musik nach. Die Ergebnisse dieser Masterarbeit bestätigen somit, dass eine unvertraute Sprache nicht als Musik interpretiert wird. Geringe oder keine vorherrschende Sprachkompetenz hat keinen Einfluss darauf, dass eine Fremdsprache beim Erkennen von Synchronität und Asynchronität als eine Sprache wahrgenommen wird. Dies erlaubt erneut die bisher ungeklärte Frage, welche Eigenschaften ein auditives Signal besitzen muss, um als Sprache und nicht als Musik oder Geräusche erkannt zu werden.

Eine Antwort auf diese Frage kann dieser Hörversuch nicht liefern. Jedoch konnte gezeigt werden, dass sich Muttersprache und unvertraute Fremdsprache im Bezug auf das Erkennen von Asynchronität nicht unterscheiden und sich auch die Lage des Punktes subjektiver Synchronität nicht mit der vorhandenen oder nicht vorhandenen Sprachkompetenz verändert.

Das Verwerfen der dritten H1-Hypothese basiert auf einem Vergleich der hier gewonnenen Daten und den Ergebnissen von Frau Heide (2009). Der bestätigte signifikante Unterschied zwischen Musik und Sprache würde einen eben solchen Unterschied zwischen Deutsch und Persisch erfordern, um eine andere Aussage treffen zu können. Die Hypothese, dass sich Persisch als unvertraute Fremdsprache wie Musik wahrgenommen wird, kann durch die statistisch vergleichbare Bewertung von Deutsch und Persisch jedoch nicht bestätigt werden.

4.2.2 Kritische Anmerkungen

An dieser Stelle sollen kritische Anmerkungen zur vorliegenden Masterarbeit genannt werden, die dazu beitragen sollen einen Überblick über den Umfang der Arbeit, die Sicherheit der Ergebnisse und bisher unberücksichtigte Einflussfaktoren zu gewinnen.

Die durch den Hörversuch gewonnenen Ergebnisse konnten einzelne Schlussfolgerungen anderer Arbeiten bestätigen und ergänzen. An dieser Stelle soll jedoch erwähnt werden, dass ein direkter Vergleich verschiedener Studien nur bedingt möglich ist. Hier wurde ein Forced-Choice-Verfahren verwendet und Teilnehmer mussten entscheiden, welches von zwei dargebotenen Videos physikalisch synchron dargeboten wurde. In einem ähnlichen Versuch von Navarra et al. (2010) wurde hingegen eine SJ-Aufgabe (synchrony order) gestellt. Probanden wurden in diesem Fall gefragt, ob ein Video synchron war oder nicht. Dies ist im Vergleich zu dem in dieser Masterarbeit verwendeten Versuchsdesign kein parameterfreier Test. Auch eine TOJ-Aufgabe (Temporal Order Judgment) oder Wortbenennungen unter Verwendung von bildgebenden Verfahren geben Einblicke in die Verarbeitung audiovisueller Sprache. Die Ergebnisse können untereinander jedoch nur begrenzt verglichen werden, da sie sich in wichtigen Faktoren wie der Aufgabenstellung, dem Versuchsdesign, dem Versuchsmaterial, dem Versuchsort, und dem Versuchsaufbau unterscheiden.

Auch der durchgeführte Hörversuch weist Fehlerquellen auf. Es nahmen ausschließlich Studenten der TU Berlin aus den Fachbereichen „Audiokommunikation und -technologie", „Sprache und Kommunikation“ sowie „Informatik“ teil. Zusätzlich lag das Alter der Teilnehmer zwischen 20 und 32 Jahren. Die gewonnenen Ergebnisse können daher nicht auf eine allgemeine Bevölkerungsgruppe angewandt werden. Hierzu wäre eine breitere Teilnehmerzahl aus verschiedenen Altersklassen nötig.

Eine weitere Fehlerquelle ist die geforderte Sprachkenntnis des Englischen. Dies war nötig, um einen Vergleich zwischen der Muttersprache, einer vertrauen und einer unvertrauten Fremdsprache durchzuführen. Die Anforderungen an die Teilnehmer gab nur Englisch als 1.Fremdsprache vor und lieferte somit keine detaillierte Auskunft über die tatsächliche Sprachkompetenz. Das junge Alter der Probanden lässt eine passive oder aktive Nutzung des Englischen vermuten. Dies wurde jedoch nicht durch einem Einstufungstest oder eine Selbsteinschätzung überprüft. Die gewonnenen Ergebnisse lassen wie oben beschrieben zwei Erklärungsansätze zu. Da im Gegensatz zum

Versuch von Navarra et al. (2010) hier ein parameterfreies Versuchsdesign gewählt wurde, ist schwer einzuschätzen, ob eine Widerlegung ihrer Ergebnisse vorliegt. Eine Untersuchung mit genauen Kennissen der Sprachniveaus ist daher angebracht.

Das verwendete Material wurde von Muttersprachlern gesprochen und im Versuch im Programm „Eiusdem Temporis“ in zufälliger Reihenfolge dargeboten, was etwa 45 Minuten pro Teilnehmer in Anspruch nahm. Hier traten Ermüdungserscheinungen der Teilnehmer auf. Eine Gewöhnung ans Material wurde durch die zufällige Sprachauswahl verhindert, doch die hohe Trialanzahl führte zu häufigen Wiederholungen, wodurch die Konzentration nachließ. Eine phasenweise Durchführung des Hörversuchs für jede Sprache hätte dies verhindern können, würde allerdings den Gewöhnungseffekt erlauben. Ein Einsatz von verschiedenen, vergleichbaren Sprachmaterialien wäre angebrachter gewesen. In diesem Versuch stand für jede Sprache ein ausgewählter Satz zur Verfügung. Die Verwendung von mehreren Sätzen und einer geringeren Trialanzahl könnte Ermüdungen der Teilnehmer verhindern.

Im Fall einer Anpassung müsste zusätzlich auf die visuelle Information durch Lippenlesen geachtet werden. Die automatische Integration der visuellen Information lässt vermuten, dass neben der Mimik auch die Lippenbewegungen eine Erwartungshaltung beim Teilnehmer erzeugen, welche einen Einfluss auf seine Entscheidung hat.

Zusätzlich sollte die Hypothesenbildung die asynchrone Sensibilität für die positive und negative Tonverschiebung berücksichtigt werden. Weiterhin sollte sich die sprachliche Abhängigkeit dieser Sensibilität in den Hypothesen widerspiegeln. Die Hypothesen sind für diese Masterarbeit allgemein formuliert worden und sollte in weiteren Versuchen an die Sprachen angepasst und genauer definiert werden.

Auch wäre eine Rückmeldung der Teilnehmer für eine Optimierung von zukünftigen Tests denkbar. Aufgrund des Testinhalts ist es nötig die Probanden über die Art des Tests aufzuklären. Die genaue Fragestellung zum Einfluss der Sprache war den Probanden in diesem Fall nicht bewusst, doch erweckte bereits das Wissen, dass das Erkennen von Asynchronität untersucht werden soll, eine gewisse Erwartungshaltung, welche die Ergebnisse beeinflusst haben könnte. Im Folgen sollen daher neben einer Zusammenfassung Anregungen für zukünftige Untersuchungen gegeben werden.

5. Fazit und Ausblick

Die vorliegende Masterarbeit konzentrierte sich auf die Frage des Erkennens von Synchronität von audiovisuellem Material in Abhängigkeit von der Sprachkompetenz. Hierfür wurde ein Hörversuch durchgeführt, indem Probanden in einem Forced-Choice-Verfahren zwei Videos in ihrer Muttersprache (Deutsch), ihrer ersten Fremdsprache (Englisch) und einer ihnen unvertrauten Fremdsprache (Persisch) präsentiert wurden. Die Aufgabe bestand darin die physikalisch synchrone Version zu erkennen. Die Synchronität am Hörerort wurde dafür vorher durch Messungen der zeitlichen Latenz von Licht und Schall sichergestellt.

Es konnten keine signifikanten Unterschiede für die sechs Zeitversätze -120ms, -80ms, -40ms, +40ms, +80ms und +120ms zwischen den Sprachen festgestellt werden. Das Erkennen von Asynchronität bzw. Synchronität gelang unabhängig von der Sprache mit äquivalenter Sicherheit. Auch die Lage des Punktes subjektiver Synchronität und eben wahrnehmbaren Unterschieds ergab keine signifikanten Unterschiede zwischen Deutsch, Englisch und Persisch. Es konnte eine Ausweitung des Synchronitätsfensters im Vergleich zum deutschen Synchronitätsfenster nachgewiesen werden, wodurch über ein größeres zeitliches Intervall audiovisuelle Information integriert werden kann.

Wie Frau Heide in ihrer Arbeit (2009) anmerkte, sollten weitere Untersuchungen zur Perzeptivität von komplexen Signalen wie Sprache durchgeführt werden. Die hier gewonnenen Ergebnisse basieren z.B. nur auf syllable-timed und stressed-timed Sprachen. In der Zukunft sollte auch ein Vergleich mit mora-timed Sprachen durchgeführt und mit den hier gewonnenen Ergebnissen verglichen werden. Zusätzlich fand dieser Hörversuch nur mit deutschen Muttersprachlern statt. Es wäre interessant zu sehen, ob ein ähnlicher Versuch mit Muttersprachlern der zu testenden Sprachen abweichende Ergebnisse liefert.

Daraus würden sich Verbesserungen für die Qualität von Film und Fernsehen ableiten lassen, da fremdsprachiges Material angepasst werden könnte, sollten zukünftige Untersuchungen Abweichungen feststellen. Frau Heide (2009) wies berechtigt darauf hin, dass Film und Fernsehen von drei Audio-Ebenen geprägt sind: Sprache, Musik und Geräusche. Die Ergebnisse dieser Arbeit erlauben die Schlussfolgerung, dass die

Sprache keinen Einfluss auf den Punkt subjektiver Synchronität hat. Auch dies ist ein wichtiger Punkt für die Produktion und die weitere Verarbeitungskette.

Einzig für eine Vergleich der männlichen und weiblichen Bewertungen konnte ein signifikanter Unterschied bei den Bewertungen festgestellt, welcher in weiteren Versuch mit einer größeren Stichprobenmenge noch einmal genauer klassifiziert werden sollte.

Der Versuchsaufbau könnte durch ein adaptives Verfahren und kleinere Intervalle erweitert werden und auch eine Untersuchung mit anderem Sprachmaterial wäre vorstellbar. Das hier verwendete Material basierte auf der in der Sprachwissenschaft genutzte Aesop-Fabel und wurde im Frontalbild von Muttersprachlern in ihrer jeweiligen Sprache aufgezeichnet. Um eine Qualitätssicherung von Film und Fernsehen zu gewährleisten, sollte hierzu weitere Untersuchungen mit unterschiedlichen Blickwickeln oder Farb- und Helligkeitsfaktoren durchgeführt werden, welche dem typischen Filmformat entsprechen. Auch ein Einfluss der Lautstärke der jeweiligen Sprache wäre als Faktor für das Erkennen von Asynchronität denkbar und wurde in diesem Versuch durch eine Normalisierung unberücksichtigt gelassen.

Eine statistisch repräsentative Aussage über den Einfluss der Sprachkompetenz auf die Wahrnehmung von audiovisuellem Material über alle Altersklassen hinweg und unter Einbezug der unterschiedlichen Sprachrhythmen kann in Zukunft helfen die Post-Produktion zu beschleunigen und eine bessere Qualität für die Zuschauer zu gewährleisten. Zusätzlich liefert es ein besseres Verständnis über die Verarbeitungsprozesse und die Ebenen der Informationsverarbeitung von audiovisuellem Sprachmaterial.

Die vorliegende Arbeit konnte einzelnen Fragestellungen zur Sprachverarbeitung von audiovisuellem Material nachgehen. Gleichzeitig wurde interessante neue Fragen und Ansätze aufgezeigt, welchen in zukünftige Untersuchungen nachgegangen werden sollte, um sowohl die Grundlagenforschung weiter zu vertiefen als auch die Qualität von Film und Fernsehen auch in Zukunft zu sichern und zu verbessern.

6. Verzeichnisse

6.1 Literaturverzeichnis

Berwick, R.C. und Friederici, A.D. und Chomsky, N. und Bolhuis, J. (2013). “Evolution, brain, and the nature of language”, In: *Trends in Cognitive Sciences* 17 (2013), S.89-98

Best, Catherine T. und Avery, Robert A.(1999). “Left-Hemisphere advantage for click consonants is determined by linguistic significance and experience”, In: *Psychological Science* 10,1 (1999), S.65-70

Bortz, Jürgen (2005). *Statistik für Human- und Sozialwissenschaftler*, 6., überarbeitete Aufl., Springer Medizin Verlag, Heidelberg

Bortz, Jürgen und Döring, Nicola (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, 4.,überarbeitete Aufl., Springer Medizin Verlag, Heidelberg

Brown, Colin M. und Hagoort, Peter (2000). *The Neurocognition of language*, Oxford University Press Inc., New York

Dixon, Norman F. und Spitz, Lydia (1980). “The detection of auditory visual desynchrony”, In: *Perception* 9 (1980), S.719-721

Dumay, N. und Frauenfelder, U.H. und Content, A. (2002). “The Role of the syllable in lexical segmentation in French: Word-Spotting Data”, In: *Brain and Language* 81 (2002), S.144-161

Eckstein, Peter P. (2012). *Angewandte Statistik mit SPSS - Praktische Einführung für Wirtschaftswissenschaftler*, 7., überarbeitete Aufl., Springer Gabler Verlag, Wiesbaden

Eysenck, Michael W. und Keane, Mark T. (2005). *Cognitive Psychology - A student's handbook*, 5.Aufl., Psychology Press Ltd., New York

Friederici, Angela D. (2002). “Towards a neural basis of auditory sentence processing”, In: *TRENDS in Cognitive Sciences* 6, 2 (2002), S.78-84

Friedrich, Hans-Jörg (2008). *Tontechnik für Mediengestalter*, Springer-Verlag, Berlin Heidelberg

IPA. (1999). *Handbook of the International Phonetic Association - A guide to the use of the International Phonetic Alphabet*, Cambridge University Press, New York

Heide, Berta Luise (2009). *Experimentelle Untersuchungen zur perzeptiven Synchronität von Bild und Ton*, Magisterarbeit, Fachbereich Audiokommunikation, TU-Berlin

Höhle, Barbara (Hrsg.) (2012). *Psycholinguistik*, 2.Aufl., Akademie Verlag GmbH, Berlin

Holman, Tomlinson (2010). *Sound for Film and Television*, 3.Aufl., Focal Press, Burlington, USA

Kroll, Judith F.; Michael, Erica; Tokowicz, Natasha und Dofour, Robert (2002). “The development of lexical fluency in a second language”, In: *Second Language Research* 18, 2 (2002), S. 137-171

Kroll, Judith F. und Stewart, Erika (1994). “Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations”, In: *Journal of memory and language* 33 (1994), S.149-174

Kutas, Marta und Hillyard, Steven A. (1980). “Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity”, In: S*cience, New Series* 207, 4427 (1980), S.203-205

Liberman, Alvin M. und Mattingly, Ignatius G. (1985). “The motor theory of speech perception revises”, In: *Cognition* 21 (1985), S.1-31

Macaluso, E. und George, N. und Dolan, R. und Spence, C. und Driver, J (2003). “Spatial and temporal factors during processing of audiovisual speech a PET study”, In: *NeuroImage* 21 (2004), S. 725-732

MacDonald, John und McGurk, Harry (1978). “Visual influences on speech perception processes”, In: *Perception & Psychophysics* 24(3) (1978), S.253-257

Marslen-Wilson, W. und Tyler, L.K. (1980). “The temporal structure of spoken language understanding”, In: *Cognition* 8 (1980), S.1-71

McGurk, Harry und MacDonald, John (1976). “Hearing lips and seeing voices”, In: *Nature* 264 (1976), S.746-748

Mueller, Jutta L. und Rüschemeyer, Shirley-Ann und Friederici, Angela D. (2006). „Aktivitätsmuster im Gehirn: Unterschiede und Gemeinsamkeiten beim Verstehen von Erst- und Zweitsprache”, In: *Perspektiven der Hirnforschung- Neuroforum* 2 (2006), S. 176-183

Näätänen, R. und Tervaniemi, M. und Sussman, E. und Paavilainen, P. und Winkler, I. (2001). “'Primitive intelligence' in the auditory cortex”, In: *TRENDS in Neurosciences* 24 (2001), S.283-288

Navarra, J. und Vatakis, A. und Zampini, M. und Soto-Faraco, S. und Humphreys, W. und Spence, C. (2005). “Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration”, In: *Cognitive Brain Research* 25 (2005), S.499-507

Navarra, J. und Alsius, A. und Velasco, I. und Soto-Faraco, S. und Spence, C. (2010). “Perception of audiovisual speech synchrony for native and non-native language”, In: *Brain Research* 1323 (2010), S. 84-93

Otake, T. und Hatano, G. und Cutler, A. und Mehler, J. (1993). “Mora or Syllable? Speech Segmentation in Japanese”, In: *Journal of memory and language* 32 (1993), S.258-278

Pisoni, David B. und Remez, Robert E. (2005). *The Handbook of speech perception*, Blackwell Publishing Ltd., USA

Rudloff, Ingo (1997). *Zusammenfassung der Diplomarbeit "Untersuchungen zur wahrgenommenen Synchronität von Bild und Ton bei Film und Fernsehen",* Diplomarbeit, Ruhr-Universität Bochum

Scherer, Klaus R. und Banse, Rainer und Wallbott, Harald G. (2001). “Emotional interferences from vocal expression correlate across languages and cultures”, In: *Journal of cross-cultural psychology* 32,1 (2001), S.76-92

Schmidt, Robert F. und Schaible, Hans-Georg (2006). *Neuro- und Sinnesphysiologie,* 5., neu bearbeitete Aufl., Springer Medizin Verlag, Heidelberg

Scott, Sophie K. und Wise, Richard J.S. (2004). “The functional neuroanatomy of prelexical processing in speech perception”, In: *Cognition* 92 (2004), S.13-45

Slavik, Karl M. und Stefan Weinzierl (2008): *Handbuch der Audiotechnik*, Springer-Verlag, Berlin Heidelberg, S. 609-685.

Stevenson, R.A. und Altieri, N.A. und Kim, S. und Pisoni, D.B. und James, T.W. (2009). “Neural processing of asynchronous audiovisual speech perception”, In: *NeuroImage* 49 (2010), S. 3308-3318

van Wassenhove, Virginie und Grant, Ken W. und Poeppel, David (2005). “Temporal window of integration in auditory-visual speech perception”, In: *Neuropsychologia* 45 (2007), S. 598-607

van Heuven, Walter J.B. und Dijkstra, Ton (1998). “Orthographic Neighborhood Effects in Bilingual Word Recognition”, In: *Journal of memory and language* 39 (1998), S.458-483

Vatakis, Argio und Navarra, Jordi und Soto-Faraco, Salvador und Spence, Charles (2007). “Audiovisual temporal adaption of speech: temporal order versus simultaneity judgments”, In: *Exp Brain Res* 185 (2008), S. 521-529

Vroomen, Jean und Stekelenburg, Jeroen J. (2011). “Perception of intersensory synchrony in audiovisual speech: Not that special”, In: *Cognition* 118 (2011), S. 78-86

Vroomen, Jean und Keetels, Mirjam (2010). “Perception of intersensory synchrony: A tutorial review”, In: *Attention, Perception, & Psychophysics* 72 (4) (2010), S.871-884

Ward, Jamie (2010). *The Student’s Guide to cognitive neuroscience*, 2.Aufl. Psychology Press, East Sussex, New York

Werker, Janet F. und Tees, Richard C. (1992). “The organization and reorganization of human speech perception”, In: *Annual Review of Neuroscience* 15 (1992), S.377-402

6.1.1 Internetquellen

[1] The Virtual Linguistics Campus (2001-2013): *Persian (Farsi): The North Wind and the Sun*
http://linguistics.online.uni-marburg.de/free/generalmodules/languages/farsi/info/file_4.html (Zugriff: 19. 08. 2012)

[2] Beck, Julia; Frenzel, Juliane; Oestreich, Raimar; Riedl, Simon (2004): *Filmsound*
http://server4.medienkomm.uni-halle.de/filmsound/ (Zugriff: 19. 02. 2013)

[3] Gegenfurtner, Karl R.; Walter, Sebastian; Braun, Doris I. (2002):*Visuelle Informationsverarbeitung im Gehirn*
http://www.allpsych.uni-giessen.de/karl/teach/aka.htm (Zugriff: 19. 02. 2013)

[4] Ruhr Uni Bochum - Guski, R. "Wahrnehmen - ein Lehrbuch" (2006): *Kap_4_2*
http://eco.psy.ruhr-uni-bochum.de/ecopsy/download/Guski-Lehrbuch/Kap_4_2.html (Zugriff: 19. 02. 2013)

[5] The Accent Lab (2011): *Farsi| The Accent Lab*
http://theaccentlab.com/wiki/accent-wiki/non-native-pronunciations-of-english/persian/ (Zugriff: 20. 02. 2013)

[6] TU Berlin (2007-2008): *TU Berlin> Institut für Sprache und Kommunikation* http://fak1-alt.kgw.tu-berlin.de/call/linguistiktutorien/phonologie/phonologie%20k3.html (Zugriff: 22. 02. 2013)

[7] Benutzer Mrmryrwrk' (2008): *Datei: Kohortenmodell-Schema.png - Wikipedia* http://de.wikipedia.org/w/index.php?title=Datei:Kohortenmodell-Schema.png&filetimestamp=20081229132221 (Zugriff: 23. 02. 2013)

[8] Lehmann, Marie-Luise (2009): *Gehirnatlas - Primäres und sekundäres Sehzentrum* http://www.gehirn-atlas.de/sehzentrum.html (Zugriff: 11. 03. 2013)

[9] Woodbury und Delgutte (2007): *Audiovisual Integration in Speech Perception* http://web.mit.edu/hst.722/www/Topics/AudiovisualIntegration.htm (Zugriff: 11.03.2013)

[10] Reiter, Thomas (2005): *Zurück zu Post-Hoc Test* http://www.reiter1.com/Glossar/Post_Hoc_Test.htm (Zugriff: 11. 03. 2013)

[11] Leard Statistics (2013): *Friedman Test in SPSS* https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php (Zugriff: 11. 03. 2013)

[12] Dr. rer. nat. Noesselt, Tömme (2005-2006): *Department of Neurolo* http://neuro2.med.uni-magdeburg.de/neurologie/en/forschung/multimodale_aufmerksamkeit.php (Zugriff: 11. 03. 2013)

6.1.2 Software

[a] http://www.fon.hum.uva.nl/praat/ (Zugriff: 19. 02. 2013)
[b] http://www.sonycreativesoftware.com/vegaspro (Zugriff: 16. 01. 2013)
[c] http://www.mediacoderhq.com/ (Zugriff: 16. 01. 2013)
[d] http://www-01.ibm.com/software/de/analytics/spss/ (Zugriff: 20. 03. 2013)
[e] http://www.mathworks.de/products/matlab/ (Zugriff: 20. 03. 2013)

6.2 Abbildungsverzeichnis

Abbildung 1: Aufbau des menschlichen Auges (Schmidt et al., 2006, S.244)..... 9
Abbildung 2: visuelles Netzwerk (Lehmann, 2009)..... 11
Abbildung 3: visuelle Informationsverarbeitung (Ward, 2010, S.110) 12
Abbildung 4: Schematische Darstellung des Ohres (Schmidt et al., 2006, S. 292)..... 14
Abbildung 5: Schematische Darstellung der zentralen Hörbahn (Schmidt et al., 2006, S.304)..... 15
Abbildung 6: 2-Wege-Theorie zur Sprachwahrnehmung und -produktion (Ward, 2010, S.228)..... 19
Abbildung 7: Kohorten-Modell (Bildquelle: [8]) 22
Abbildung 8: Revised-Hierarchical-Model (Höhle, 2012, S.180; nach Kroll und Stewart 1994) 24
Abbildung 9: Sprachbezogene EKP-Werte: (a) N400, (b) very early left-anterior negativity (ELAN), (c) P600 (Friederici, 2002)..... 26
Abbildung 10: Transkription der Aesop-Fabel auf Deutsch (Handbook of the IPA, 1999, S.89)..... 34
Abbildung 11: Phonemische Transkription der Aesop-Fabel auf Deutsch (Handbook of the IPA, 1999, S.88)..... 34
Abbildung 12: Transkription der Aesop-Fabel auf Englisch (Handbook of the IPA, 1999, S.44)..... 35
Abbildung 13: Phonemische Transkription der Aesop-Fabel auf Englisch (Handbook of the IPA, 1999, S.44)..... 35
Abbildung 14: (phonemische) Transkription der Aesop-Fabel auf Persisch ([1], Mina Fallahi)..... 36
Abbildung 15: möglicher Progammstatus - "Eiusdem Temporis"..... 40
Abbildung 16: Screenshot des Programms "Eiusdem Temporis" 41
Abbildung 17: Versuchsaufbau im MiniLab der TU Berlin..... 42
Abbildung 18: Oszilloskopbessung 43
Abbildung 19: Datentransformation von Excel zu SPSS..... 46
Abbildung 20: Balkendiagramm der summierten korrekten Synchronurteile für Deutsch (d), Englisch (e) und Persisch (p)..... 48
Abbildung 21: Häufigkeitsverteilung richtiger Synchronurteile für die Zeitversätze -120ms und +120ms; 50
Abbildung 22: Häufigkeitsverteilung richtiger Synchronurteile für den Zeitversaz -40ms; blau = Deutsch, grün = Englisch, rot = Persisch..... 50

Abbildung 23: Liniendiagramm Mittelwerte fehlerhafter Synchronurteile in Prozent..... 51
Abbildung 24: Kurvenanpassung am Beispiel der Bewertung des deutschen Materials 57
Abbildung 25: Gegenüberstellung der Fehlurteile von Männern und Frauen 60
Abbildung 26: Excel-Grafik der PSS- und JND-Werte für Männern und Frauen 61
Abbildung 27: Excel-Grafik der JND-Werte für Deutsch (d), Englisch (e) und Persisch (p) 66
Abbildung 28: Sprachrelevante Regionen (Berwick et al., 2013, S.94)..... 69
Abbildung 29: Neurocognitives Modell der auditiven Satzverarbeitung (Friederici, 2002, S.79) 70
Abbildung 30: Schmalband-Spektogramm für die deutsche Aufnahme 86
Abbildung 31: Schmalband-Spektogramm für die englische Aufnahme 86
Abbildung 32: Schmalband-Spektogramm für die persische Aufnahme 87
Abbildung 33: Schaltbild der Konstruktion zur Messung der Helligkeitsschwankungen88
Abbildung 34: Foto der Konstruktion zur Messung der Helligkeitsschwankungen 89

6.3 Tabellenverzeichnis

Tabelle 1: Frameraten für Film und Video mit Timecodes und Sample Raten (Holman, 2010, S.142) 4
Tabelle 2: MPEG Audiocodecs im Überblick (Friedrich, 2008, S.196) 5
Tabelle 3: Vergleich der Anzahl an Indikatoren für Deutsch, Englisch und Persisch ... 37
Tabelle 4: Vergleich für Energie, Frequenz, Jitter und Shimmer für Deutsch, Englisch und Persisch 38
Tabelle 5: Summe korrekter Synchronurteile für Deutsch (d), Englisch (e) und Persisch (p) 48
Tabelle 6: Exceltabelle der Häufigkeiten korrekter Synchronurteile, inkl. der Mittelwerte 49
Tabelle 7: Exceltabelle der Mittelwerte fehlerhafter Synchronurteile in Prozent 51
Tabelle 8: Ergebnisse der einfaktoriellen ANOVA 52
Tabelle 9: Ergebnisse des Kolmogorov-Smirnov-Anpassungstest für Normalverteilung 53
Tabelle 10: Ergebnisse des parameterfreien Friedman-Tests 54
Tabelle 11: Ergebnisse des Post-Hoc Wilcoxon-Tests 55
Tabelle 12: Tabelle der Zusammenhangsmaße von Sprache und prozentualen Synchronurteilen für die ANOVA 56
Tabelle 13: PSS und JND-Werte für Deutsch, Englisch und Persisch 58
Tabelle 14: SPSS-Tabelle der PSS- und JND-Varianzanalyse 58
Tabelle 15: Zusammenhangsmaße für Sprache und PSS- bzw. JND-Werte 59
Tabelle 16: Tabelle der PSS- und JND-Werte für Männer und Frauen 60
Tabelle 17: Auszug aus der Post-Hoc-ANOVA-Tabelle der Vergleiche zeitlicher Versätze 62
Tabelle 18: An die verwendeten Zeitversätze angepasste Ober- und Untergrenzen der Synchronitätsfenster 66
Tabelle 19: Breite des Synchronitätsfensters für Deutsch, Englisch und Persisch auf Basis der JND-Werte 67

7. Anhang

7.1 Farbwahrnehmung

Wie bereits im Kapitel 2.1.1 beschrieben, ist das P-System für die Verarbeitung von Farbinformationen zuständig. Die eigentliche Farbwahrnehmung entsteht im Gehirn. Wie sie sich allerdings genau zusammensetzt, wird versucht durch unterschiedliche Theorien zu erklären. Das Farbsehen beruht auf den Zapfentypen, welche auf Licht unterschiedlicher Wellenlängen reagieren. Die Dreifarbentheorie von Young, Helmholtz und Maxwell aus dem 19.Jahrundert postuliert, dass bei der im Gegensatz zur physiologischen, additiven Farbmischung, bei der durch Mischung von Komplementärfarben Weiß entstehen kann, eine subtraktive Farbmischung durchgeführt wird. Physikalisch betrachtet werden Anteile des weißen Lichts absorbiert und die verbleibende Wellenlänge bestimmt die wahrgenommene Farbe. So würde eine Mischung aus langwelligem Blau und kurzwelligem Gelb die Farbe Grün im mittleren Wellenlängenbereich ergeben.

Ewald Hering legte im selben Jahrhundert seine Gegenfarbentheorie als Alternative dar. Der zufolge liegen Gegenfarbenneuronen für Rot-Grün und Gelb-Blau bei den Horizontalzellen der Netzhaut vor. Diese sind ebenfalls in den P-Zellen im CGL wiederzufinden, obwohl sie hier nicht farbspezifisch sind. Erst Doppelgegenfarbneuronen reagieren auf Farbkontraste zwischen Farbzentrum und Farbperipherie und lassen sich erstmals in der primären Sehrinde finden.

Eine weitere Theorie aus dem 19.Jahrhundert ist die Kries-Zonentheorie, welche die bereits genannten Theorien vereint. Sie beschreibt, dass sowohl die Verrechnung in der Netzhaut als auch den höheren Sehrindenabschnitten zur Farbwahrnehmung beiträgt und durch einen Vergleich der Wellenlängenanteile ermittelt wird. Hier spielt noch eine Objektkomponente und eine Beleuchtungskomponente eine wichtige Rolle, da sich das Farbempfinden eines Objektes je nach Beleuchtungsart (Tageslicht, Glühlampe, etc.) ändert. Dies ist beim Farbfilm zu bemerken, da eine Kamera keinen Korrekturmechanismus wie im menschlichen Gehirn besitzt, wodurch dasselbe Objekt abhängig von der Beleuchtung in unterschiedlichen Farben gezeigt wird. Das menschliche Gehirn hingegen erkennt die unterschiedliche Beleuchtung und nimmt die Farbe als konstant wahr, auch wenn die Wellenlängen durch Beleuchtung und Reflexion an der Objektoberfläche leicht abweichen. Dieses Ereignis wird als Farbkonstanz bezeichnet (Schmidt et al., 2006, S. 282-285; Ward, 2010, S. 111).

7.2 Praat-Analyse

In den folgenden Screenshots aus der Sprachanalysesoftware Praat sind jeweils Schmalband-Spektrogramme der aufgezeichneten Sätze für den Hörversuch dargestellt. Zusätzlich sind Pulses (blau oben), Intensitäten (gelb) und Tonhöhen (blau unten) eingezeichnet. Die Screenshots befinden sich ebenfalls auf der beiliegenden CD in der Powerpoint-Datei " Praat_Spektrogrammen_Screenshots.pptx".

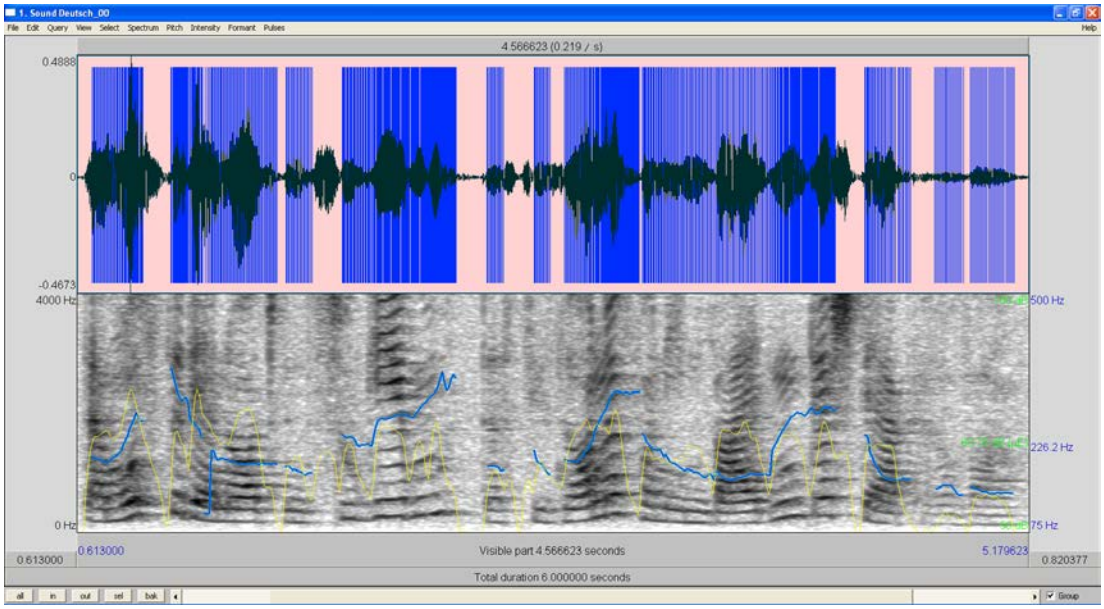

Abbildung 30: Schmalband-Spektogramm für die deutsche Aufnahme

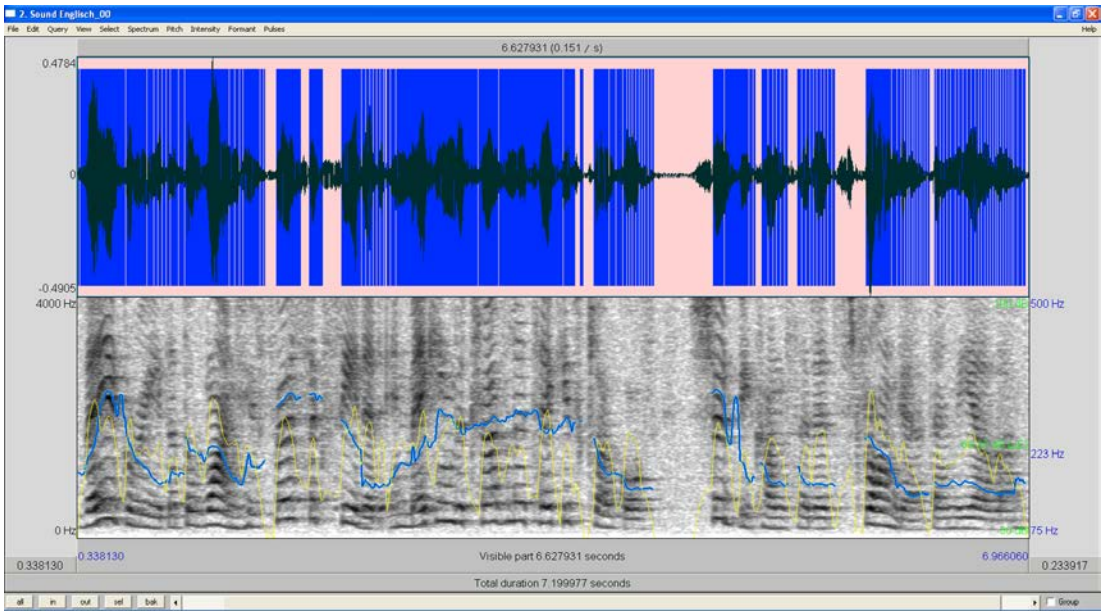

Abbildung 31: Schmalband-Spektogramm für die englische Aufnahme

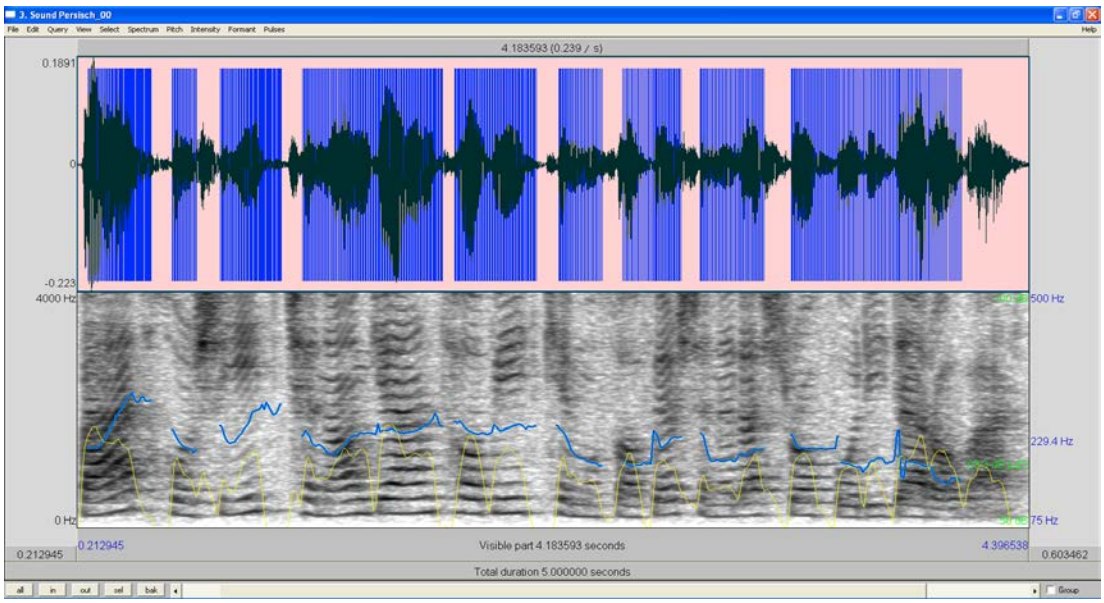

Abbildung 32: Schmalband-Spektogramm für die persische Aufnahme

7.3 „Eiusdem Temporis"

Das Programm "Eiusdem Temporis" von Stefan Schütz kann in einer separaten zip-Datei auf der beigelegten CD gefunden werden.

7.4 Schaltbild

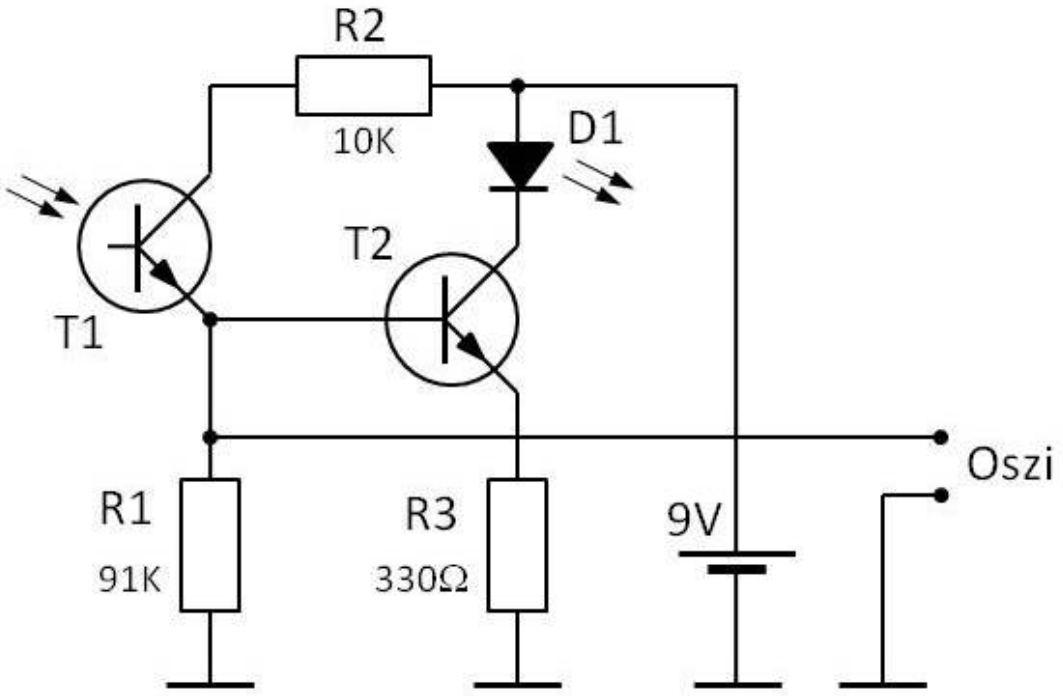

Abbildung 33: Schaltbild der Konstruktion zur Messung der Helligkeitsschwankungen

Die obige Schaltung zeigt die für die Synchronitätsmessung verwendete Konstruktion. Sie dient der Gewinnung eines elektrischen Signals aus der Helligkeitsänderung des wiedergegebenen Videos. Mithilfe eines Zweistrahl-Oszillografen kann die Zeitdifferenz zwischen dem vom Kondensator-Mikrofon erfassten akustischen Signal und der vom Fototransistor erfassten Helligkeitsänderung gemessen und die zeitliche Verschiebung von Bild und Ton bestimmt werden. Dadurch ist das wichtigste Bauelement in der Schaltung der Fototransistor T1. Er erlaubt beim Überschreiten einer bestimmten Beleuchtungsstärke, welche durch den Monitor gegeben ist, einen Stromfluss vom Kollektor zum Emitter, wodurch am Widerstand R1 eine Spannung gemessen werden kann. Der Spannungssprung am Widerstand dient als Zeitmarke für die Messung mit dem Zweistrahl-Oszillografen.

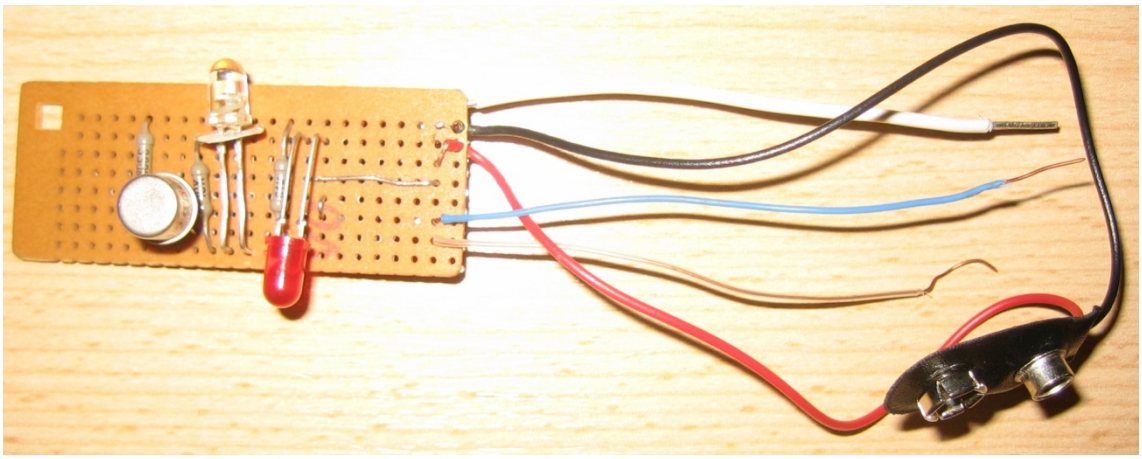

Abbildung 34: Foto der Konstruktion zur Messung der Helligkeitsschwankungen

Der Schalttransistor T2 und die Leuchtdiode D1 sind für die Einrichtung des Aufbaus und für Kontrollzwecke von Bedeutung. Der genannte Spannungssprung an R1 wird an der Basis von T2 als Eingangssignal verwendet, steuert den Transistor durch und erlaubt einen Stromfluss über Kollektor und Emitter sowie über den Widerstand R3, welche D1 zum Leuchten bringt. Dieser Teil der Schaltung wurde vor der eigentlichen Messung verwendet, um den Fototransistor so zu platzieren, dass der Helligkeitswechsel eindeutig erkannt wird. Während der Messung diente er der optischen Kontrolle für das Funktionieren der Schaltung.