



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Lieben oder leben? Diskrimination des deutschen /i:/-/e:/-
Vokalkontrasts durch Hörer/innen mit L1 Japanisch“

verfasst von / submitted by

David Eppensteiner BA BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 814

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Deutsch als Fremd- und Zweitsprache

Betreut von / Supervisor:

Dipl.-Päd. Dr. Anke Sennema MA

Inhaltsverzeichnis

1	Einleitung.....	5
1.1	Erkenntnisinteresse und Zielsetzung	5
1.2	Gang der Untersuchung	7
1.3	Anmerkungen zur Begriffsverwendung	8
2	Wie nehmen wir (fremdsprachliche) Laute wahr?.....	9
2.1	Eigenschaften des sprachlichen Signals	10
2.2	Das native Phoneminventar	11
2.3	Entwicklungspfade für die Unterscheidung fremdsprachlicher Kontraste.....	14
2.4	Zusammenhang zwischen Produktion und Perzeption	15
2.5	Kategoriale vs. kontinuierliche Wahrnehmung	17
3	Einflussfaktoren für die Wahrnehmung von Vokalen der L2.....	19
3.1	Cross-linguistic Influence.....	20
3.2	Ähnlichkeit	22
3.3	Spracherfahrung in / Sprachkontakt mit der L2	25
3.4	Alter	27
3.5	Aufmerksamkeit	28
4	Kontrastierung des deutschen und japanischen Vokalsystems	30
4.1	Artikulatorische Grundlagen	30
4.2	Das deutsche Vokalsystem	31
4.3	Das japanische Vokalsystem	33
4.4	Kontrastierung beider Vokalsysteme.....	35
4.5	Der /i:/-/e:/-Vokalkontrast	36
5	Vorhersagen phonetischer Modelle	40
5.1	Das Perceptual Assimilation Model (PAM).....	40
5.2	Das Speech Learning Model (SLM) und das PAM-L2.....	42

5.3	Hypothesen für den empirischen Teil.....	46
6	Empirischer Teil.....	47
6.1	Forschungsdesign	47
6.2	Visualisierung des Forschungsdesigns	50
6.3	Maßnahmen zur Sicherstellung der Gütekriterien.....	51
6.4	Experiment 1 – lieben oder leben – Diskrimination.....	53
6.4.1	Proband/inn/en	53
6.4.2	Stimuli	54
6.4.3	Ablauf.....	56
6.4.4	Ergebnisse	57
6.5	Experiment 2 – ABC der Familiennamen – Identifikation.....	66
6.5.1	Proband/inn/en	66
6.5.2	Stimuli	66
6.5.3	Ablauf.....	68
6.5.4	Ergebnisse	69
7	Zusammenfassung und Diskussion der Ergebnisse	70
8	Resümee und Ausblick	80
9	Literaturverzeichnis	82
10	Anhang.....	88

Abbildungsverzeichnis

Abbildung 1: Formantkarte des Deutschen; Mittelwerte der Formanten aller Sprecher (aus Sendlmeier/Seebode 2007).....	10
Abbildung 2: Entwicklungspfade für die Unterscheidung von Kontrasten (aus Mazuka et al. 2013: 193)	14
Abbildung 3: Konfusionsmatrix; richtig und falsch zugeordnete Identifikationen in Prozent (aus Kerschhofer-Puhalo 2014: 242).	24
Abbildung 4: Identifikationsrate von /e:/ nach Erstsprachen gegliedert (aus Kerschhofer-Puhalo 2014: 575)	25
Abbildung 5: Japanisches und deutsches Vokalsystem (deutsche Vokale sind mit schwarzen, japanische mit roten Punkten gekennzeichnet, nach eLaut Japanisch und Deutsch).....	35
Abbildung 6: Artikulation des [i] des Deutschen (a)-(b) Mundformen (Front- und Seitenansicht), (c) Zungenstellung (Wängler 1976, aus Kaneko 2005: 38).....	36
Abbildung 7: Artikulation des [e] des Deutschen (a)-(b) Mundformen (Front- und Seitenansicht), (c) Zungenstellung (Wängler 1976, aus Kaneko 2005: 40).....	37
Abbildung 8: Artikulation des japanischen i-Lautes: (a) Lippenstellung, (b) Mundformenfoto (Seitenansicht), (c) Zungenstellung (aus Kaneko 2005: 10)	38
Abbildung 9: Artikulation des japanischen e-Lautes: (a) Lippenstellung, (b) Mundformenfoto, (c) Zungenstellung (aus Kaneko 2005: 10)	38
Abbildung 10: PAM-Vorhersagen	41
Abbildung 11: Vergleich von /i:/ (durchgezogene Linie) und /e:/ (strichlierte Linie) im /pVt/-Kontext (aus Kerschhofer-Puhalo 2014: 177).....	44
Abbildung 12: Visualisierung des Forschungsdesigns.....	50
Abbildung 13: Oszillogramm der Stimuli ['li:bŋ] (oben) und ['le:bŋ] (unten).....	55
Abbildung 14: ExperimentMFC-Folie während des Diskriminationsexperiments.....	56
Abbildung 15: Diskriminationsrate für Teil 1 in Prozent korrekt (ohne SSN)	58
Abbildung 16: Diskriminationsrate für Teil 2 in Prozent korrekt (SNR 0dB).....	60
Abbildung 17: Diskriminationsrate für Teil 3 in Prozent korrekt (SNR -5dB)	61
Abbildung 18: Vergleich der drei Experimentteile für Gruppe A	62
Abbildung 19: Vergleich der drei Experimentteile für Gruppe BC	63
Abbildung 20: Fehlerverteilung Gruppe A	64
Abbildung 21: Fehlerverteilung Gruppe BC.....	65
Abbildung 22: Spektogramm von /e:/ (oben) und /i:/ (unten) in Isolation.....	67

Abbildung 23: ExperimentMFC-Folie während des Identifikationsexperiments	68
Abbildung 24: Vokaltrapez Deutsch (deutsche Vokale sind in schwarzer, japanische in roter Farbe)	71
Abbildung 25: Ergebnisse von Ito (rot eingekreist) (aus Mazuka et al. 2013: 196).	77

Tabellenverzeichnis

Tabelle 1: Vokale des Deutschen (nach Dieling 1992: 119).....	32
Tabelle 2: Vokale des Japanischen (nach Dieling 1992: 86)	33
Tabelle 3: Akustische Messungen der beiden Zielvokale am Beispiel eines Minimalpaares..	54
Tabelle 4: Maße der zentralen Tendenz und Standardabweichung für Teil 1 (ohne SSN).....	58
Tabelle 5: Maße der zentralen Tendenz und Standardabweichung für Teil 2 (0dB SNR)	59
Tabelle 6: Maße der zentralen Tendenz und Standardabweichung für Teil 3 (-5dB SNR).....	61
Tabelle 7: Ergebnisse des Identifikationsexperiments (in absoluten Werten)	69

1 Einleitung

1.1 Erkenntnisinteresse und Zielsetzung

Denn daß niemand den andern versteht, daß keiner bei denselben Worten dasselbe, was der andere, denkt, daß ein Gespräch, eine Lektüre bei verschiedenen Personen verschiedene Gedankenfolgen aufregt, hatte ich schon allzu deutlich eingesehen (Goethe 1998: 738).

Sprechen ist ein handlungsorientierter performativer Akt. Gerade mündliche Sprache ist daher von Missverständnissen verschiedenster Art geprägt, die im besten Fall Stirnrunzeln, im schlimmsten Fall heftige Auseinandersetzungen hervorrufen. Oder um es mit Ursula Hirschfeld (1995: 177) zu sagen: „Abweichungen verschiedener Art, Störungen in der Produktion, Übertragung und Perzeption von Sprache stellen an sich keine Besonderheit dar – sie sind die Regel“. In dieser Arbeit soll es um auditive Verständnisschwierigkeiten in der Fremd- oder Zweitsprache gehen, die wesentlich durch den erstsprachlichen Einfluss (in diesem Fall Japanisch) auf die Wahrnehmung bedingt sind. Um mein Erkenntnisinteresse und meine Motivation für diese Untersuchung darzulegen, möchte ich mit einer kleinen Anekdote beginnen:

Vor etwa vier Jahren war ich das erste Mal in Japan, um mein Japanisch zu verbessern und dabei gleichzeitig die japanische(n) Kultur(en) und alles was dazugehört kennenzulernen. Zu diesem Zeitpunkt hatte ich erst seit etwa sechs Monaten im Selbststudium Japanisch gelernt. Meine Sprachkenntnisse waren also – gelinde gesagt – noch bescheiden. Auf meiner Reise, die mich von Tokio bis nach Hiroshima führte, habe ich allerdings eine interessante Entdeckung gemacht. In zahlreichen Gesprächen mit japanischen Sprecher/inne/n, etwa beim Check-In in einem Hotel oder beim Bestellen eines Gerichts im Restaurant, vernahm ich oftmals den Ausdruck „ええっと“ (gesprochen in etwa [ɛ:tɔ]), kannte aber zu diesem Zeitpunkt dessen Bedeutung noch nicht. Als ich einen japanisch sprechenden Koreaner, den ich auf der Reise kennengelernt hatte, darauf angesprochen habe, was „ええっと“ übersetzt bedeutet, verstand er mich zuerst nicht. Dies lag daran, dass ich fälschlicherweise nicht [ɛ:tɔ], sondern [e:tɔ] gesagt habe. Ich habe also den japanischen e-Laut durch das deutsche gespannt-lange e /e:/ ersetzt, was bei meinem Gesprächspartner Verwirrung hervorgerufen hat, da er es wiederum als [i:tɔ] wahrgenommen hat. Im weiteren Verlauf meines Aufenthalts in Japan und noch viele Male danach kam es in Unterhaltungen mit Japaner/inne/n immer wieder zu solchen kleinen „Missverständnissen“, bei denen ein langes e als i wahrgenommen worden war, weshalb ich zu

dem Schluss gekommen bin, dass Japaner/innen offensichtlich Schwierigkeiten haben, das lange, halbgeschlossene und ungerundete e (= [e:]) als solches zu identifizieren, da dieses in der japanischen Sprache nicht vorkommt. Es wurde meiner Erfahrung nach meist als geschlossenes, ungerundetes i (= [i:]) wahrgenommen.

Aus diesen Beobachtungen erwuchs das Forschungsinteresse die Wahrnehmung bzw. genauer gesagt die Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts durch Hörer/innen mit Erstsprache Japanisch empirisch zu erforschen. Es sollte herausgefunden werden, ob dieser Kontrast in einem AXB-Diskriminationsexperiment Probleme bereitet. Im Zuge meines Auslandspraktikums in Japan hatte ich Gelegenheit das beobachtete Phänomen experimentell zu untersuchen. Japanische Student/inn/en einer Universität nahe Tokio wurden dazu getestet. Als Basis für meine Untersuchung dient ein Experiment, welches ich im Zuge eines Phonetikseminars vor vier Jahren erstellt habe und für die Masterarbeit umfassend überarbeitet habe. So wurden beispielsweise dem ursprünglichen Experimentdesign mittels Störgeräuschen verschiedene Schwierigkeitsgrade hinzugefügt, um die Wahrnehmung unter erschwerten Bedingungen testen zu können.

Ein Schwerpunkt dieser Arbeit liegt außerdem auf der Bestimmung und Diskussion von Faktoren, welche voraussichtlich wesentlichen Einfluss auf die fremdsprachliche Lautwahrnehmung nehmen. So soll etwa untersucht werden, ob die sprachliche Erfahrung in der Zweitsprache für besagten Vokalkontrast eine Rolle spielt. Dazu werden zwei Gruppen japanischer Student/inn/en – mit und ohne Deutschkenntnissen – getestet. Gleichzeitig werden dabei die Vorhersagen zweier phonetischer Modelle, des Perceptual Assimilation Models (PAM bzw. PAM-L2) von Best/Tyler und des Speech Learning Models (SLM) von Flege überprüft. Zusätzlich wird in einem weiteren kleinen Experiment die Identifikation der beiden Vokale durch japanische Deutschlernende getestet, um ein umfassenderes Bild fremdsprachlicher Wahrnehmung zu erhalten. Die Forschungsfragen, die sich aus dem Erkenntnisinteresse ergeben, lauten wie folgt:

1. Können japanische L1-Hörer/innen (mit und ohne DaF-Erfahrung) den deutschen /i:/-/e:/-Vokalkontrast – unter den durch das Experiment vorgegebenen Bedingungen – unterscheiden? Wenn ja, wie gut?
2. Welche Faktoren/Variablen haben Einfluss auf die fremdsprachliche Wahrnehmung der beiden Vokale? Spielt die sprachliche Erfahrung in der Zweitsprache eine Rolle?
3. (Inwiefern) verändert sich das Ergebnis unter erschwerten Bedingungen (SNR 0dB/-5dB) im Vergleich zu deutschen L1-Hörer/inne/n?

Um das Rätsel des (mir) unbekanntes Wortes aufzulösen: „ええっと“ kann mit „Hmm...“ übersetzt werden. Scheinbar habe ich meine Kommunikationspartner/innen durch meine Fragen oft zum Nachdenken gebracht.

1.2 Gang der Untersuchung

Es handelt sich bei dieser Studie um eine quantitative experimentelle Forschung, die durch Theorie gestützt und wesentlich ergänzt wird. So soll etwa der erste Teil der zweiten Forschungsfrage „Welche Faktoren/Variablen haben Einfluss auf die fremdsprachliche Wahrnehmung der beiden Vokale?“ vorrangig mit Rückgriff auf andere Literatur besprochen werden. Die Arbeit besteht im Kern aus einem Theorieteil (Kapitel 2-5), der in das Thema einführt und den Forschungsstand zum Thema beinhaltet und einen empirischen Teil (Kapitel 6). Der Aufbau der Masterarbeit soll nachfolgend genauer beschrieben werden.

Der Theorieteil gliedert sich in vier Teile. Das erste Theoriekapitel (Kapitel 2) führt an das Thema heran und setzt sich ganz allgemein mit der Frage auseinander wie fremdsprachliche Laute bzw. Sprache im Allgemeinen wahrgenommen wird. Dazu wird zuerst das sprachliche Signal selbst betrachtet. Danach werden die verschiedenen Wege, die die fremdsprachliche Lautwahrnehmung nehmen kann, nachgezeichnet. Außerdem wird die Wechselbeziehung zwischen Rezeption und Produktion thematisiert und kategoriale von kontinuierlicher Wahrnehmung abgegrenzt. Grundsätzliche Konzeptionen wie etwa das Perceptual Narrowing und diverse Wahrnehmungsmodelle werden dabei erläutert.

In Kapitel 3 sollen wesentliche Einflussfaktoren auf die fremdsprachliche Lautwahrnehmung bestimmt werden, um die bereits angesprochene zweite Forschungsfrage so weit als möglich beantworten zu können. Als zentrale Variablen für die L2-Lautwahrnehmung wurden identifiziert: Cross-Linguistic Influence (CLI), Ähnlichkeit, sprachliche Erfahrung in bzw. Sprachkontakt mit der Fremd- oder Zweitsprache, Alter und Aufmerksamkeit. Dazu kommen später thematisierte methodologische Faktoren, die durch den Experimentmodus bestimmt sind und vokalspezifische Merkmale.

Im nächsten Abschnitt (Kapitel 4) werden das deutsche und japanische Vokalsysteme gegenübergestellt. Hier soll auf vokalinhärente artikulatorische Merkmale eingegangen werden. Nach einer Einleitung über artikulatorische Grundlagen wird zuerst jedes Vokalsystem für sich dargestellt. Danach werden beide Systeme kontrastiert dargestellt und der für diese Arbeit entscheidende /i:/- /e:/-Vokalkontrast in artikulatorischer Hinsicht beleuchtet.

Im letzten Theoriekapitel (Kapitel 5) werden schließlich auf Grundlage von drei phonetischen Modellen (PAM, PAM-L2 und SLM) Hypothesen für das nachfolgende Diskriminationsexperiment entwickelt. Damit wird eine Brücke zum empirischen Teil geschlagen.

Die generierten Hypothesen werden im empirischen Teil (Kapitel 6) überprüft. Nach der Darstellung des Forschungsdesigns, worin Aufbau und Ablauf der beiden Experimente genau beschrieben werden, folgen Maßnahmen zur Wahrung der Gütekriterien. Danach werden die Ergebnisse im Sinne einer transparenten und nachvollziehbaren Arbeitsweise präsentiert und visualisiert.

In der Interpretation und Diskussion der Resultate wird schließlich ein Bogen zur Einleitung gespannt und die eigene Forschung kritisch betrachtet (Kapitel 7). Im abschließenden Resümee und Ausblick (Kapitel 8) sollen dann nochmal die wesentlichen in der Arbeit gewonnenen Erkenntnisse zusammengefasst und offen gebliebene Fragen angesprochen werden. Der Anhang ist (bis auf das Abstract) auf CD-ROM ausgelagert.

1.3 Anmerkungen zur Begriffsverwendung

An dieser Stelle sollen ein paar grundsätzliche Termini und deren Verwendungsweise angeführt werden. Abkürzungen werden ab der ersten Verwendung des Begriffs im Text in Klammer angegeben.

Als *Erstsprache* oder *L1* wird in dieser Arbeit jene Sprache verstanden, die als erstes erworben wurde. Es ist auch möglich, mehrere Erstsprachen zu erlangen, wenn diese gleichzeitig erworben wurden (Bilingualität oder Multilingualität). Der Begriff *Muttersprache* wird soweit als möglich vermieden, um etwaige damit verbundene Konnotationen zu unterbinden. Anstatt *muttersprachlich* wird darüber hinaus der aus dem Englischen entlehnte Terminus *nativ* bevorzugt. So wird beispielsweise vom nativen oder erstsprachlichen Phoneminventar anstatt vom muttersprachlichen Phoneminventar gesprochen.

Zweitsprache oder *L2* bezeichnet in Analogie zur L1 die Sprache, die nach der Erstsprache gelernt oder erworben wurde. Der Zusammenhang ist also ein chronologischer. Dabei muss es sich nicht zwingend um die erste gelernte Fremdsprache handeln. L2 wird in dieser Arbeit als Sammelbegriff für jegliche nach der L1 gelernte Fremd- oder Zweitsprache verwendet. Die Begriffe *Vokal* und *Selbstlaut* werden in dieser Arbeit synonym verwendet. Als *naive* Hörer/innen werden jene Personen bezeichnet, die keinerlei sprachlichen Kontakt mit der betreffenden Sprache haben.

2 Wie nehmen wir (fremdsprachliche) Laute wahr?

In diesem einführenden Teil sollen einige grundlegende Konzepte und wesentliche Annahmen und Erkenntnisse zur Wahrnehmung von Vokalen bzw. Lauten allgemein und in der Fremd- bzw. Zweitsprache nachgezeichnet werden. Der Weg führt dabei vom Allgemeinen (Kap. 2, der Wahrnehmung von Sprache) über das Konkrete (Kap. 3, der Wahrnehmung von Vokalen in der L2) zum Speziellen (der Wahrnehmung des deutschen /i:/-/e:/-Vokalkontrasts; vor allem in Kap. 4, wenn das japanische Vokalsystem dem deutschen gegenübergestellt wird). Der Rahmen für die Darstellung des Forschungsstandes wird durch die Forschungsfragen vorgegeben:

1. Können japanische L1-Hörer/innen (mit und ohne DaF-Erfahrung) den deutschen /i:/-/e:/-Vokalkontrast – unter den durch das Experiment vorgegebenen Bedingungen – unterscheiden? Wenn ja, wie gut?
2. Welche Faktoren/Variablen haben Einfluss auf die fremdsprachliche Wahrnehmung der beiden Vokale? Spielt die sprachliche Erfahrung in der Zweitsprache eine Rolle?
3. (Inwiefern) verändert sich das Ergebnis unter erschwerten Bedingungen (SNR 0dB/-5dB) im Vergleich zu deutschen L1-Hörer/inne/n?

Zuerst eine kleine „Ent-täuschung“. Die Frage, wie wir fremdsprachliche Laute wahrnehmen, kann bis heute nicht zufriedenstellend beantwortet werden. Die komplexen Prozesse der Sprachwahrnehmung erfolgen weitgehend automatisch und unbewusst (vgl. Höhle 2010: 40). Was genau in unserem Gehirn abläuft, wenn ein fremdsprachlicher Laut unser Ohr erreicht, ist bis zum heutigen Tage unklar, da wir nicht direkt in den Kopf hineinblicken, sondern nur getätigte Aussagen analysieren und im Nachhinein dann daraus Theorien entwickeln können. Dasselbe gilt für die Laute der Erstsprache. Bislang ist nicht geklärt, „welche speziellen akustischen Hinweise im sprachlichen Signal dem Hörer die Identifizierung von bestimmten Sprachlauten ermöglichen“ (Höhle 2010: 43). Auch diese Arbeit kann keine eindeutige Antwort darauf geben. Gegenstand der Phonetik ist die gesprochene Sprache. Der Fokus dieser Arbeit liegt auf der Sprachwahrnehmung, die Produktion von Sprache soll nur soweit als erforderlich thematisiert werden, da sonst der Rahmen dieser Untersuchung gesprengt werden würde. Disziplinen, welche sich mit der Sprachwahrnehmung beschäftigen, sind die auditorische Phonetik und die Psycholinguistik (vgl. Reetz/Jongman 2009: 1).

2.1 Eigenschaften des sprachlichen Signals

Am Anfang steht das sprachliche Signal, welches „gleichzeitig [...] Produkt der menschlichen Artikulation und [...] Ausgangspunkt für den Wahrnehmungs- und Verständnisprozess gesprochener Sprache darstellt“ (Höhle 2010: 41). In Form einer Schallwelle erreicht es zuerst das Hörorgan und gelangt danach zum Gehirn, wo die Verarbeitung und Kategorisierung erfolgt. „Die Psycholinguistik geht davon aus, dass das menschliche Sprachvermögen auf einem komplexen System beruht, das das Zusammenwirken verschiedener Wissensstrukturen und Verarbeitungsprozesse in einer präzisen und schnellen zeitlichen Abstimmung erfordert“ (Höhle 2010: 12). Wie dieses genau aussieht, darüber kann nur spekuliert werden. Ein Modell, das die beim Hören von Sprache ablaufenden Wahrnehmungsprozesse zu erklären versucht, wird im nächsten Kapitel dargestellt (siehe Kap. 3.5).

Was wir allerdings aus der Betrachtung des akustischen Signals wissen: Gesprochene Sprache ist durch *mangelnde Invarianz* und *Nicht-Linearität* charakterisiert (vgl. Höhle 2013: 43, Kerschhofer-Puhalo 2014: 59). Mit dem Begriff *mangelnde Invarianz* wird darauf referiert, dass sich das Signal von Sprecher/in zu Sprecher/in und hinsichtlich der akustischen Merkmale mitunter stark unterscheidet. Verantwortlich dafür sind neben individuellen sprecherbedingten Unterschieden etwa die Position im Wort bzw. innerhalb einer Silbe und die sogenannte *Koartikulation*, d.h. die Beeinflussung der Produktion von Lauten durch vorhergehende oder nachfolgende Lautsegmente (vgl. Höhle 2013: 43). Gerade bei Vokalen sind die Grenzen zwischen den einzelnen Selbstlauten aufgrund der verschiedenen Sprechertypen fließend, wie ein Blick auf die Formantkarte des Deutschen zeigt (vgl. Huang 2011, siehe Abb. 1).

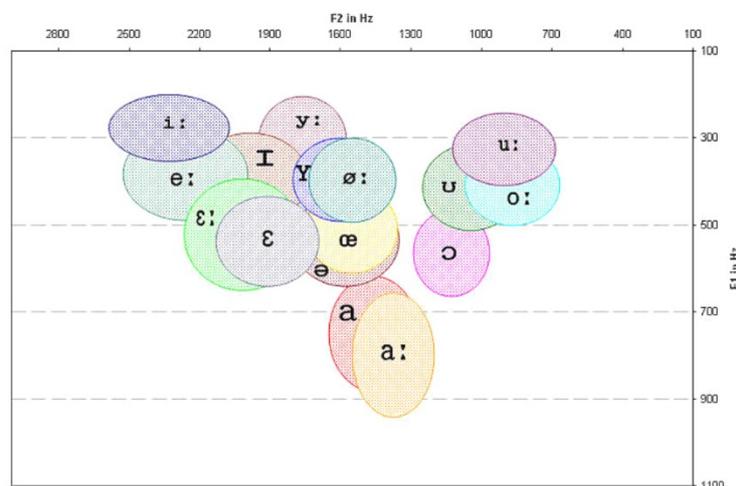


Abbildung 1: Formantkarte des Deutschen; Mittelwerte der Formanten aller Sprecher (aus Sendlmeier/Seebode 2007).

Formanten sind Resonanzfrequenzen spezifischer Obertöne, welche bestimmte artikulatorische Gegebenheiten von Vokalen (Öffnungswinkel des Mundes, Zungenlage und Lippenrundung) abbilden (vgl. Sendlmeier/Seebode 2007). In der Regel werden für deren Messung die Frequenzen in der Vokalmitte verwendet. Auf diesem Diagramm sind die durchschnittlichen Frequenzen der ersten beiden Formanten abgebildet, welche als wesentliche Kenngrößen für die Vokalqualität gelten (vgl. Reetz/Jongman 2009: 184f). Etwa am Beispiel des /i:/-/e:/-Vokalkontrasts, der uns für diese Arbeit besonders interessiert, kann man sehen, wie sich die Vokalbereiche teilweise überlappen. Erstaunlicherweise gelingt es Hörer/innen aber trotz individueller Sprecherunterschiede (bedingt durch Geschlecht, Alter, Dialekt etc.) und situationsbedingten Unterschieden in der Realisierung, erstsprachliche Laute zu identifizieren (vgl. Reetz/Jongman 2009: 252f, Kerschhofer-Puhalo 2014: 60; zu Modellen, die diesen Sachverhalt zu erklären versuchen, siehe Kap. 2.4).

Das zweite Charakteristikum Nicht-Linearität des Sprachsignals verweist auf die Kontinuität von Sprache (vgl. Höhle 2010: 45, Reetz/Jongman 2009: 251). Damit ist gemeint, dass die Phoneme einer Äußerung nicht bestimmten Abschnitten im akustischen Signal zugeordnet werden können. Es besteht also kein 1:1-Verhältnis. Stattdessen vermengen sich die akustischen Eigenschaften einzelner Laute. Phoneme sind nur abstrakte, (vereinfacht) gedachte Kategorien zur Einordnung von Lauten. „[S]peech is not a sequence of isolated sound segments but rather a continuous flow of larger sound structures that are eventually perceived by listeners“ (Reetz/Jongman 2009: 7). So wird beispielsweise die Opposition zwischen verschiedenen Verschlusslauten (z.B. /b/ vs. /d/) durch die Formantenübergänge zum folgenden Vokal bestimmt (vgl. Höhle 2010: 45). Darüber hinaus werden in mündlicher Sprache zwischen Wörtern faktisch keine Pausen gemacht, was dem/der L2-Hörer/in die Wahrnehmung zusätzlich erschwert (vgl. Reetz/Jongman 2009: 3). Nach den akustischen Merkmalen des sprachlichen Signals soll im nächsten Kapitel die Bedeutung der Erstsprache für die Wahrnehmung angesprochen werden.

2.2 Das native Phoneminventar

Der Einfluss der Erstsprache auf den Erwerb weiterer Sprachen ist unbestritten. Abhängig von der L1 werden Laute unterschiedlich wahrgenommen. Die Zuordnung fremdsprachlicher Laute wird durch das sogenannte native (oder erstsprachliche) Phoneminventar entscheidend beeinflusst. Die Phoneme als kleinste bedeutungsunterscheidende Elemente der Sprache (vgl.

Ernst 2004: 91) stellen die abstrakten Kategorien dar, welche die gehörten sprachlichen Laute in ein sprachliches Lautsystem, das Phoneminventar, gliedern. Dieses ist von Sprache zu Sprache unterschiedlich und beispielsweise durch sprachliche Erfahrung auch veränderbar (siehe Kap. 3.3). Die phonetischen Kategorien der Erstsprache wirken wie ein Sieb, das vertraut klingende Laute aus der Menge des Gehörten herausfiltert (vgl. Barkowski et al. 2014: 142, Rausch R./Rausch I. 2002: 47f). Fremdsprachliche Laute, die den Phonen der L1 ähneln, stellen daher oft eine größere Schwierigkeit dar, da sie unter Umständen gar nicht als verschieden erkannt werden (vgl. Rausch R./Rausch I. 2002: 90, Flege et al. 1994: 3623).

Zahlreiche Studien haben sich mit dem Einfluss der Erstsprache auf die Wahrnehmung auseinandergesetzt. So werden beispielsweise Schwierigkeiten in der Unterscheidung des englischen /r/-/l/-Kontrasts durch Japaner/innen dadurch erklärt, dass beide Laute im Japanischen in eine einzige phonetische Kategorie fallen. Durch phonetische Übungen kann die Diskrimination aber mitunter entscheidend verbessert werden (vgl. Best et al. 2001: 775).

Wann bildet sich das native Phoneminventar heraus? Das weitverbreitete Konzept des *Perceptual Narrowing* (auch *Early Tuning* genannt) geht davon aus, dass sich bei Kleinkindern etwa mit Vollendung des ersten Lebensjahrs die Wahrnehmung auf die Phoneme der Erstsprache spezialisiert (vgl. Mazuka et al. 2013: 192, Best/McRoberts 2003: 184, Polka/Werker 1994: 421, Barkowski et al. 2014: 141).¹ Werden nach diesem Zeitpunkt neue Sprachen bzw. Laute gelernt, wird auf das native Phoneminventar zurückgegriffen. Die Schwelle scheint bei 6 Monaten oder früher für Vokale und bei 10 bis 12 Monaten für Konsonanten zu liegen (vgl. Polka/Werker 1994: 423, 427). Zuvor sind Babys universale „Sprachversther“, sprich sie können alle Laute aller Sprachen unterscheiden – so zumindest die Annahme. Dem Konzept zufolge sollte die Diskrimination von nichtnativen Kontrasten nach dem ersten Lebensjahr deutlich schlechter werden – was allerdings nicht immer und für alle Kontrastpaare der Fall ist. Die Diskrimination nichtmuttersprachlicher Kontraste im Erwachsenenalter ist nicht automatisch schlecht wie das *Perceptual Narrowing* suggeriert. So konnte beispielsweise ein dentaler versus lateraler Konsonantenkontrast in der Sprache isiZulu, der im Englischen nicht vorhanden ist, von englischsprechenden Erwachsenen und 14 Monate alten Kleinkindern problemlos unterschieden werden (vgl. Best/McRoberts 2003: 184f).

¹ Die Einstimmung auf die Erstsprache erfolgt allerdings bereits ab der Geburt. So können Babys beispielsweise bereits kurz nach der Geburt die Prosodie der Erstsprache von der Satzmelodie einer anderen Sprache unterscheiden und sie zeigen schon wenige Stunden nach der Entbindung eine Präferenz für die Stimme der Mutter (vgl. Polka/Werker 1994: 422).

Aber auch der frühe Kontakt mit der L2 (< 1 Jahr) ist nicht zwingend für eine gute oder schlechte Diskrimination entscheidend (vgl. Best et al. 2001: 776). Englisch- und Französischlernende Kleinkinder von 6-8 Monaten und 10-12 Monaten konnten etwa den interdentalen Frikativ nicht vom alveolaren Plosiv unterscheiden [ð] – [d], welche im Englischen kontrastiv sind, im Französischen aber nicht (vgl. Best/McRoberts 2003: 185). Folglich scheint sich in diesem Fall die Diskrimination bei Englisch-Sprecher/innen erst irgendwann nach dem ersten Lebensjahr zu verbessern. Die Bedeutung des Faktors Spracherfahrung für zahlreiche Kontraste soll im folgenden Kapitel noch angesprochen werden (siehe Kap. 3.3).

Diese widersprüchlichen Funde möchte das *Perceptual Assimilation Model* (in der Folge kurz als PAM bezeichnet) von Cathy Best erklären (vgl. Best et al. 2001, Best/McRoberts 2003). Das Konzept geht davon aus, dass fremdsprachliche Lautkontraste abhängig vom Phoneminventar der L1 unterschiedlichen oder gleichen phonetischen Kategorien zugeordnet und je nachdem besser oder schlechter unterschieden werden. Dabei spielt allerdings nicht nur das erstsprachliche Phoneminventar, sondern auch das Verhältnis des Kontrastpaares zueinander eine Rolle. Das *Speech Learning Model* (SLM) von James Flege geht prinzipiell in eine ähnliche Richtung, betrachtet aber vorrangig die Produktion von L2-Lauten durch erfahrenere Lernende. Die Ähnlichkeit des neu zu lernenden Lautes in der L2 im Vergleich zu Lauten der L1 spielt dabei eine gewichtige Rolle (vgl. Bohn 1998: 4f). Vorhersagen beider Modelle sollen in einem späteren Kapitel noch herausgearbeitet werden (siehe Kap. 5).

Beide Konzeptionen können aber nicht alle Ergebnisse von Diskriminationsexperimenten zufriedenstellend erklären. Als stellvertretendes Beispiel soll hier die Studie von Mazuka et al. (2013) angeführt werden. Darin wurde die Diskrimination von drei deutschen Vokalkontrasten (/bi:k/-/be:k/, /bu:k/-/bo:k/ und bu:k/-/by:k/) durch japanische Kleinkinder (mit 4,5 und 10 Monaten) und Erwachsene untersucht.² Ein Ergebnis der Studie war, dass Erwachsene im Gegensatz zu den Kleinkindern alle Kontraste problemlos unterscheiden konnten, was mit den Ergebnissen meines ersten Experiments, welches ich vor einigen Jahren durchgeführt habe, übereinstimmt, den Vorhersagen des PAM aber zuwiderläuft. Im folgenden Kapitel sollen die verschiedenen möglichen Wege, welche die Wahrnehmung von L2-Kontrasten im Kindes- und Erwachsenenalter nehmen kann, näher beschrieben werden.

² Dies ist zudem die einzige Studie, die ebenfalls die Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts durch japanische Hörer/innen experimentell untersucht hat.

2.3 Entwicklungspfade für die Unterscheidung fremdsprachlicher Kontraste

Mazuka et al. (2013: 193) führen vier mögliche Entwicklungspfade hinsichtlich der Diskrimination nicht-muttersprachlicher Kontraste an (siehe Abb.2): 3. *decline*, 4. *experience-independent maintenance*, 5. *non-discrimination* und 6. *improvement without exposure*.³

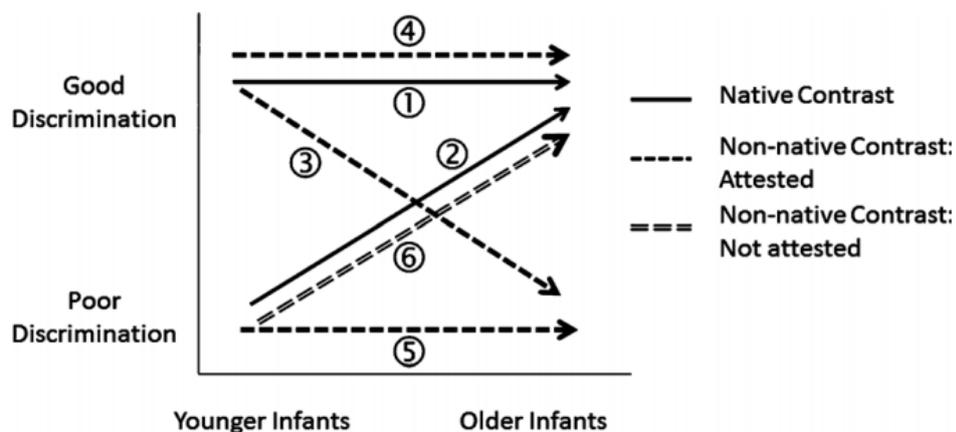


Abbildung 2: Entwicklungspfade für die Unterscheidung von Kontrasten (aus Mazuka et al. 2013: 193)

Der Pfad *decline* verläuft gemäß dem Perceptual Narrowing: Kontraste können als Babys unterschieden werden, später aber nicht mehr. Wie bereits erwähnt, verschlechtert sich die Diskrimination eines fremdsprachlichen Kontrasts aber nicht zwingend nach dem ersten Lebensjahr. Die Unterscheidung eines fremdsprachlichen Kontrasts kann sich mit, aber auch ohne Kontakt mit der Zielsprache verbessern (*improvement without exposure*)⁴. So konnten beispielsweise in einer Studie von Polka und Werker englischsprachige Erwachsene zwei deutsche Vokalkontraste (/dYt/ – /dUt/ und /dyt/ – /dut/) weit besser unterscheiden als 10-12 Monate alte Kleinkinder (vgl. 1994: 427). Oder wie im Fall der bereits angesprochenen Zulu-Klickkontraste ist es auch möglich, dass die Kontraste, die von Babys diskriminiert werden konnten, auch später noch von Erwachsenen problemlos unterschieden werden können (*experience-independent maintenance*) (vgl. Best et al. 2001: 776). Natürlich besteht auch die

³ Für muttersprachliche Kontraste (für diese Arbeit nicht relevant) kann die Diskrimination entweder gleichbleiben (1. *experience-contingent maintenance*) oder sich verbessern (2. *enhancement*).

⁴ Für diesen Fall gibt es zu diesem Zeitpunkt allerdings noch kaum experimentelle Untersuchungen (vgl. Mazuka et al. 2013: 194).

Möglichkeit, dass die Kontraste weder davor noch danach unterschieden werden können (*non-discrimination*).

Kommen wir zu dem bereits zuvor angesprochenen Experiment zurück: Interessanterweise konnten in der Studie von Mazuka et al. (2013: 192) 4,5 Monate alte japanische Babys den /bi:k/-/be:k/-Kontrast im Experiment nicht unterscheiden, im Gegensatz zu 10 Monate alten japanischen Kleinkindern und Erwachsenen, die damit keine Schwierigkeiten hatten. Es handelt sich hierbei also um *improvement without exposure* (Verbesserung der Diskrimination ohne Kontakt mit der Zielsprache), was sowohl dem PAM als auch dem Perceptual Narrowing widerspricht. Über die Ursache für dieses Ergebnis konnten die Autoren nur spekulieren. Sie äußern die Vermutung, dass entweder eine Vielzahl von Faktoren dafür verantwortlich oder die japanischen Kleinkinder eventuell über die Eltern mit Vokalen in der Erstsprache in „Berührung“ gekommen sind, die dem deutschen /i:/ und /e:/ ähneln (vgl. Mazuka et al. 2013: 206):

We speculate that it may be necessary to consider a combination of factors simultaneously; articulatory complexity of the contrast, the developmental changes in infants' ability to pay attention to finer acoustic/articulatory differences and to the combination of articulatory gestures, in addition to learning specific properties of Japanese native vowel categories (Mazuka et al. 2013: 205).

In Kapitel 3 sollen deshalb einige Faktoren herausgegriffen werden, die für die Wahrnehmung fremdsprachlicher Laute im Allgemeinen und Vokalen im Speziellen zentral sind. Zuvor soll aber noch auf die Beziehung zwischen Produktion und Perzeption, sowie dem Unterschied zwischen kategorialer und kontinuierlicher Wahrnehmung eingegangen werden.

2.4 Zusammenhang zwischen Produktion und Perzeption

Sprachwahrnehmung und Artikulation sind zwei Seiten einer Medaille. Über das genaue Verhältnis zwischen Produktion und Perzeption fremdsprachlicher Laute ist man sich aber bis heute uneins (vgl. Bohn: 1998: 1, Mokari & Werner 2017). Es wird allerdings vielfach davon ausgegangen, dass nur das, was gehört und wahrgenommen wird, in der Folge auch selbst ausgesprochen werden kann, weshalb einer Ausspracheschulung gezieltes Hörtraining vorausgehen sollte (vgl. Barkowski et al. 2014: 142). Die L1-spezifische Wahrnehmung hat selbstverständlich auch Einfluss auf die Artikulation fremdsprachlicher Laute (vgl. Rausch R./Rausch I. 2002: 88, Flege et al. 1994: 3624, Kerschhofer-Puhalo 2014: 17).

Diverse Modelle versuchen das Verhältnis zwischen Perzeption und Produktion von Sprache und das Problem der Invarianz bei der Sprachwahrnehmung zu klären. Die sogenannte (überarbeitete) *Motortheorie* von Liberman et al. geht etwa davon aus, dass für die Wahrnehmung von Lauten sog. *artikulatorische Gesten* eine zentrale Rolle spielen (vgl. Höhle 2009: 50). Darunter werden invariante „Artikulationsprogramme“ verstanden, welche die motorische Seite der Produktion von Lauten im Vokaltrakt regeln und welche auch entscheidend für die Perzeption sind. Die angesprochene Invarianz des Sprachsignals ergibt sich dem Modell zufolge erst aus der tatsächlichen Ausführung (= Performanz) (vgl. Höhle 2009: 50). Der/die Hörer/in versucht demnach im Wahrnehmungsprozess die intendierten Artikulationsmuster aus der konkreten Realisation eines Lautsegments zu ermitteln und daraus auf den entsprechenden Laut zu schließen.

Auch die *Theorie des direkten Realismus*, ein allgemeines Wahrnehmungsmodell, geht nach Fowler von einem engen Zusammenhang zwischen Wahrnehmung und Produktion aus (vgl. Höhle 2009: 50f). Hier wird allerdings die tatsächliche Artikulation und nicht die (vorgeschaltete) beabsichtigte Artikulationsgeste ins Blickfeld genommen. Das Ohr nimmt demzufolge nicht physikalische Änderungen des sprachlichen Signals, sondern artikulatorische Gesten wahr (vgl. Höhle 2009: 50f). Dieser Theorie widerspricht allerdings die Tatsache, dass es mehrere Wege gibt, um ein bestimmtes akustisches Signal zu artikulieren (vgl. Reetz/Jongman 2009: 264). Außerdem erklären beide der angeführten Modelle nicht, wie die artikulatorischen Gesten konkret aus dem Sprachsignal heraus wahrgenommen werden können.

Weitere Modelle gehen von einer größeren Bedeutung der akustischen Information für die Identifikation von Lauten aus. So sind beispielsweise nach der *Theorie der Merkmalsdetektoren* von Stevens/Blumstein invariante akustische Merkmale im sprachlichen Signal vorhanden, welche die Bestimmung des Artikulationsortes ermöglichen (vgl. Höhle 2009: 51). Diese fanden sich bei der Analyse spektraler Information in sehr kleinen Zeitfenstern von 10-25 Millisekunden (vgl. Höhle 2009: 51).

Exemplar-basierte Modelle nach Goldinger wiederum gehen davon aus, dass „eine perzeptuelle Kategorie durch alle bislang gehörten Exemplare dieser Kategorie mit all ihren akustischen Eigenschaften repräsentiert wird“ (Höhle 2009: 51). Jedes dieser Exemplare beinhaltet detaillierte akustische Information, aber auch linguistische und indexikalische Merkmale. Normalisierung ist deshalb nach diesem Ansatz nicht erforderlich, da keine Invarianz gegeben ist (vgl. Höhle 2009: 52). Im nächsten Unterkapitel sollen (abhängig vom Lauttyp) verschiedene Wahrnehmungsmodi besprochen und diverse Theorien zur Identifikation von Vokalen erläutert werden.

2.5 Kategoriale vs. kontinuierliche Wahrnehmung

Konsonanten werden kategorial, Vokale eher kontinuierlich wahrgenommen – so lautet der allgemeine Konsens (vgl. Höhle 2010: 47, Reetz/Jongman 2009: 268, Polka/Werker 1994: 422, Polka 1995: 1287, Levy/Strange 2008: 142, Huang 2011, Pisoni 1973, Kerschhofer-Puhalo 2014: 62-64). Was heißt das?

Kategoriale Wahrnehmung bedeutet, dass ein Hörer denselben akustischen Unterschied zwischen zwei Lauten dann besser wahrnehmen kann, wenn die beiden Laute unterschiedlichen phonologischen Kategorien (z.B. /b/ und /p/) zugeordnet werden als wenn sie derselben phonologischen Kategorie angehören (z.B. verschiedene Realisierungen von /b/) (Höhle 2010: 45).

Die Kategoriegrenzen verlaufen hierbei abrupt (vgl. Reetz/Jongman 2009: 266). Kategoriale und kontinuierliche Perzeption bezeichnen zwei unterschiedliche Wahrnehmungsmodi, die bei der Diskrimination von Lauten zum Tragen kommen. Konsonanten – insbesondere Stopkonsonanten – weisen innerhalb einer phonetischen Kategorie nur wenig Variation auf und können daher innerkategorial nicht unterschieden werden (vgl. Huang 2011, Pisoni 1975: 1). Anders verhält es sich mit Vokalen, die auch innerhalb einer phonetischen Kategorie relativ gut unterschieden werden können, was unter anderem mit der großen Variabilität von Vokalen erklärt wird (vgl. Cutler et al 2005: 34, Kerschhofer-Puhalo 2014: 63). Die Kategorienübergänge sind hier deutlich „unschärfer“ und verlaufen ineinander (vgl. Polka/Werker 1994: 422). Auch fremdsprachliche Vokale können oft relativ gut unterschieden werden (vgl. Polka/Werker 1994: 422). Insgesamt haben akustisch-auditive Hinweise offenbar einen stärkeren Einfluss auf die Bestimmung von Vokalen als phonetisch-kategorische Information (vgl. Levy/Strange 2008: 142, Flege et al. 1994: 3623).

Fujisaki und Kawashima (1970) haben für die Wahrnehmung von Sprache ein Modell entwickelt, welches zwei verschiedene Verarbeitungsmodi bzw. Repräsentationen im Kurzzeitgedächtnis annimmt: ein *phonetisches Kurzzeitgedächtnis*, welches für die kategorische Zuordnung phonetisch unterschiedlicher Laute zuständig ist und das vor allem bei Konsonantenkontrasten zum Einsatz kommt und ein *auditorisches Kurzzeitgedächtnis*, welches die innerkategorische Differenzierung bei Vokalkontrasten übernimmt (vgl. Pisoni 1973: 2f, Pisoni 1975: 3f).

Anders als Diskriminationsexperimente suggerieren, sollte die kategoriale Wahrnehmung aber auch nicht zu stark bewertet werden, da Hörer/innen trotz allem Unterschiede in der Güte von Lauten erkennen können, was darauf hindeutet, dass nicht nur phonemspezifische kategoriale Information wahrgenommen wird (vgl. Höhle 2010: 47).

Wie gelingt es Hörer/innen aber Vokale trotz der großen Varianz in der Realisierung (bedingt durch sprecherspezifische, situationsbedingte Variablen und den lautlichen Kontext) zu erkennen? Eine mögliche Antwort darauf liefern Normalisierungstheorien (vgl. Reetz/Jongman 2009: 254). Unterschieden wird zwischen zwei großen Strömungen: *extrinsischer* und *intrinsischer Normalisierung*.

Extrinsische Normalisierung geht davon aus, dass der umgebende Kontext für die Einschätzung und Bestimmung von wahrgenommenen Vokalen eine zentrale Rolle spielt (vgl. Reetz/Jongman 2009: 254). Ladefoged und Broadbent (1957) konnten in einer Studie nachweisen, dass es einen Unterschied macht, ob Wörter isoliert (in diesem Fall /bVt/) oder in einem Satzkontext präsentiert werden (mit dem vorgeschalteten Satz „Please say what this word is“) (vgl. Reetz/Jongman 2009: 254). Die Wahrnehmung des Zielvokals änderte sich abhängig vom umgebenden Satzkontext bzw. von vorhergehenden Vokalen, was als mögliches Indiz für extrinsische Normalisierung gesehen werden kann. Vokale in Isolation können nur schwer identifiziert werden (vgl. Cutler et al 2005: 34). Gegen extrinsische Normalisierung spricht allerdings, dass Vokale auch dann akkurat bestimmt werden, wenn verschiedene Sprecherstimmen zufällig vermischt werden (vgl. Reetz/Jongman 2009: 254).

Intrinsische Normalisierung bezeichnet als Gegenposition dazu die Annahme, dass jegliche Information, die für die Vokalidentifikation erforderlich ist, bereits im Vokal selbst vorhanden ist. Überlappungen zwischen den Vokalen sollen nach dieser Ansicht faktisch nicht existieren oder zumindest signifikant reduziert werden, wenn die akustischen Eigenschaften auf eine andere Art und Weise (beispielsweise mit anderer Skalierung, also nicht linear oder durch Zuhilfenahme von anderen Parametern wie die Grundfrequenz F0) gemessen bzw. dargestellt werden (vgl. Reetz/Jongman 2009: 255).⁵

Möglicherweise sind aber auch zusätzlich zu den Formanten weitere akustische Hinweise für die Identifikation von Vokalen relevant. Zu bedenken ist in diesem Zusammenhang, dass diese Konzeptionen auf Formantenfrequenzen beruhen, die in der Mitte des Vokals gemessen wurden. Mögliche relevante Effekte und Informationen werden dadurch möglicherweise nicht oder zu wenig berücksichtigt, da sich Hörer/in bei der Wahrnehmung voraussichtlich nicht nur auf den unveränderlichen „steady-state“-Teil des Vokals beziehen (vgl. Reetz/Jongman 2009: 256). Eine Studie von Hillenbrand und Neary (1999) mit zwei synthetisch erzeugten Vokaltypen – mit und ohne Formantenübergänge – scheint diese Ansicht zu bestärken. Die

⁵ Meist werden Vokale als Verhältnis zwischen den ersten beiden Formanten dargestellt (vgl. Formantkarte, Abb. 1).

Identifikation von den synthetischen Vokalen mit Formantenübergänge, welche natürlichen näherkommen, war dabei besser (89%) als die der steady-state-Vokale (74%) (vgl. Reetz/Jongman 2009: 256).

3 Einflussfaktoren für die Wahrnehmung von Vokalen der L2

Um der Antwort auf die Frage nach dem Wie der fremdsprachlichen Lautwahrnehmung etwas näher zu kommen, soll an dieser Stelle auf einzelne Variablen eingegangen werden, die darauf wesentlich Einfluss nehmen können. Auf diese Weise soll die zweite Forschungsfrage – Welche Faktoren/Variablen haben Einfluss auf die fremdsprachliche Wahrnehmung der beiden Vokale? – theoretisch fundiert (und soweit als möglich) beantwortet werden. Alle Einflussfaktoren ausfindig zu machen und erschöpfend darzustellen, wird nicht möglich sein, daher muss eine Auswahl von voraussichtlich essentiellen Variablen getroffen werden. Bohn (vgl. 1998: 1) teilt die Einflussfaktoren für die Wahrnehmung fremdsprachlicher Laute in drei Variablengruppen:

1. Lernenden- und L1-Variablen, die den Lerner/die Lernerin und dessen/deren sprachlichen Hintergrund charakterisieren (u.a. lautliche Eigenschaften der L1, Spracherfahrung mit der L2, Alter, Geschlecht etc.),
2. L2-Variablen, welche das gehörte lautliche Material beschreiben (akustische Eigenschaften der L2-Laute), sowie
3. methodische Variablen, die das Vorgehen in Wahrnehmungsexperimenten darstellen.

Eine ähnliche, etwas spezifischere Kategorisierung findet sich bei Kerschhofer-Puhalo (vgl. 2014: 233). Die Wahrnehmung von Lauten der L2 ist ihr zufolge abhängig von

1. Faktoren, die sich auf das experimentelle Setting beziehen (entspricht Bohns dritter Variablengruppe *methodologische Variablen*),
2. sprachbezogene Faktoren, die sich aus der Beziehung zwischen L1 und L2 und der wahrgenommenen Distanz bzw. Ähnlichkeit ergeben (Cross-linguistic Influence),
3. Faktoren, die sich auf vokalspezifische Charakteristika beziehen (akustisch-phonetische Eigenschaften der Stimuli)
4. individuelle Variablen wie Alter, Sprachverwendung in L1 und L2 und Kontakt mit der L2

In Anlehnung an diese beiden Einteilungen sollen in diesem und späteren Kapiteln folgende Faktoren hervorgehoben werden, die für die Vokalwahrnehmung und diese Untersuchung als besonders relevant angesehen werden:

- A. **Cross-linguistic Influence:** beschreibt den wechselseitigen sprachlichen Einfluss zwischen L1 und L2; damit zusammenhängend die wahrgenommene **Ähnlichkeit**/Distanz zwischen Vokalkontrasten, woraus indirekt auf die „Schwierigkeit“ eines Vokalkontrasts geschlossen werden kann
- B. **vokalspezifische Charakteristika** des sprachlichen Materials
- C. individuelle **Lernendenvariablen**, v.a. **Spracherfahrung** und **Alter**
- D. **methodologische Faktoren**, die durch den Experimentmodus bestimmt sind; besondere Beachtung soll hierbei dem Faktor **Aufmerksamkeit** geschenkt werden.

Eine Besprechung der Einflussfaktorengruppen A und C erfolgt in knapper Form in den einzelnen Unterkapiteln (Kap. 3.1 – 3.4). Auf die artikulatorischen Merkmale des /i:/-/e:/-Vokalkontrasts (Variablengruppe B) wird in Kapitel 4 detailliert eingegangen, da diese eine Grundlage für das PAM darstellen.⁶ Methodologische Einflussfaktoren (D) – abgesehen von Aufmerksamkeit – werden im empirischen Teil angesprochen (siehe Kap. 3.5). Weitere Lernendenvariablen, die unter Umständen ebenfalls eine Rolle spielen können, auf die hier aber nicht umfassend eingegangen werden kann und soll, sind unter anderem affektive Variablen wie Motivation, Einstellung gegenüber der L2, Geschlecht und der umstrittene Faktor „Talent“ (vgl. Bohn 1998: 9f).

3.1 Cross-linguistic Influence

L2-Lernende nehmen Laute der Zweitsprache anders wahr als dies Muttersprachler/innen tun. Wie bereits angesprochen spielt das Phoneminventar der L1 dabei eine gewichtige Rolle. Das native Phoneminventar bildet den mentalen Rahmen für die Wahrnehmung von gesprochener Sprache. Bereits 1931 hat Polivanov auf den Einfluss der L1 auf die Wahrnehmung fremdsprachlicher Laute hingewiesen (vgl. Bohn 1998: 2). Als „language consciousness“ bezeichnet er in seiner Untersuchung L1-spezifische Zuordnungen von fremdsprachlichen Kontrasten, welche als eine der Hauptursachen für Wahrnehmungsschwierigkeiten und

⁶ Für eine Diskussion der akustisch relevanten Merkmale für die Vokalwahrnehmung siehe das Kap. von Raphael (2005) im Handbook of Speech Perception. Darin werden Erkenntnisse und Fragen zu den akustischen „Cues“, die zur Identifikation von Vokalen benötigt werden, diskutiert und die Normalisierung von Vokalen besprochen.

fremdsprachlichen Akzent bestimmt werden kann. So können beispielsweise chinesische Lernende (zumindest zu Beginn des L2-Erwerbs) nicht zwischen russischem stimmhaften /b/ und nicht-aspiriertem /p/ unterscheiden, da aufgrund ihrer Erstsprache beide Laute für sie vollkommen ident klingen (vgl. Bohn 1998: 2f). Jedoch werden mitunter nicht nur Unterschiede wahrgenommen, die für muttersprachliche Lautkontraste von Relevanz sind, was den Einfluss der Erstsprache etwas relativiert (vgl. Best/Tyler 2007: 18).

In den 1950ern wurden von Weinreich und Lado schließlich *Transfer* und *Interferenz* als zentrale Ursachen für Schwierigkeiten im L2-Erwerb bestimmt (vgl. Kerschhofer-Puhalo 2014: 30). Mit dem Begriff *Transfer* wird darauf referiert, dass linguistische Elemente aus der L1 in die L2 übernommen werden. Dieser Transfer kann positiv oder negativ sein. Positiver Transfer trägt zum Erfolg in der Zielsprache bei, während negativer Transfer (auch als Interferenz bezeichnet) zu sprachlichen Fehlern und Normverstößen führt (vgl. Kerschhofer-Puhalo 2014: 30f). Auf Grundlage dieser Unterscheidung wurde die sogenannte *Kontrastivhypothese* (auf Contrastive Analysis Hypothesis, kurz CAH) entwickelt, demzufolge Unterschiede in Sprachen Schwierigkeiten im L2-Erwerb vorhersagen können (vgl. Kerschhofer-Puhalo 2014: 31). Aber nicht nur Unterschiede, sondern auch Ähnlichkeiten zwischen Sprachen können erschwerend wirken. Dies gilt in besonderem Maße für die phonologische Sprachebene (siehe Kap. 3.2).

Die L1 hat dabei nicht nur Auswirkungen auf die Ausbildung der L2 und L3, der Einfluss geht in beide Richtungen (vgl. Flege 2007: 363f). Um der wechselseitigen Einflussnahme zwischen Erst- und Zweitsprachen Rechnung zu tragen, wird heute deshalb oftmals nicht von Transfer, sondern von *Cross-linguistic Influence* (CLI) gesprochen:

[...] cross-linguistic influence (CLI) will be preferred to account not only for direct transfer of linguistic structures from one language to another but also to refer to the wide range of implicit and explicit assumptions and strategies in L2 performance, including mental processes that are invoked when learners are establishing mental representations of vowel categories and relationships of similarity and contrast (Kerschhofer-Puhalo 2014: 34).

Mit dem Begriff wird auch ein größeres Feld umschrieben. Es werden nicht nur der Output, sondern auch die mentalen Prozesse und Repräsentationen von Lautkategorien des Lernenden ins Blickfeld genommen. Sowohl Unterschiede als auch Gemeinsamkeiten spielen dabei eine Rolle. Hier nicht behandelt, aber zumindest erwähnt werden soll zudem die Debatte um mögliche Sprachuniversalien, die unter Umständen Einfluss auf die Diskrimination von Lautkontrasten haben könnten (vgl. Kerschhofer-Puhalo 2014: 40). Nachfolgend soll das im Zusammenhang von CLI und für diese Untersuchung bedeutsame Konzept der Ähnlichkeit thematisiert werden.

3.2 Ähnlichkeit

Ähnlichkeit ist ein entscheidender Faktor, um die relative Schwierigkeit der Diskrimination phonetischer Kontraste zu bestimmen (vgl. Kerschhofer-Puhalo 2014: 35). Wie die Ergebnisse eines Wahrnehmungsexperiments von Flege et al. (1994) nahelegen, sind für die phonetische Ähnlichkeit u.a. die Ähnlichkeit artikulatorischer Gesten, auditorisch-akustische Unterschiede des Lautkontrastpaares und deren kategorialer Status (ob beide einer oder zwei unterschiedlichen Phonemen zugeordnet werden) entscheidend (vgl. Flege et al. 1994: 3623, 3639). Das Gegenstück zur Ähnlichkeit ist die *Distanz*.

Nadja Kerschhofer-Puhalo hat es sich in ihrer Dissertation (2014) zur Aufgabe gemacht, das Konzept der Ähnlichkeit operationalisierbar zu machen und es im Kontext fremdsprachlichen Lernens zu modellieren. Untersucht werden sollten dabei die Langzeitrepräsentationen der Vokalkategorien in der L2 Deutsch (vgl. Kerschhofer-Puhalo 2014: 221). Dazu wurde ein großer Korpus von Konfusionsdaten, d.h. falsch zugeordnete L2-Kategorien aus einem Vokal-Identifikationstest mit 173 Deutschlernenden zehn unterschiedlicher Erstsprachen untersucht und ausgewertet⁷. Aus den Konfusionsdaten wurden Werte *perzipierter Ähnlichkeit* und *Distanz* zwischen deutschen Vokalkategorien für die unterschiedlichen Erstsprachen errechnet. Dabei wurde nicht nur interlinguale (zwischen L1 und L2), sondern auch intralinguale Ähnlichkeit (innerhalb) der L2 betrachtet (vgl. Kerschhofer-Puhalo 2014: 98). Die gewonnenen Erkenntnisse der Untersuchung sollen hier mit Blick auf den /i:/-/e:/-Vokalkontrast zusammengefasst präsentiert werden. Zuvor müssen aber einige Begriffe bestimmt werden.

Perzipierte Ähnlichkeit wird von der Autorin folgendermaßen definiert:

From an auditory perspective, perceptual similarity could [...] be defined as lack of perceptual distinctiveness of L2 contrasts due to the learners' experience with L1 and L2 (Kerschhofer-Puhalo 2014: 111).

Entscheidend für die perzipierte Ähnlichkeit sind also im Gegensatz zur phonetischen Ähnlichkeit nicht nur akustisch-phonetische Merkmale des sprachlichen Materials, sondern auch die Spracherfahrung in L1 und L2, die wiederum mit der zuvor angesprochenen Variable CLI zusammenhängt. Der Faktor Spracherfahrung wird im nächsten Unterkapitel behandelt.

⁷ Albanisch, Arabisch, Englisch, Farsi, Ungarisch, Mandarin, Polnisch, Rumänisch, Serbokroatisch (BKS) und Türkisch

Vokalkontraste zeichnen sich weiters durch Asymmetrie aus, d.h. manche Vokalpaare sind sich ähnlicher als andere (vgl. Kerschhofer-Puhalo 2014: 115). Dementsprechend sind auch nicht alle nichtmuttersprachlichen Kontraste gleich schwierig zu identifizieren oder zu unterscheiden (vgl. Best/Tyler 2007: 16f). Die perzeptuelle Salienz gilt dafür als weiterer entscheidender Faktor, der allerdings nicht einfach zu bestimmen ist. Sie hängt ebenfalls mit der Spracherfahrung und mit den akustisch-phonetischen Merkmalen zusammen, die für die Identifikation und Unterscheidung von Vokalen erforderlich sind (z.B. Länge des Vokals, Tonhöhe, Formantenverläufe etc.) (vgl. Kerschhofer-Puhalo 2014: 111): “[S]aliency [...] usually refers to characteristics of the stimulus materials in interaction with listeners’ maturational level and linguistic experience” (Strange 2011: 457). Die akustisch-phonetischen Details allein können nicht die mentalen Repräsentationen von Lautkategorien im Langzeitgedächtnis vorhersagen. Es besteht also kein direkter Zusammenhang zwischen

acoustic cues in the input and perceptual response (output) [...]. The relationships of signal-inherent properties, mental representations of categories and perceptual responses depend on multiple factors that vary language-specifically (Kerschhofer-Puhalo 2014: 611).

Mit Salienz wird vereinfacht gesagt eine bestimmte sprachspezifische perzeptuelle Gewichtung (= Auffälligkeit) dieser phonemischen Hinweise (Cues) beschrieben. Dabei gilt: Je niedriger die perzeptuelle Salienz, desto schwieriger sind ähnliche Vokale voneinander zu unterscheiden:

The discrimination of similar vowels will be more difficult due to the larger number of shared properties and/or due to the lower perceptual saliency of those cues that differentiate the members of a contrast (Kerschhofer-Puhalo 2014: 119).

Diese Gewichtung ist unter anderem von der Erstsprache abhängig (vgl. Guion/Pederson 2007: 57f). Je nach L1 geben die Hörer/innen mitunter auf unterschiedliche akustisch-phonetische Cues Acht, die es ihnen ermöglicht, die Vokale zu bestimmen oder zu diskriminieren. Beispielsweise verwenden Sprecher/innen von Tonsprachen wie Mandarin die Steigung der Tonhöhe F0 zur Identifikation von Vokalen, während das bei Nicht-Tonsprachen nicht der Fall ist (vgl. Guion/Pederson 2007: 64).

Abbildung 3 zeigt die im Identifikationstest falsch und richtig (fett gedruckt) zugeordneten Kategorisierungen aller L2-Hörer/innen in Prozent. /e:/ wurde beispielsweise durchschnittlich von allen Hörer/innen zu 55 % richtig erkannt und zu 12 % als /i:/ identifiziert.

%	a	ɑ	ɛ	ɛ̃	e	ɪ	i	ɔ	o	ʊ	u	æ	ø	ɤ	y	⟨ei⟩	⟨eu⟩	⟨au⟩	wrong	corr	
a	76	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	76
ɑ	15	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	84
ɛ	0	1	33	9	49	0	1	0	0	0	0	0	0	0	0	5	0	0	67	33	
ɛ̃	1	0	7	74	15	1	0	0	0	0	0	0	0	0	0	1	0	0	26	74	
e	0	0	15	7	55	4	12	0	0	0	0	0	1	0	1	5	0	0	45	55	
ɪ	0	0	1	10	4	68	11	0	0	0	0	0	0	2	1	1	0	0	32	68	
i	0	0	0	1	3	14	80	0	0	0	0	0	0	0	1	1	0	0	20	80	
ɔ	0	0	0	0	0	0	0	74	20	1	0	2	1	0	0	0	1	0	26	74	
o	0	0	0	0	0	0	0	8	61	4	12	2	6	1	2	0	1	1	39	61	
ʊ	0	0	0	0	0	0	0	9	3	66	13	2	1	5	1	0	0	0	34	66	
u	0	0	0	0	0	0	0	0	2	15	71	1	1	4	6	0	0	0	29	71	
æ	0	0	2	5	3	0	0	2	1	2	1	59	14	6	2	0	1	0	41	59	
ø	0	0	0	0	0	0	0	0	1	2	5	7	45	9	28	0	1	0	55	45	
ɤ	0	0	0	0	0	1	0	0	0	8	3	9	4	58	15	0	1	0	42	58	
y	0	0	0	0	0	0	2	0	0	2	3	1	3	16	71	0	1	0	29	71	
Σ	6	7	4	7	9	6	7	6	6	7	7	6	5	7	9	.9	.5	.1	35	64	

Abbildung 3: Konfusionsmatrix; richtig und falsch zugeordnete Identifikationen in Prozent (aus Kerschhofer-Puhalo 2014: 242).

Wird ein Vokal der L2 nicht als verschieden zu einem Selbstlaut der Erstsprache erkannt, kann es zu Substitutionen kommen. Bei der Wahl des Substituts und der wahrgenommenen Schwierigkeit spielt neben der Salienz voraussichtlich auch die Häufigkeit des Vorkommens der jeweiligen Lautkategorie eine Rolle:

Substitutions are phonological processes by which one sound is substituted with another similar sound. In cases, where more than one category come into consideration as substitute, the “more salient”, more frequently occurring and “less marked” sound category [...] is expected to be selected as perceptual substitute [...] (Kerschhofer-Puhalo 2014: 159).

Ein Blick auf die Konfusionsdaten zeigt, dass /e:/ von allen nichterstsprachlichen Hörer/innen im Schnitt vergleichsweise oft mit /i:/ substituiert worden ist (vgl. Kerschhofer-Puhalo 2014: 592). Eine Beobachtung, die ich vielfach auch bei japanischen L1-Sprecher/innen in der alltäglichen Kommunikation gemacht habe. Die Substitution von /e:/ nach /i:/ tritt dabei signifikant häufiger auf als eine Ersetzung in die andere Richtung (12% vs. 3%, vgl. Kerschhofer-Puhalo 2014: 689). Die Schwierigkeit des Vokalkontrasts kann möglicherweise auch auf spektrale Ähnlichkeit zurückgeführt werden kann:

[...] some contrasts, e.g. [...] /i:/-/e:/ [...] show considerable spectral similarity and may therefore be difficult in cross-language perception even if they are not considered to be „similar“ to each other under phonological perspective.

/i:/ gilt gegenüber /e:/ als robustere Kategorie, da die Lautklasse in fast allen Sprachen vorhanden ist (vgl. Kerschhofer-Puhalo 2014: 571, Strange 2011: 458). Die nachfolgende Tabelle (Abb. 4) zeigt die Identifikationen für das deutsche /e:/ für die unterschiedlichen

Erstsprachen. In der Untersuchung wurde /e:/ als /i:/ besonders oft von Hörer/innen mit L1 Polnisch identifiziert (35%):

/e:/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	i	i:	ɔ	o:	ʊ	u:	æ	ø:	ɣ	y:	<ei>	<eu>	<au>	#
Albanian			15	14	68	1	1												206
Arabic	1	3	24	15	44	4	10												176
English	1	2	30	15	43	3	5											1	261
Farsi			13	21	31	6	17					1						13	72
Hungarian			10	1	79	2	7											1	410
Mandarin	2		7	10	50	3	13											14	107
Polish			7	3	24	11	35							1	2	4		11	552
Romanian			18	6	72		3												216
SBC			21	6	58	1	8					1	2			1		1	591
Turkish			10	4	63	2	11											10	418
German			7		93														324
all L2			15	7	55	4	12							1	1	5			3063

Abbildung 4: Identifikationsrate von /e:/ nach Erstsprachen gegliedert (aus Kerschhofer-Puhalo 2014: 575)

Über die Ähnlichkeit kann indirekt auf die Schwierigkeit eines Kontrasts geschlossen werden: Je ähnlicher sich zwei Vokale sind, desto schwieriger sind sie zu identifizieren bzw. diskriminieren. Ein weiterer wichtiger Faktor für die relative Schwierigkeit der Lautwahrnehmung in der L2 ist die Aufmerksamkeit (*attention*), die in Kapitel 3.5 erläutert wird.

3.3 Spracherfahrung in / Sprachkontakt mit der L2

„Perception differs in important ways between naïve listeners and those who have experience with the stimulus contrasts as elements of a second language” (Best/Tyler 2007: 14). Eine Studie von MacKain, Best und Strange (1981) hat gezeigt, dass durch sprachliche Erfahrung in der L2 Wahrnehmungsschwierigkeiten weitgehend überwunden werden können (vgl. Bohn 1998: 3). Untersucht wurde der englische /r/-/l/-Kontrast bei japanischen Lernenden. Je erfahrener diese waren, desto besser konnten sie die vorgegebenen Stimuli unterscheiden und zur richtigen phonemischen Kategorie zuordnen.

Dass die Spracherfahrung aber nicht zwangsläufig zu einer veränderten Wahrnehmung fremdsprachlicher Laute beiträgt, zeigt eine Untersuchung von Bohn und Flege (1990) (vgl. Bohn 1998: 4). Auch nach mehrjähriger Spracherfahrung in der L2 haben deutschsprachige Lernende den englischen /i/-/I/-Kontrast auf dieselbe Weise differenziert wie deutsche Lernende mit geringer Spracherfahrung im Englischen. Der englische /ɛ/-/æ/-Vokalkontrast

wurde hingegen von deutschen Lernenden, die weniger als ein Jahr in den USA verbracht haben nach anderen Kriterien differenziert, als von jenen Hörer/innen, die mehr als fünf Jahre in den USA gelebt haben, welche den Kontrast ebenso wie L1-Englischsprachige aufgrund von spektralen Merkmalen unterschieden (vgl. Bohn 1998: 4). Bei diesem Vokalkontrast hatte also wiederum die Spracherfahrung in der L2 Einfluss auf die fremdsprachliche Wahrnehmung.

Die angesprochenen Unterschiede in der Wahrnehmung versucht Fleges *Speech Learning Model* (SLM) zu erklären (vgl. Bohn 1998: 4f, siehe Kap. 5.2). Dem Modell zufolge spielt die Ähnlichkeit der L2-Laute zu jenen der L1 eine große Rolle. Wenn sich neue Laute der L2 hinreichend stark von jenen der L1 unterscheiden, wird eine neue Kategorie gebildet (wie beim /ɛ/-/æ/-Kontrast), ähneln sich diese aber zu sehr, ist dies nicht der Fall (siehe /i/-/I/-Kontrast). Flege spricht in diesem Zusammenhang von der Äquivalenzklassifikation. Die L2-Laute werden dabei als mehr oder weniger gute Vertreter einer Kategorie der L1 wahrgenommen.

Fleges Modell zeigt, dass nicht in allen Fällen fremdsprachliche Wahrnehmungsschwierigkeiten durch Spracherfahrung überwunden werden können. Allerdings merken viele Forscher/innen an, dass Spracherfahrung nicht quantifiziert werden kann (vgl. Flege 2007: 363, Bohn 1998: 5, Kerschhofer-Puhalo 2014: 276f). Ab welcher Dauer hat man genügend Spracherfahrung gesammelt? Spielen die Dauer des Aufenthalts im Zielsprachenland (*length of residence*, LOR), das Alter ab dem Deutsch gelernt wird (*age of beginning*, AOL) oder gezieltes phonetisches Training eine Rolle? Bedeutsamer als LOR oder AOL scheint die Menge und Qualität des sprachlichen Inputs in der L2 zu sein (vgl. Kerschhofer-Puhalo 2014: 277), wobei dabei natürlich auch bedacht werden muss, dass Input nicht gleich Intake ist. Wie kann die sprachliche Kompetenz (*level of proficiency*) in der L2 als Teilmenge und wesentliche Einflussgröße am besten gemessen werden? In dieser Arbeit wird von zwei Faktoren ausgegangen, um die Spracherfahrung indirekt zu bestimmen. Einerseits wird die bisherige Dauer des L2-Erwerbs und der individuelle Sprachhintergrund ermittelt, andererseits sind die Student/innen in unterschiedliche Niveaustufen eingestuft, wodurch zumindest eine grobe Einteilung der Sprachkompetenz in der L2 möglich gemacht wird.

Eine andere Form der Spracherfahrung findet sich in der Studie von Tees und Werker (1984). Darin wurde gezeigt, dass Personen mit L1 Englisch, die in ihrer frühen Kindheit (< 1 Jahr) mit Hindi in Kontakt gekommen sind, kontrastierende Laute in dieser Sprache auch noch im Erwachsenenalter unterscheiden konnten, obwohl sie ab dem zweiten Lebensjahr einsprachig englisch aufgewachsen sind (vgl. Bohn 1998: 7). Für manche Kontraste kann also auch bereits der kurzfristige Kontakt mit der Zielsprache – in der Kindheit oder auch im Erwachsenenalter – die Diskrimination nichtmuttersprachlicher Kontraste positiv beeinflussen (vgl. Best et al.

2001: 776). Der nur schwer bestimmbare Faktor Sprachkontakt bzw. Spracherfahrung spielt mitunter eine nicht unwesentliche Rolle, um fremdsprachliche Laute überhaupt erst als „anders“ identifizieren zu können.

Dass auch nicht-spezifische Spracherfahrung für die Diskriminierung von fremdsprachlichen Lauten von Bedeutung sein kann, legt das Ergebnis der Untersuchung von Polka (1992) nahe. Hörer/innen mit Farsi konnten einen Kontrast unterscheiden, der nicht in ihrer Erstsprache vorkommt, indem sie ihre „Erfahrung“ über die Artikulationsorte nutzen konnten (velar vs. uvular, vgl. Bohn 1998: 6). Damit wird der Faktor sprachliche Erfahrung mit der L2 etwas relativiert. Zusammenfassend kann gesagt werden, dass die Spracherfahrung in der L2 mit Einschränkungen (abhängig vom jeweiligen L2-Kontrast und der L1) Einfluss auf die Wahrnehmung von fremdsprachlichen Lauten hat.

3.4 Alter

Mit Lennebergs *Critical Period Hypothesis* wurde die Diskussion angefacht, ob man nur als Kind (bei Lenneberg zwischen ca. 2 und 12 Jahren) eine Sprache vollständig erwerben kann (vgl. Bohn 1998: 7, Flege 2007: 357). Bohn meint, dass es beinahe „unstrittig [ist], dass Kinder Sprachen erfolgreich erwerben, während Erwachsene meist nicht erfolgreiche Lerner sind“ (Bohn 1998: 7). Wie bereits angesprochen, können Kleinkinder nach dem bis heute einflussreichen Konzept des Perceptual Narrowing bis zum Alter von ca. acht Monaten alle Konsonantenkontraste aller Sprachen (nicht nur der Erstsprache) unterscheiden. Danach erfolgt eine Spezialisierung auf die Laute der L1. Allerdings haben die Wahrnehmungsverluste keine biologischen Ursachen wie eine Untersuchung von Werker und Logan (1985) nachgewiesen hat (vgl. Bohn 1998: 7). Erwachsene sind demnach nur durch die L1 geprägt, weshalb sie nicht auf akustische Detailinformationen achten, die für die Unterscheidung fremdsprachlicher Kontraste relevant wäre (vgl. Bohn 1998: 7f). Wenn die Erwachsenen allerdings fremdsprachliche Stimuli in einem auditiven und nicht sprachspezifisch-phonetischen Modus wahrgenommen hatten, konnten sie die Kontraste unterscheiden (siehe dazu auch Kap. 3.5). Flege (vgl. 2007: 357) hat einen starken Einfluss des AOL (dort *AOA*, *age of arrival* genannt) auf Akzent und Produktion bei koreanischen L1-Sprecher/innen in der L2 Englisch festgestellt, die durchschnittlich 15 Jahre (von 7 bis 30 Jahren) in den Vereinigten Staaten gelebt haben. Wie Flege selbst anmerkt, ist dies aber nur ein möglicherweise relevanter Faktor für die Bemessung der sprachlichen Kompetenz in der Zweitsprache. So spielt beispielsweise die Sprachnutzung in der L2 eine wesentliche Rolle (vgl. Flege 2007: 358). Neben der eben

genannten Studie werden immer wieder Ergebnisse präsentiert, die darauf hinweisen, dass Wahrnehmungsprobleme durch Erfahrung mit der L2 überwunden werden können (siehe Kap. 3.3). Das heißt, dass auch für Erwachsene ein vollständiger L2-Spracherwerb möglich ist. Umgekehrt gibt es auch Fälle von Erwachsenen, die in der Kindheit begonnen haben eine Zweitsprache zu lernen und diese mit Akzent sprechen (vgl. Flege 2007: 358).

Anstatt von weniger erfolgreichem Lernen sollte man meines Erachtens besser von anderem Lernen sprechen: Erwachsene lernen Sprachen anders als Babys, können aber dennoch (zumindest teilweise) ersprachliches Niveau in der L2 erreichen (vgl. Bohn 1998: 7f). Die Lernfähigkeit Erwachsener ist definitiv gegeben. Auch in der Studie von Kerschhofer-Puhalo (vgl. 2014: 286) konnte kein signifikanter Zusammenhang zwischen AOL und Performanz festgestellt werden. Das vielfach zitierte „earlier is better“ (vgl. Flege 2007: 353) konnte nicht nachgewiesen werden. Entscheidend schienen stattdessen die Spracherfahrung bzw. Sprachkompetenz und die Quantität und Qualität des Inputs gewesen zu sein – ein Ergebnis, das folgende These von Best und Tyler stützt:

The contribution of age to L2 perceptual learning occurs primarily through interactions with length of residence (LOR), relative usage of L1/L2 [...] and relative quantity and quality of input from native L2 speakers (Best/Tyler 2007: 15).

3.5 Aufmerksamkeit

In diesem Abschnitt soll das *Automatic Selective Perception* (ASP) Modell von Strange (2011) vorgestellt und dessen Implikationen für diese Untersuchung deutlich gemacht werden. Dazu müssen zuerst einige Begriffe definiert werden. Wahrnehmung wird bei Strange als sinnsuchende Aktivität beschrieben, bei der der Hörer/die Hörerin semantisch zu verstehen versucht, was der/die Sprecher/in mitteilen will (vgl. Strange 2011: 457). Aufmerksamkeit steuert die Filterung des kontinuierlichen Sprachstroms nach Informationen, die für das Verständnis benötigt werden. Im Zusammenhang mit Aufmerksamkeit sind drei Termini von Bedeutung: selektive Wahrnehmung (*selective perception*), Fokus (*attentional focus*) und automatische Prozesse (*automatic processes*) (vgl. Strange 2011: 459). Selektive Wahrnehmung bezieht sich auf die Erkennung und Gewichtung von phonetisch relevanter Information. Der Begriff Fokus meint die Ausrichtung auf eine bestimmte Art von Information abhängig von den Zielen und Absichten des Hörers/der Hörerin. Automatische Prozesse referiert auf perzeptuelle Aktivitäten, die unbewusst erfolgen, keinen Fokus und damit geringe kognitive Ressourcen benötigen.

Das ASP unterscheidet zwischen zwei unterschiedlichen Wahrnehmungsmodi – einem phonologischen und einem phonetischen Modus:

The phonological mode of perception is normally utilized by adult listeners when processing continuous L1 speech utterances produced by speakers with the same or similar dialect as their own. This mode enables the listener to detect sufficient phonologically relevant contrastive information for word-form identification. [...] The phonetic mode of perception involves the attunement to context-dependent allophonic details, i.e., accessing the listener's knowledge of phonetic and phonotactic patterns that is necessary to produce phonetic sequences appropriately in the L1 dialect/language (Strange 2011: 460).

Während der phonologische Modus also bei der Wahrnehmung erstsprachlicher Äußerungen zum Einsatz kommt, wird der phonetische Modus vor allem am Beginn des Zweitspracherwerbs oder bei einem unbekanntem Dialekt aktiviert. Wahrnehmung im phonetischen Modus benötigt Aufmerksamkeit (bei Strange *attentional focus* genannt) und damit kognitive Ressourcen. Dieser Modus kommt zum Beispiel bei Diskriminationsexperimenten (wie in dieser Arbeit) zum Einsatz und sorgt in ruhiger Umgebung mit einfachen Stimuli bei kurzem Interstimulusintervall (ISI, 0-750ms) für relativ gute Unterscheidungsraten – was mit dem phonologischen Modus nicht möglich wäre (vgl. Strange 2011: 460, Mora 2007: 2).

Die erstsprachliche Wahrnehmung verläuft im Gegensatz dazu weitgehend automatisiert. Mittels hochgradig automatisierter perzeptueller Routinen (*selective perception routines*, SPR) werden einzelne Wörter der L1 im phonologischen Modus ohne viel kognitiven Aufwand erkannt und verstanden – auch unter schlechten Hörbedingungen (vgl. Strange 2011: 460). Diese SPRs haben sich durch jahrelange Erfahrung mit der Erstsprache herausgebildet und sorgen für eine schnelle und robuste Verarbeitung.

Hinsichtlich Verarbeitung und Speicherung des Gehörten unterscheidet Strange zwischen 1. sogenannten Traces im Kurzzeitgedächtnis, 2. Repräsentationen lautlicher Kategorien im Langzeitgedächtnis und 3. perzeptuelle Routinen im prozeduralen Gedächtnis (vgl. Strange 2011: 459). Die im Langzeitgedächtnis gespeicherten Repräsentationen lautlicher Kategorien beschreiben das „Wissen“ des Hörers/der Hörerin über die Beziehungen phonetisch-phonologischer Einheiten. Das prozedurale Gedächtnis bezieht sich hingegen auf perzeptuelle Strategien und auf das Wie der Wahrnehmung. Traces sind phonetische „Spuren“, die im Kurzzeitgedächtnis verweilen und ebenfalls für die Diskrimination und Identifikation nutzbar gemacht werden können. Diese phonetischen Informationen sollten daher vor allem bei Diskriminationsexperimenten mit kurzem ISI (wie in dieser Untersuchung) zum Tragen kommen.

4 Kontrastierung des deutschen und japanischen Vokalsystems

In diesem Kapitel wird das deutsche Vokalsystem dem japanischen gegenübergestellt. Dabei sollen insbesondere die e- und i-Vokale beider Sprachen verglichen werden.⁸ Dafür wird auch die Produktion in den Blick genommen, um mögliche Schwierigkeiten für L2-Lernende aufzuzeigen. Unterschiede wie Gemeinsamkeiten in den differierenden Phoneminventaren und innerhalb des /i:/-/e:/-Vokalkontrasts sollen dabei herausgearbeitet werden.

4.1 Artikulatorische Grundlagen

Der wohl augenfälligste Unterschied zwischen der deutschen und der japanischen (Aus-)Sprache zeigt sich in der bereits erwähnten Schwierigkeit, die Japaner/innen mit den Phonemen /r/ und /l/ (vor allem in der Produktion) haben. So wird etwa das deutsche Wort *Blätter* [ˈblɛtɐ] von japanischen Muttersprachler/innen in der Regel genauso wie *Bretter* [ˈbrɛtɐ] ausgesprochen. Die Ursache liegt darin, dass die „r“- und „l“-Laute im Deutschen zwei Phoneme bilden, während sie im Japanischen Phone einer einzigen Kategorie darstellen, also nicht distinktiv sind. Das heißt, es wird in der japanischen Aussprache nicht zwischen [r] bzw. [R] und [l] unterschieden. Umgekehrt ist es mir im Selbststudium schwergefallen, das japanische „r“ zu erlernen, welches gefühlt eine Mischung aus den Lauten [R], [l] und [d] darstellt. Dies nur als ein Beispiel für einen offenkundigen Unterschied zwischen beiden Sprachen, der den Einfluss des nativen Phoneminventars auf das Sprechen und Hören aufzeigt. Aber auch (scheinbar) gleich klingende Laute verschiedener Sprachen weisen zumindest geringfügige qualitative Unterschiede auf (vgl. Rausch R./Rausch I. 2002: 28). Die phonetische IPA-Transkription gaukelt hier mitunter eine Gleichheit vor, die bei näherer Betrachtung gar nicht oder zumindest nur bedingt gegeben ist. So wird ein japanisches „i“ etwas anders produziert als ein deutsches (siehe Kap. 4.4 und 4.5), auch wenn wohl kaum jemand einen Unterschied zwischen beiden Lautprodukten erkennen würde.

⁸ Die folgenden Ausführungen basieren zu weiten Teilen auf der Seminararbeit zum selben Thema, die ich vor etwa zwei Jahren geschrieben habe.

Betrachtet man die Erstsprache für sich, gleicht ebenfalls kein Laut dem anderen. Selbst wenn ein und derselbe erstsprachliche Laut mehrmals hintereinander artikuliert wird, sind die resultierenden Lautrealisationen jedes Mal etwas anders (vgl. Rausch R./Rausch I. 2002: 28).

Im Folgenden soll der Fokus auf Vokale gelegt werden. Peter Ernst definiert diese in seiner „Einführung in die germanistische Sprachwissenschaft“ folgendermaßen:

Vokale sind **Öffnungslaute**. Die Stimmlippen vibrieren und erzeugen einen Stimmtönen. Dieser wird im Mundraum durch die Einstellung der beweglichen Artikulationsorgane (Zunge, Unterkiefer, Lippen) so modifiziert, dass die für die verschiedenen Vokale typischen Frequenzmuster entstehen (Ernst 2004: 74f).

Vokale sind also (in der Regel) stimmhafte Laute, die durch eine Öffnung im Mund- bzw. Rachenraum erzeugt werden. Im Unterschied zu Konsonanten wird bei der Artikulation eines Selbstlauts aber kein Verschluss gebildet, das heißt, der Luftstrom wird nicht durch ein Hindernis manipuliert (vgl. Wiese 2011: 26, Rausch R./Rausch I. 2002: 24). Die Vokalqualität ist abhängig von der „Höhe/Öffnung des Vokals [gesteuert durch Unterkiefer bzw. Zunge, Anmerkung D.E.], die Lage der Zunge auf der horizontalen Achse und die Beteiligung oder Nichtbeteiligung der Lippenrundung [...]“ (Wiese 2011: 38). Dazu kommt im Deutschen als weiteres distinktives Merkmal die Länge des Selbstlauts, die als Vokalquantität bezeichnet wird.

4.2 Das deutsche Vokalsystem

Durch die Unterscheidung in Qualität und Länge ergibt sich eine große Zahl an unterschiedlichen Vokalen im Deutschen: „Die deutsche Aussprache [...] kennt [infolgedessen] 17 verschiedene Vokallaute, das sind weitaus mehr als die meisten anderen Sprachen aufweisen“ (Barkowski et al. 2014: 152).⁹ Abgesehen von den beiden Schwa-Lauten (die immer unbetont sind) und dem langen offenen /ɛ:/¹⁰ bilden die geschlossenen mit den offenen Vokalen Minimalpaare, die sich ausschließlich in der Quantität und Gespanntheit voneinander

⁹ Manche Darstellungen kommen auf eine andere Zahl, da die Schwa-Laute (e-Schwa und a-Schwa bzw. das vokalisierte R) nicht immer zu den Vokalen gezählt werden. Diphthonge sind hier nicht miteingerechnet.

¹⁰ Das /ɛ:/ wird zur Unterscheidung der Graphien <e> und <ä> benutzt, in vielen Gebieten aber heute zugunsten von /e:/ aufgegeben (z.B. bei Gewähr – Gewehr, vgl. Kohler 1995: 172f). Der Umlaut hat damit in der gesprochenen Sprache oft keine lautliche Markierung mehr.

unterscheiden (siehe Tabelle 1; z.B. „Miete – Mitte“, vgl. Rausch R./Rausch I. 2002: 25). Der Silbenakzent liegt immer auf gespannten Vokalen, ungespannte Vokale sind immer unbetont.

	vorn		zentral			hinten
kurz	i	ʏ				ʊ
	ɛ		œ	ə	ɐ	ɔ
			a			
lang	i:	y:				u:
	e:	ɛ:	ø:			o:
			ɑ:			

Tabelle 1: Vokale des Deutschen (nach Dieling 1992: 119)

Die Vokale des Deutschen können – neben der Quantität – in Vorder- und Hinterzungenvokale sowie in verschiedene Öffnungsgrade (geschlossen, mittel und offen) eingeteilt werden und gerundet oder ungerundet sein. So lassen sich etwa die beiden deutschen Vokale /i:/ und /e:/, die uns in dieser Arbeit interessieren, nach Kohler beispielsweise (vgl. 1995: 171f) folgendermaßen binär kodieren (etwas vereinfacht):

/i:/ +*gespannt* +*vorn* +***geschlossen*** -*offen* -*gerundet*

/e:/ +*gespannt* +*vorn* -***geschlossen*** -*offen* -*gerundet*

Man kann an dieser Klassifikation sehr gut erkennen, dass sich beide Vokale aus artikulatorischer Sicht betrachtet lediglich hinsichtlich des Öffnungsgrades voneinander unterscheiden: /i:/ ist ein hoher Vorderzungenvokal (geschlossen, ungerundet und gespannt), /e:/ ein mittelhoher Vorderzungenvokal, der im Vergleich dazu weniger geschlossen ist. Moosmüller (2007) und Kerschhofer-Puhalo (2014) unterscheiden stattdessen zwischen präpalatalem /i:/ und mittelpalatalem /e:/ (vgl. Kerschhofer-Puhalo 2014: 174). Die geschlossenen kurzen Vokale (z.B. /i/), welche in der obigen Tabelle 1 nicht angeführt sind, kommen hauptsächlich in Fremdwörtern und Diphthongen vor (z.B. im Wort „Simulation“, vgl. Rausch R./Rausch I. 2002: 25).

4.3 Das japanische Vokalsystem

Japanisch kommt im Gegensatz zum Deutschen mit nur fünf verschiedenen Vokalen aus (vgl. Kubozono 2015: 38f). Anders als im Deutschen sind die japanischen Vokale weder gespannt noch gerundet, insgesamt ist der Lippenöffnungsgrad relativ gering ausgeprägt¹¹ (vgl. Kaneko 2005: 9, Kaneko/Neyer 1987: 88, Vance 1987: 10). Dementsprechend fungiert die Lippenrundung in dieser Sprache nicht als distinktives Merkmal (vgl. Kaneko/Neyer 1987: 100). „Runde“ Vokale wie etwa das deutsche „ö“ /ø:/ und „ü“ /y:/ existieren im Japanischen nicht und müssen von japanischen Deutschlernenden erst erlernt werden. Laut Dieling bereitet dabei besonders die erforderliche Lippenstülpung Schwierigkeiten, da sie eventuell als unziemlich gilt (vgl. Dieling 1992: 86). Die beiden Schwa-Laute stellen ebenfalls eine Besonderheit der deutschen Sprache dar (vgl. Kaneko/Neyer 1987: 101). Infolgedessen wird der unbetonte e-Schwa [ə] (z.B. in „Gedanke“ [gə'daŋkə]) von Japaner/innen meist durch /ɛ/ substituiert (vgl. Dieling 1992: 86).

Ein typisches Charakteristikum der japanischen Sprache ist, dass Vokale vor Nasalen immer nasaliert werden (vgl. Dieling 1992: 86). Desweiteren werden die hohen Vokale [i] und [u] devokalisiert (= „verschluckt“), wenn sie keinen Akzent tragen, nicht morpheminitial stehen und ihnen kein stimmhaftes Phonem folgt, weshalb sie als phonetisch „schwache“ Vokale bezeichnet werden (vgl. Kaneko/Neyer 1987: 88, Kubozono 2015: 39). So wird beispielsweise der Stadtteil Asakusa [asak'sa] ausgesprochen.

vorn	zentral	hinten
i		u
ɛ		ɔ
	a	

Tabelle 2: Vokale des Japanischen (nach Dieling 1992: 86)¹²

¹¹ Sakuma (1973) zufolge sind die Lippen bei ɔ schwach gerundet (vgl. Vance 1987: 10).

¹² Tabelle wurde um den u-Laut u ergänzt, welcher in der Übersicht von Dieling gefehlt hat. Die japanischen Vokale decken sich nicht völlig mit den deutschen Vokalen, weshalb ich in weiterer Folge zum Beispiel von jap. e und nicht ɛ spreche, um die Laute unterscheiden zu können.

Historisch betrachtet spielt die Vokalquantität in der japanischen Sprache keine entscheidende Rolle (vgl. Kubuzono 2015: 39). Jedoch ist die Vokallänge im Japanischen für die Unterscheidung in vielen Fällen durchaus relevant – ersichtlich etwa am Minimalpaar おばあさん (= „obaasan“; jap. für „Großmutter, alte Frau“) vs. おばさん (= „obasan“; jap. für „Tante, Frau im mittleren Alter“, vgl. Kubozono 2015: 39). Hauptquelle für Wörter mit langen Vokalen sind Lehnwörter aus dem Chinesischen oder Englischen, im „ursprünglichen“ Japanisch sind diese nur selten anzutreffen (vgl. Kubuzono 2015: 39).

Den Autoren Kaneko und Neyer zufolge gibt es im Japanischen aber bis heute keine langen Vokale (und auch keine Diphthonge, vgl. Kaneko/Neyer 1987: 98). Diese Ansicht ist allerdings umstritten. Diskutiert wird, ob es sich bei „Doppelvokalen“ wie in おかあさん (= „okaasan“; jap. für „Mutter“) „um Folgen mehrerer gleicher Vokale handelt [...], oder ob man ein Dehnungsphonem anzusetzen hat, um den qualitativen Unterschied zu erklären“ (Kaneko/Neyer 1987: 88). Kaneko und Neyer plädieren für erstere Möglichkeit und stellen langen deutschen Vokalen gedoppelte japanische gegenüber, die sich allerdings phonetisch voneinander unterscheiden (z.B. / i :/ - / ii /, vgl. Kaneko/Neyer 1987: 98, Vance 1987: 13f).

Vance (vgl. 1987: 13) und andere Autor/innen (z.B. Kubozono 2015: 39) gehen im Gegensatz dazu sehr wohl von langen Vokalen in der japanischen Sprache aus, welche einen Kontrast zu den kurzen Vokalen darstellen. Neueren Erkenntnissen zufolge spielen neben der Vokallänge unter anderem auch die Tonhöhe (engl. „pitch“) für die Perzeption von langen Vokalen eine Rolle (vgl. Kubozono 2015: 40). Lange Vokale im Japanischen unterscheiden sich nach Kaneko nur in quantitativer Hinsicht von ihren „kurzen“ Gegenstücken (vgl. Kaneko 2005: 16f). Kasuya und Sato (1989) zufolge wird der Langvokal im Regelfall zudem nicht doppelt so lange, sondern nur 60% länger als der Kurzvokal ausgesprochen (vgl. Kaneko 2005: 17). Ob es nun im Japanischen lange Selbstlaute gibt oder nicht: Fest steht auf jeden Fall, dass der im Deutschen vorhandene „Gegensatz offen – geschlossen [...] im Japanischen nicht“ existiert (Dieling 1992: 86). In den folgenden Kapiteln sollen das deutsche und das japanische Vokalsystem gegenübergestellt und auf Gemeinsamkeiten und Unterschiede eingegangen werden.

4.4 Kontrastierung beider Vokalsysteme

Da sich die für das deutsche Vokalsystem charakteristische Opposition gespannt-ungespannt zwischen langen und kurzen Vokalen im Japanischen nicht findet, könnten gegebenenfalls Probleme in der Wahrnehmung auftreten. Lange Vokale im Deutschen sind immer gespannt, japanische weder gespannt noch gerundet. Hinsichtlich der Produktion von japanischen Vokalen gibt es meist geringfügige Unterschiede im Vergleich zum äquivalenten deutschen Selbstlaut.¹³ So wird das japanische „i“ etwas höher realisiert als das deutsche, während die Spannung der Lippenmuskeln lockerer ist. Das japanische „e“ liegt etwa auf der Höhe zwischen /e:/ und /ɛ:/, während das deutsche höher zu liegen scheint. Das japanische „a“ entspricht qualitativ dem deutschen. Das japanische „o“ ähnelt dem deutschen /ɔ/, es hat allerdings nur geringe Lippenrundung. Ebenso fehlt beim japanischen „u“ die Lippenrundung. Es wird als /u/ realisiert, während das deutsche „u“ als /ʊ/ bzw. /u/ ausgeführt wird.

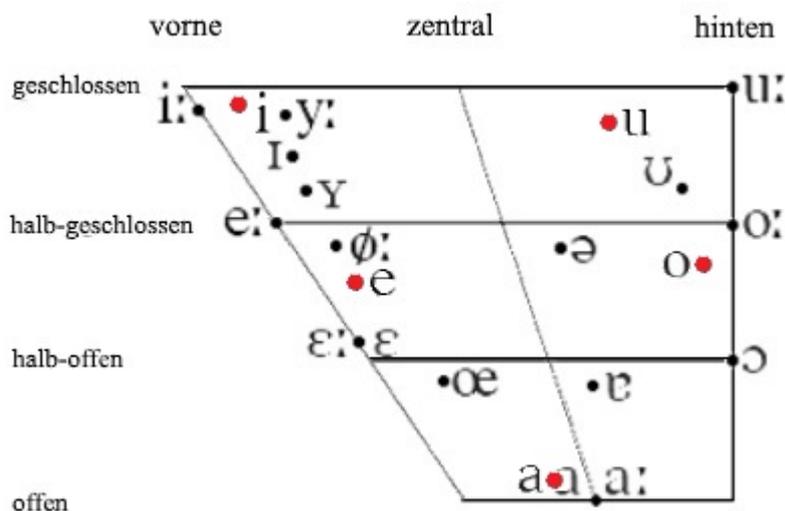


Abbildung 5: Japanisches und deutsches Vokalsystem (deutsche Vokale sind mit schwarzen, japanische mit roten Punkten gekennzeichnet, nach eLaut Japanisch und Deutsch)

Abbildung 5 zeigt eine Gegenüberstellung der Vokalsysteme beider Sprachen. Das japanische und das deutsche Vokaltrapez wurden dazu übereinandergelegt (Vokaltrapeze nach eLAUT Japanisch und Deutsch). Die roten Punkte kennzeichnen die japanischen Vokale, die schwarzen die deutschen Vokale. Aufgrund des wesentlich komplexeren und umfangreicheren Vokalinventars verwundert es nicht, wenn Helga Dieling (1992: 86) zu dem Schluss kommt,

¹³ Vgl. Kaneko/Neyer 1987: 100f für die folgenden Ausführungen.

dass „das vergleichsweise komplizierte Vokalinventar des Deutschen für Japanischsprechende ein schwieriges Pensum darstellt“. Es ist generell ein schwieriges Unterfangen, neue Vokale zu erlernen, da sich die Zunge kontinuierlich bewegen lässt und nie auf einer Position „einrastet“. Somit stellt sowohl die Produktion, als auch die Rezeption eines fremden Vokals eine besondere Herausforderung dar, welche gezielt trainiert werden muss (vgl. Wiese 2011: 36).

4.5 Der /i:/-/e:/-Vokalkontrast

Im Folgenden interessieren wir uns ausschließlich für den Kontrast zwischen den deutschen Phonemen /i:/ und /e:/, der in dieser Arbeit genauer beleuchtet werden soll. Zur besseren Vergleichbarkeit werden auch das jap. i und e aus artikulatorischer Perspektive betrachtet. Während sich das /i:/ in ähnlicher Ausführung in fast allen Sprachen der Welt findet und daher meist nicht eigens erlernt werden muss, stellt das fast ausschließlich im Deutschen vorkommende /e:/ für viele Deutschlernende möglicherweise eine Hürde dar (vgl. Kaneko 2005: 65). Das gespannt-lange (deutsche) i wird in der „Phonetik für Ausländer“ von Rudolf und Ilka Rausch mit folgenden Worten beschrieben (zur grafischen Veranschaulichung siehe Abb. 6):

Das [i:] ist ein hoher Vorderzungenvokal. Die Vorderzunge wölbt sich zum vorderen Hartgaumen, z.T. auch in Richtung der Alveolen auf. Die seitlichen Zungenränder haben Kontakt mit den Backenzähnen, wohingegen der vordere Zungenrand Kontakt mit den unteren Schneidezähnen hat. Die Lippen sind leicht abgehoben und weisen einen geringen Breitzug auf. Das Gaumensegel ist gehoben (Rausch R./Rausch I. 2002: 249).

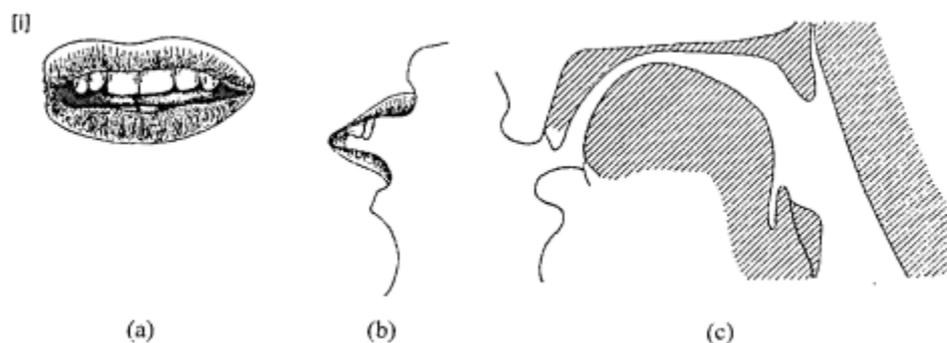


Abbildung 6: Artikulation des [i] des Deutschen (a)-(b) Mundformen (Front- und Seitenansicht), (c) Zungenstellung (Wängler 1976, aus Kaneko 2005: 38)

Es wird in „betonter Stellung lang, in unbetonter Stellung lang oder kurz ausgesprochen (*biete* [ˈbi:tə], *Böcklin* [ˈbœkli:n], *Iran* [ʔiˈʁa:n])“ (Kaneko 2005: 38). Das gespannt-lange (deutsche) *e* unterscheidet sich davon nur geringfügig und wird im Gegensatz dazu folgendermaßen definiert – die artikulatorischen Unterschiede zum deutschen „i“ wurden von mir hervorgehoben:

Das gespannt-lange *e* ist ein **mittelhoher** Vorderzungenvokal. **Die Zunge wölbt sich daher im vorderen Bereich gegen den harten Gaumen, jedoch nicht so stark wie beim [i:]. Der Zahnreihenabstand ist relativ klein, jedoch etwas größer als beim [i:].** Ein leichter Lippenbreitzug (Mundwinkel nach außen) ist möglich. Der vordere Zungenrand hat Kontakt mit den unteren Schneidezähnen. Das Gaumensegel ist gehoben. [e:] wird mit relativ großer Artikulationsspannung und mit großer Präzision gesprochen (Rausch R./Rausch I. 2002: 280).

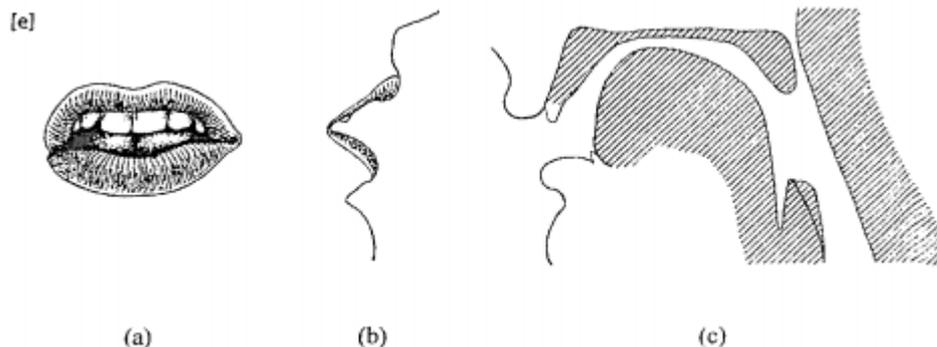


Abbildung 7: Artikulation des [e] des Deutschen (a)-(b) Mundformen (Front- und Seitenansicht), (c) Zungenstellung (Wängler 1976, aus Kaneko 2005: 40)

[i:] unterscheidet sich also lediglich hinsichtlich der Zungenhebung bzw. des Öffnungsgrades des Unterkiefers von [e:]. Das gespannt-lange *e* ist etwas weiter geöffnet als das gespannt-lange *i* (und geschlossener als das ungespannt-lange [ɛ]). Abgeleitet von [i:] wird [e:] durch eine geringfügig größere Öffnung des Unterkiefers artikuliert. „Zu beachten ist dabei, daß die Vorderzunge der Bewegung des Unterkiefers vollkommen passiv folgt. Auch die Lippen passen sich dem veränderten Zahnreihenabstand völlig passiv an“ (Rausch R./Rausch I. 2002: 281). Bei zu großem Zahnreihenabstand klingt das [e:] wie ein [ɛ:] (vgl. Rausch.R./Rausch I. 2002: 281).

Betrachten wir jetzt den japanischen /i-/e/-Kontrast. Die Artikulation des japanischen *i* (= ungerundeter geschlossener Vorderzungenvokal) wird folgendermaßen ausgeführt (siehe Abb. 8 für eine grafische Darstellung):

Die Lippen bilden einen schmalen Spalt, aber ihre Spannung ist nicht stark. Der Zahnreihenabstand ist gering. Die Zungenspitze liegt an den unteren Schneidezähnen an. Die oberen Zähne sind zu sehen. Der

Zungenrücken wölbt sich stark nach vorn gegen den harten Gaumen. Die Zungenspannung ist nicht so stark wie beim deutschen [i]. Der Zungenrand berührt eine größere Fläche am Gaumen und an den oberen Backenzähnen. Das Velum ist gehoben (Kaneko 2005: 9).

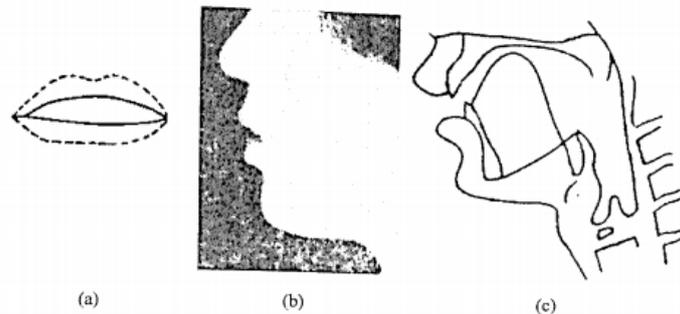


Abbildung 8: Artikulation des japanischen i-Lautes: (a) Lippenstellung, (b) Mundformenfoto (Seitenansicht), (c) Zungenstellung (aus Kaneko 2005: 10)

Als Hauptunterschied zum deutschen [i] wird also die Gespanntheit angegeben. Im Gegensatz dazu wird das japanische e (= ungerundeter mittlerer Vorderzungenvokal) folgendermaßen beschrieben (siehe Abb. 9):

Die Lippenöffnung ist geringer und oval. Die Muskelspannung ist viel leichter als beim französischen geschlossenen [e], aber ein wenig stärker als beim englischen [e] (in *men*). Die Zungenspitze liegt an den unteren Schneidezähnen. Die Oberzähne sind zu sehen. [...] Der Zungenrand berührt die Innenseite der oberen Backenzähne. Das Velum ist gehoben (Kaneko 2005: 9f).

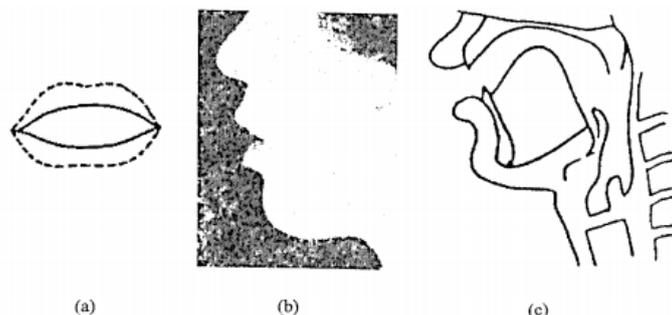


Abbildung 9: Artikulation des japanischen e-Lautes: (a) Lippenstellung, (b) Mundformenfoto, (c) Zungenstellung (aus Kaneko 2005: 10)

Wie man anhand der beiden Abbildungen und den Artikulationsausführungen erkennen kann, unterscheiden sich beide Vokale vor allem im Öffnungsgrad und damit einhergehend in der Lippenstellung (ähnlich wie beim deutschen Vokalkontrast). Kaneko hat das jap. i mit dem deutschen verglichen und kommt zu dem Schluss, dass sich beide Vokale nur geringfügig unterscheiden. Das japanische i liegt ihr zufolge zwischen dem deutschen [i] und [ɪ]:

Zungenhöhe, Berührungsfläche des Zungenrandes und Lippenlage beim japanischen [i] sind denen des Deutschen sehr ähnlich. Die Zungenspannung (Vorverlagerung der Zungenwurzel) ist jedoch nicht so stark wie die des deutschen [i]. Im proportionalen Verhältnis zwischen F_1 und F_2 ist das japanische [i] dem des deutschen [i] näher, das F_1 - F_3 -Verhältnis des japanischen [i] aber ist mit dem des deutschen [i] fast identisch. Im Ganzen steht das japanische [i] zwischen dem deutschen [i] und [ɪ] (Kaneko 2005: 62).

Das japanische e positioniert sich wie bereits erwähnt zwischen dem deutschen [e:] und [ɛ:] bzw. [ɛ] (siehe Abb. 5) und unterscheidet sich vom deutschen gespannt-langen e folgendermaßen:

Die Lippenstellung und der Zahnreihenabstand des japanischen [ɛ] [bei mir als jap. e bezeichnet; Anmerkung D.E.] sind fast identisch mit denen des deutschen [e], jedoch ist die Zungenhöhe des japanischen [ɛ] viel tiefer und die Berührungsfläche des Zungenrandes viel geringer und sie liegt weiter hinten als die des deutschen [e]; die Artikulation des japanischen [ɛ] ist (bis auf den Grad der Lippen) fast identisch mit der des deutschen [ɛ]. Die Formantenwerte des japanischen [ɛ] liegen sehr nahe an denen des deutschen [ɛ] (Kaneko 2005: 64).

Das /e:/ wurde allerdings (wie von mir beobachtet) von japanischen Muttersprachler/innen oft als /i:/ (und nicht als jap. e) wahrgenommen. Dies liegt vermutlich daran, dass sich bei dem gespannt-langen i und dem gespannt-langen e nur der Öffnungsgrad des Kiefers unterscheidet (und diese Laute somit sehr ähnlich sind), während das ungespannt-lange e [ɛ:] nicht nur offener, sondern auch mit geringerer Artikulationsspannung gesprochen wird, womit es dem jap. e ähnlicher ist (vgl. Rausch R./Rausch I. 2002: 283). [ɛ:] kontrastiert also in zwei Merkmalen mit den beiden anderen Lauten und ist somit höchstwahrscheinlich auch leichter von diesen zu unterscheiden. Das gespannt-lange e wird aus eigener Beobachtung bei der Artikulation durch Japaner/innen in der Regel durch das jap. e substituiert, da das Graphem <e> mit dem bekannten e-Laut des erstsprachlichen Lautinventars assoziiert wird. Diese Substitution als mögliche Quelle von Missverständnissen wird auch von Kaneko diskutiert (vgl. Kaneko 2005: 65).

Um diese Verwechslungen zu vermeiden, muss meiner Meinung nach erst einmal das Bewusstsein geschaffen werden, dass es sich dabei um zwei unterschiedliche Laute handelt, bevor man den neuen Vokal durch Übungen erlernen kann. Rudolf und Ilka Rausch propagieren hierbei den Weg über die Artikulationsseite: „Gelegentlich zeigt sich [...], daß Erfolge nur über die Artikulation und deren ständige Korrektur zu erzielen ist. Der artikulatorische Unterschied zum vergleichbaren muttersprachlichen Laut ist u. U. leichter zu erfassen als die aus ihnen resultierenden akustischen Ergebnisse“ (Rausch R./ Rausch I. 2002: 90).

5 Vorhersagen phonetischer Modelle

In diesem Kapitel sollen zwei phonetische Modelle, das Perceptual Assimilation Model (PAM) von Best bzw. dessen Weiterentwicklung für L2-Lernende (PAM-L2) von Best und Tyler und das Speech Learning Model (SLM) von Flege erläutert und die Vorhersagen beider Modelle für die Wahrnehmung des /i:/-/e:/-Vokalkontrasts durch japanische Sprecher/innen mit und ohne Erfahrung in der L2 Deutsch dargestellt werden. Davon ausgehend sollen – mit Bezug auf den theoretischen Teil – Hypothesen für den empirischen Teil generiert werden.

5.1 Das Perceptual Assimilation Model (PAM)

Das Perceptual Assimilation Model (PAM) ist ein Modell zur Beschreibung fremdsprachlicher Wahrnehmung. Es geht davon aus, dass artikulatorische Gesten für die Wahrnehmung entscheidend sind. Diese sind abhängig von den Artikulationsorganen (*articulatory organs*), sowie vom Artikulationsort (*constriction locations*) und vom Artikulationsgrad (*constriction degree*) (vgl. Best et al. 2001: 777). Die perzeptuelle Assimilation richtet sich nach der Ähnlichkeit dieser Faktoren zwischen einem muttersprachlichen und einem fremdsprachlichen Laut:

Its central premise is that mature listeners have a strong tendency to perceptually assimilate non-native phones to the native phonemes they perceive as most similar (Best/McRoberts 2003: 186).

Ursprünglich wurde das Modell entwickelt, um die unerwartet gute Unterscheidung von Zulu-Klicklauten durch amerikanische Hörer/innen mit L1 Englisch zu erklären (vgl. Best et al. 2001: 777). Die damit einhergehende entscheidende Erkenntnis war: „Adult discrimination of non-native speech contrasts is *not* uniformly poor“ (Best et al 2001: 776). Das heißt für naive Hörer/innen sind nicht alle fremdsprachlichen Lautkontraste gleich schwierig zu differenzieren (vgl. Best/Tyler 2007: 16).

Das PAM unterscheidet drei Möglichkeiten wie ein fremdsprachlicher Laut assimiliert werden kann: als kategorisiertes Exemplar eines Lautes der L1, als unkategorisierter Laut, der zwischen zwei nativen Phonemen fällt oder als nichtassimilierbarer nichtsprachlicher Laut, der keine Ähnlichkeit zu einem muttersprachlichen Laut aufweist. Für nichtmuttersprachliche Kontraste, die als sprachliche Laute identifiziert worden sind, sind dementsprechend folgende paarweise Assimilationstypen möglich (vgl. Best et al. 2001: 777):

- a) *Single-Category* [SC] *Assimilation*: beide L2-Laute werden einem L1-Phonem zugeordnet
- b) *Category-Goodness* [CG] *Difference*: die L2-Laute werden einem L1-Phonem zugerechnet, ein L2-Laut klingt allerdings eher wie das L1-Phonem als der andere
- c) *Two-Category* [TC] *Assimilation*: die L2-Laute werden zwei unterschiedlichen L1-Phonemen zugeordnet

Das Modell geht davon aus, dass nicht-erstsprachliche kontrastierende Phone besser unterschieden werden können, wenn diese zu zwei L1-Kategorien (Phonemen) zugeordnet werden, als wenn diese zu einer Kategorie zugeordnet werden. *Category-Goodness Difference* bezeichnet einen bestimmten Grad der Unterschiedlichkeit der kontrastierenden Laute und liegt daher von der Einordnung her zwischen der SC und der TC Assimilation. Die Phone werden dabei ebenfalls einer Kategorie zugeordnet, allerdings wird hierbei genauer differenziert. Dies ist der Fall, wenn von zwei Lauten einer mehr wie ein bekanntes muttersprachliches Phonem klingt als der andere. Zur Veranschaulichung der PAM-Vorhersagen sind hier nochmals alle drei Kategorien, nach dem Grad der vermuteten Diskriminierung geordnet, abgebildet (von TC = am besten bis SC = am schlechtesten):



Abbildung 10: PAM-Vorhersagen

Der Grund, weshalb die Englisch sprechenden Amerikaner/innen im vorher erwähnten Experiment von Best et al. die für sie unbekanntes Zulu-Laute unterscheiden konnten, ist, dass sie diese als Geräusche (*Non-Assimilable* [NA] *nonspeech sound*) wahrgenommen hatten, die nicht mit muttersprachlichen Phonemen in Verbindung gebracht und dem PAM zufolge gut bis sehr gut diskriminiert werden können (vgl. Best et al. 2001: 777f).

Kommen wir zu dem für diese Arbeit relevanten /i:/-/e:/-Vokalkontrast: Die Frage, zu welchem Assimilationstyp dieser Kontrast gezählt werden kann, ist nicht ganz einfach zu beantworten, da das deutsche /e:/ im Vokalraum in etwa in der Mitte zwischen jap. e und jap. i liegt (siehe Abb. 5). Da das /e:/ aber aus eigener Erfahrung im Gespräch mit Japaner/innen allerdings in der Regel nicht als jap. e, sondern als jap. i wahrgenommen worden war, kann die *Two-Category Assimilation* aller Voraussicht nach ausgeschlossen werden. Gestützt wird diese Beobachtung durch die bereits erwähnte Untersuchung von Ito (2010), die festgestellt hatte,

dass das deutsche /i:/ von japanischen Student/innen ohne Deutschkenntnisse zu 98% und das deutsche /e:/ zu 94 % als japanisches /i:/ verschriftlicht wurde (vgl. Mazuka et al. 2013: 196). Beide Laute werden also vermutlich mit dem jap. i in Verbindung gebracht.¹⁴ Würde es sich bei besagtem Vokalkontrast um eine *Single-Category Assimilation* handeln, kann von einer relativ schlechten Erkennungsrate ausgegangen werden. Bei dem Diskriminationsexperiment von Mazuka et al. (2013) wurde der /bi:k/-/be:k/-Kontrast von japanischen Erwachsenen mit 92% allerdings sehr gut erkannt, weshalb wahrscheinlich eher von einem *Category-Goodness Difference* ausgegangen werden kann. Demnach sollte die Erkennungsrate bei Japaner/innen ohne sprachlicher Erfahrung in der L2 Deutsch prinzipiell gut, aber nicht ausgezeichnet (*native-like*) sein. Die erste Hypothese lautet daher:

H1: Hörer/innen mit L1 Japanisch haben leichte Schwierigkeiten den deutschen /i:/-/e:/-Vokalkontrast zu diskriminieren. Die Erkennungsrate für die Gruppe der japanischen Student/inn/en ohne Spracherfahrung in der L2 Deutsch ist voraussichtlich gut, aber nicht ausgezeichnet (< 90%, gemäß CG des PAM).

Diese Annahme gilt aber nur für naive japanische Hörer/innen, sprich ohne Kontakt bzw. Erfahrung mit der Zielsprache. Um Voraussagen für „spracherfahrene“ Experimentteilnehmer/innen machen zu können, soll im nächsten Unterkapitel ein Blick auf das Speech Learning Model (SLM) von Flege und das PAM-L2 von Best und Tyler (2007) geworfen werden, welches SLM und PAM verknüpft und sprachliche Erfahrung (mit)berücksichtigt.

5.2 Das Speech Learning Model (SLM) und das PAM-L2

Das Speech Learning Model (SLM) von James Flege geht davon aus, dass neue Lautkategorien grundsätzlich in jedem Alter gebildet werden können (vgl. Flege 2007: 367). Der Zweitspracherwerb ist dem Modell zufolge also nicht durch eine kritische Periode bestimmt.:

The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning (Best/Tyler 2007: 24).

¹⁴ Die weitgehende Ähnlichkeit von jap. i und dem deutschen /i:/ wurde bereits angesprochen.

Der Fokus des SLM liegt auf der Lautproduktion, obgleich das Modell auch für die Vorhersage für Wahrnehmungsprobleme in der L2 verwendet wird. Es wird angenommen, dass Erst- und Zweitsprache in einem gemeinsamen phonologischen Raum (common phonological space) existieren und sich gegenseitig beeinflussen (vgl. Flege 2007: 366, CLI Kap. 3.1).

Entscheidend für die Diskrimination ist die phonetische Ähnlichkeit des L2-Lauts zu den Lauten der L1. Ein L2-Laut wird entweder als *neu* oder *ähnlich* wahrgenommen. Wenn sich ein Laut hinreichend stark von den Phonemen der Erstsprache unterscheidet, wird er von L2-Lernenden mit genügend fremdsprachlicher Erfahrung als unterschiedlich wahrgenommen und eine neue Lautkategorie wird entwickelt (vgl. Flege 2007: 367). Je unähnlicher also ein L2-Laut den erstsprachlichen Phonemen ist, desto leichter kann er unterschieden werden (vgl. Kerschhofer-Puhalo 2014: 87f). Ist dies nicht der Fall, greift die sogenannte Äquivalenzklassifikation und es wird keine neue Wahrnehmungskategorie gebildet (vgl. Bohn 1998: 4f):

By hypothesis, equivalence classification is most likely to prevent a new category from being established for an L2 sound when neighboring L1 categories are fully developed and when the L2 sound is perceived to be phonetically similar to a neighboring L1 sound (Flege 2007: 367).

Phonetische Unterschiede werden aber trotzdem wahrgenommen (!) (vgl. Flege 2007: 367), weshalb für den /i:/-/e:/-Vokalkontrast prinzipiell von einer guten Unterscheidung ausgegangen werden kann, selbst wenn keine neue Kategorie formiert worden ist. Abhängig davon, ob eine neue Kategorie gebildet wird oder nicht, unterscheidet Flege zwischen zwei Wahrnehmungsprozessen – *Phonetic Category Assimilation* und *Phonetic Category Dissimilation* (vgl. Flege 2007: 366f). Wichtig für diese Untersuchung ist ersterer Fall.

Das Modell geht davon aus, dass im Falle der Äquivalenzklassifikation L2-Lernende zuerst den ähnlichsten L1-Laut zur Artikulation von L2-Lauten benutzen, ohne ihn in irgendeiner Form zu modifizieren. Wenn sich der betreffende zweitsprachliche Laut hörbar von seinem L1 Gegenstück unterscheidet, setzt allerdings nach und nach ein Lernprozess ein und der Laut wird (mit zunehmender Spracherfahrung) Stück für Stück modifiziert. Für beide Laute der L1 und L2 dient allerdings nur eine Repräsentation im Langzeitgedächtnis – phonetische Kategorie genannt, die den Input in der Erst- und Zweitsprache berücksichtigt (vgl. Flege 2007: 368). Dieser Fall trifft vermutlich für die Produktion des deutschen /i:/-/e:/-Vokalkontrasts durch japanische Sprecher/innen zu.

Wie in Kapitel 4 gezeigt und in Abbildung 11 ersichtlich, sind sich /e:/ und /i:/ bzw. jap. i in artikulatorischer und akustischer Hinsicht sehr ähnlich (veranschaulicht durch die Formanten F1-F4, vgl. Kerschhofer-Puhalo 2014: 213), was vermutlich zu Schwierigkeiten in der

Wahrnehmung führen kann. Dazu kommt, dass die /e/-Phoneme in Japanisch und vielen Sprachen offener sind als das deutsche /e:/ und eher dem deutschen /ɛ/ ähneln:

German /e:/ is very constricted as compared to the /e/-phonemes in many other languages (e.g. Polish, SerBoCroatian, Albanian) resembling more an open [ɛ] (similar to German /ɛ/) than a more constricted [e]-quality. Difficulties in the differentiation of German /i:/ and /e:/ can therefore be expected in L2 perception (Kerschhofer-Puhalo 2014: 177).

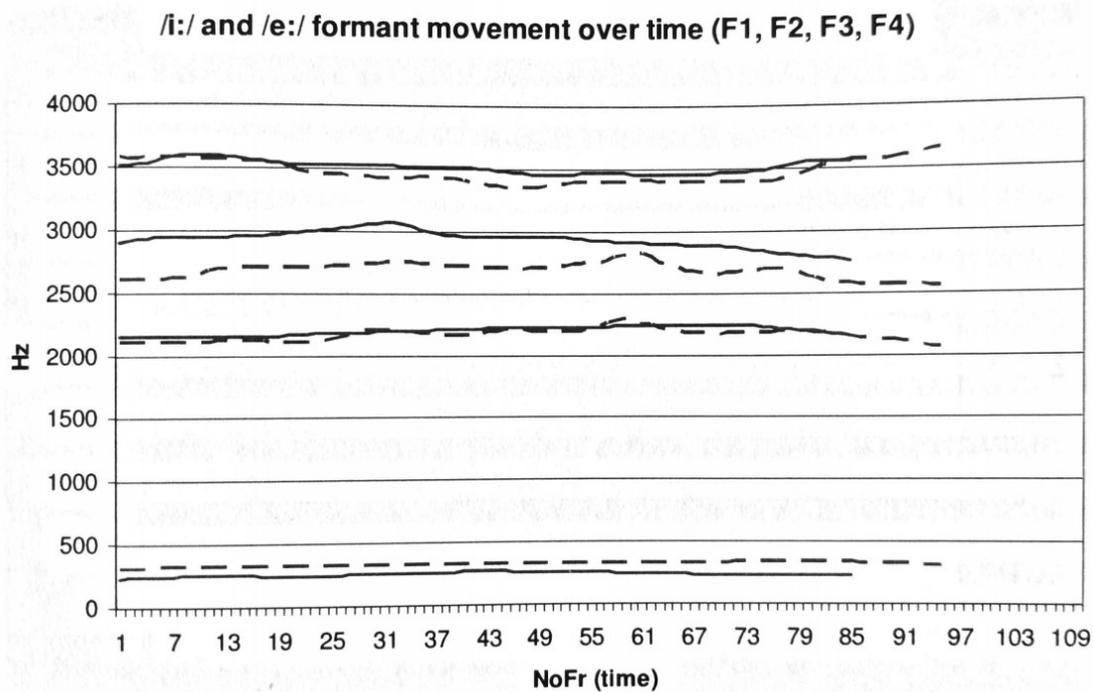


Abbildung 11: Vergleich von /i:/ (durchgezogene Linie) und /e:/ (strichlierte Linie) im /pVt/-Kontext (aus Kerschhofer-Puhalo 2014: 177).

Gestützt wird diese Annahme durch die bereits an anderer Stelle angesprochenen Forschungsergebnisse von Nadja Kerschhofer-Puhalo (2014), die festgestellt hat, dass sich in den Konfusionsdaten ihres Korpus insgesamt eine starke Tendenz zeigt, /e:/ als /i:/ zu perzipieren, während umgekehrte Substitutionen deutlich seltener sind (vgl. Kerschhofer-Puhalo 2014: 8, 689). Die Gesamtheit der nicht-nativen Hörer/innen (mit 12 unterschiedlichen Erstsprachen) nahm das deutsche /e:/ zu 12% als i: wahr (vgl. Kerschhofer-Puhalo 2014: 242):

Within the class of front non-rounded vowels, the greatest spread of responses over several categories is observed in the data for German /e:/, demonstrating significant difficulties in vowel differentiation within the class of front non-rounded vowels. [...] Substantial portions of wrong identifications for /e:/ are due to substitutions of /e:/ > /i:/ that are observed in all sub-samples except Romanian and German. Attractive substitutes for /e:/ were /i:/ (12.3%), /ɪ/ (3.6%) and <ei> (4.7%) (Kerschhofer-Puhalo 2014: 573).

Mit zunehmender Spracherfahrung sollte die Unterscheidung beider Vokale signifikant verbessert werden. Werfen wir, bevor die zweite Hypothese formuliert wird, einen Blick auf das PAM-L2, der Weiterentwicklung des Perceptual Assimilation Models, das PAM und SLM miteinander verknüpft und sprachliche Erfahrung in der Zweitsprache berücksichtigt. Grundsätzlich wird auch von Best und Tyler konstatiert, dass perzeptuelles Lernen möglich ist:

The findings suggest that the answer is a qualified „yes“. Perceptual learning occurs for some L2 contrasts, but seems to depend on their phonological and phonetic relationship to the L1, specifically on perceived similarities vs. dissimilarities to L1 phonemes (Best/Tyler 2007: 20).

Spracherfahrene Hörer/innen diskriminieren Kontraste dabei signifikant besser als unerfahrene, jedoch weniger gut als muttersprachliche Hörer/innen (vgl. Best/Tyler 2007: 20). Entscheidend ist laut PAM-L2 hierfür die phonetische und phonologische Ähnlichkeit (im Gegensatz zum SLM). Vier Fälle werden für L2-Kontraste unterschieden (vgl. Best/Tyler: 28-30):

- 1) Nur ein L2-Laut wird einer L1-Kategorie zugeordnet
- 2) Beide L2-Laute werden einer L1-Kategorie zugeordnet, aber einer wird als ähnlicher als der andere wahrgenommen (entspricht CG Difference)
- 3) Beide L2-Laute werden als gleich gute Exemplare derselben L1-Kategorie zugeordnet (entspricht SC Assimilation)
- 4) keine phonologische Assimilation zwischen L1 und L2

Betrachten wir die ersten beiden Möglichkeiten (Fall 3 und Fall 4 kann nach den vorherigen Ausführungen für diese Studie ausgeschlossen werden). Für den ersten Fall wird kein weiteres perzeptuelles Lernen erfolgen, da die Diskrimination von Anfang an keine Probleme bereitet, weil sich das Kontrastpaar auf phonetischer und/oder phonologischer Ebene hinreichend weit voneinander unterscheidet. Es handelt sich hierbei nach PAM-Termini entweder um eine TC Assimilation oder um einen kategorisiert-unkategorisiert-Unterschied (vgl. Best/Tyler 2007: 28).

Wird ein gradueller Unterschied zwischen beiden L2-Lauten im Vergleich zum erstsprachlichen Phonem festgestellt und der ähnlichere der beiden Laute einer L1-Kategorie zugeordnet (Fall 2), wird für diesen voraussichtlich keine neue phonetische Kategorie geformt. Für den anderen Laut sollte hingegen nach und nach eine neue L2-Kategorie gebildet werden:

We speculate that, initially, it would be perceptually learned as new L2 phonetic variant of the L1 phonological category. With continued exposure, the learner should learn to perceive the lexical-functional contrasts between the L2 phones, and to develop a new phonological category for the phonetically “deviant phone” The likelihood of the learner developing a new phonetic category for the “better-fitting” L2 phone depends upon the degree of its perceived similarity to the L1 category. Two

possibilities arise: a good exemplar will be assimilated to a common L1/L2 phonetic category, with little further perceptual differentiation from the L1 entity, whereas a new phonetic category is likely to be perceptually learned for a less-good exemplar of the L1 category (Best/Tyler 2007: 29).

Die Unterscheidung sollte sehr gut, aber schlechter als bei einer TC-Assimilation sein. Dieser Fall trifft wahrscheinlich für die /i:/-/e:/-Diskrimination durch Hörer/innen mit L1 Japanisch zu. Mit ausreichend sprachlicher Erfahrung sollte für das unterschiedlichere [e:]-Phon eine neue Kategorie formiert werden, während das [i:] voraussichtlich zum jap. i assimiliert und als „guter Vertreter“ der phonetischen Kategorie auch mit Spracherfahrung assimiliert bleiben wird, weshalb als zweite Hypothese festgehalten werden kann:

H2: Die Diskrimination der Gruppe der japanischen Student/inn/en mit Spracherfahrung ist signifikant besser als jene ohne sprachlicher Erfahrung in der L2 Deutsch und voraussichtlich annähernd auf erstsprachlichem Niveau.

5.3 Hypothesen für den empirischen Teil

Die beiden aus den phonetischen Modellen gewonnenen Hypothesen (PAM für H1, PAM-L2 und SLM für H2), die es im empirischen Teil zu überprüfen gilt, sollen zugunsten besserer Übersichtlichkeit hier nochmals angeführt werden:

H1: Hörer/innen mit L1 Japanisch haben leichte Schwierigkeiten den deutschen /i:/-/e:/-Vokalkontrast zu diskriminieren. Die Erkennungsrate für die Gruppe der japanischen Student/inn/en ohne Spracherfahrung in der L2 Deutsch ist voraussichtlich gut, aber nicht ausgezeichnet (< 90%, gemäß CG des PAM).

H2: Die Diskrimination der Gruppe der japanischen Student/inn/en mit Spracherfahrung ist signifikant besser als jene ohne sprachlicher Erfahrung in der L2 Deutsch und voraussichtlich annähernd auf erstsprachlichem Niveau.

Mithilfe des Diskriminationsexperiments sollen beide Hypothesen bestätigt bzw. falsifiziert werden und darauf aufbauend Schlüsse für die Wahrnehmung des /i:/-/e:/-Vokalkontrasts durch japanische Hörer/innen gezogen werden.

6 Empirischer Teil

Wissenschaft erhebt den Anspruch, über die subjektiven Bewertungen und das Alltagswissen des Einzelnen hinauszureichen. Sie behauptet diesen Anspruch, weil sie ihr Wissen über die untersuchten Gegenstände in methodisch kontrollierter und systematischer Weise gewinnt und nicht aufgrund individueller Einzelerfahrungen (Schmelter 2013: 33).

Im Sinne einer transparenten Arbeitsweise sollen in diesem Abschnitt die durchgeführten Experimente beschrieben und die methodische Vorgehensweise nachvollziehbar präsentiert werden. Der empirische Teil ist folgendermaßen gegliedert: Zuerst wird das Forschungsdesign als Ganzes dargestellt, visualisiert und die Experimente hinsichtlich ihres Nutzens für diese Arbeit beleuchtet. Danach folgen Überlegungen zur Sicherstellung der Gütekriterien und die einzelnen Experimente werden im Detail beschrieben. Das für diese Untersuchung zentrale Diskriminationsexperiment wird zuerst angeführt. Danach folgt ein Identifikationsexperiment worin die Wahrnehmung des deutschen /e:/ und /i:/ bei japanischen Deutschlernenden untersucht wird.

6.1 Forschungsdesign

An dieser Stelle sollen nochmals die Forschungsfragen in Erinnerung gerufen werden, auf deren Grundlage das Experimentdesign entwickelt wurde:

1. Können japanische L1-Hörer/innen (mit und ohne DaF-Erfahrung) den deutschen /i:/-/e:/-Vokalkontrast – unter den durch das Experiment vorgegebenen Bedingungen – unterscheiden? Wenn ja, wie gut?
2. Welche Faktoren/Variablen haben Einfluss auf die fremdsprachliche Wahrnehmung der beiden Vokale? Spielt die sprachliche Erfahrung in der Zweitsprache eine Rolle?
3. (Inwiefern) verändert sich das Ergebnis unter erschwerten Bedingungen (SNR 0dB/-5dB) im Vergleich zu deutschen L1-Hörer/innen?

Bei der vorliegenden Studie handelt es sich um eine quantitative experimentelle Forschung, d.h. Daten sollen unter kontrollierten Bedingungen erhoben werden. Die erste Frage soll mithilfe eines AXB-Diskriminationsexperimentes beantwortet werden. Die durch das Experiment gewonnenen Ergebnisse sollen Aufschluss darüber geben wie gut (oder schlecht) 2 Gruppen¹⁵

¹⁵ Ursprünglich waren 3 Gruppen geplant (ohne Deutschkenntnisse, Anfänger/innen und Fortgeschrittene). Da viele Proband/inn/en der Gruppe B allerdings bereits an der Oberschule Deutschunterricht hatten, wurde dieser Teil auch der fortgeschrittenen Gruppe (C) hinzugefügt und zu einer Gruppe fusioniert (nachfolgend BC genannt).

japanischer L1-Sprecher/innen (in diesem Fall Studierende an einer japanischen Universität) den deutschen /i:/-/e:/-Kontrast diskriminieren können. Dabei sollen die im theoretischen Teil entwickelten Hypothesen überprüft werden (siehe Kapitelende). Die beiden Gruppen unterscheiden sich hinsichtlich deren Deutschkenntnisse bzw. Kontakt mit der L2 voneinander. Unterschiede in den Resultaten beider Gruppen sollen den Einfluss sprachlicher Erfahrung aufzeigen (Frage 2). Die zweite Forschungsfrage wurde außerdem bereits im Theorieteil besprochen und kann aufgrund des Umfangs nicht vollständig empirisch erforscht werden.

Als Anregung für das AXB-Experimentdesign dieser Untersuchung fungiert eine Studie von Best et al. (2001), worin die Diskrimination von nicht-muttersprachlichen Konsonantenkontrasten untersucht wurde. Anstatt von Konsonanten wurden allerdings Vokalkontraste und anstelle von Silben wurden 24 Minimalpaare als Stimuli verwendet, die sich lediglich im /i:/-/e:/-Kontrast voneinander unterscheiden (gemischt mit 12 anderen, die für diese Arbeit nicht relevant sind). Eine genaue Beschreibung des Ablaufs erfolgt in Kapitel 6.4. Das Forschungsdesign dieser Studie basiert auf einem Experiment, welches ich vor fast drei Jahren im Zuge eines Phonetikseminars entwickelt habe. Meine Annahme, dass die Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts den Japanisch sprechenden Proband/inn/en Probleme bereitet, konnte darin nicht bestätigt werden. Dem Rätsel der Abweichung zwischen den in der „realen Welt“ gemachten eigenen Erfahrungen und den Ergebnissen des Experiments konnte ich damals nicht auf die Spur kommen.

Für die Masterarbeit sollte das Experiment daher grundlegend überarbeitet werden. In meinem revidierten Design habe ich drei Schwierigkeitsstufen eingefügt (mit und ohne Störgeräusch), um die Bedingungen des Hörens zu erschweren und möglicherweise größere Unterschiede herausarbeiten zu können. Als Störgeräusch wurde ein sogenanntes Speech Shaped Noise (SSN) in unterschiedlicher Stärke (0dB SNR und -5dB SNR) eingesetzt, das mithilfe eines Praatskripts aus einer vielstimmigen Aufnahme eines Stimmengewirrs in einer Bar erzeugt wurde. Für die Wahl der Störstärke habe ich mich an diversen Studien mit Störgeräuschen orientiert (vgl. Hazan/Simpson 1996, Cooper/Cutler 2004, Bent et al. 2010, Jin/Liu 2014). Die Ergebnisse der unterschiedlichen Teile sollen in der Auswertung miteinander verglichen werden (Frage 4). Es wird davon ausgegangen, dass beide Gruppen im Vergleich zur Kontrollgruppe unter erschwerten Bedingungen signifikant schlechter abschneiden (vgl. Cutler 2004, → H3). Als Referenz und wesentliche Grundlage für das Diskriminationsexperiment dient außerdem die bereits mehrmals angesprochene Studie von Mazuka et al. (2013), worin die Diskrimination von drei deutschen Vokalkontrasten (/bi:k/-/be:k/, /bu:k/-/bo:k/ und /bu:k/-/by:k/) durch japanische Kleinkinder und Erwachsene experimentell untersucht wurde.

Ein Test von Ito (2010), welcher in derselben Studie angeführt, der für mich allerdings leider nicht greifbar ist (da auf Japanisch und unveröffentlicht) fungiert als Anregung für das zweite Experiment (Identifikationsexperiment). Darin sollten japanische Student/inn/en deutsche Wörter mit japanischen Kana aufschreiben. Die Autorin kam zu dem Ergebnis, dass das deutsche /i:/ zu 98 % und das /e:/ zu 94 % als japanisches /i/ verschriftlicht wurde, was wiederum mit meinen im Gespräch mit Japaner/inne/n gemachten Erfahrungen übereinstimmt, dem Ergebnis der oben genannten Studie aber widerspricht (vgl. Mazuka et al. 2014: 196). Möglicherweise liegt der Widerspruch darin, dass es sich einmal um eine Diskrimination und einmal um eine Identifikation handelt. Aus diesem Grund habe ich zusätzlich zu dem Diskriminationsexperiment ein zweites Experiment durchgeführt, um gesondert die Identifikation zu testen. Getestet wurden Studierende eines Deutschkurses der japanischen Keio-Universität (verschiedene Niveaustufen gemischt). Buchstabiert vorgelesene und möglichst unbekannte österreichische Familiennamen sollten dabei notiert werden.

Nachfolgend sollen folgende drei Hypothesen zur Wahrnehmung des deutschen /i:/-/e:/-Vokalkontrasts durch japanische Hörer/innen experimentell geprüft werden:

H1: Hörer/innen mit L1 Japanisch haben leichte Schwierigkeiten den deutschen /i:/-/e:/-Vokalkontrast zu diskriminieren. Die Erkennungsrate für die Gruppe der japanischen Student/inn/en ohne Spracherfahrung in der L2 Deutsch ist voraussichtlich gut, aber nicht ausgezeichnet (< 90%, gemäß CG des PAM).

H2: Die Diskrimination der Gruppe der japanischen Student/inn/en mit Spracherfahrung ist signifikant besser als jene ohne sprachlicher Erfahrung in der L2 Deutsch und voraussichtlich annähernd auf erstsprachlichem Niveau.

H3: Beide Gruppen schneiden signifikant schlechter unter erschwerten Bedingungen (SNR 0dB/-5dB) ab, während die Unterscheidungsrate der Kontrollgruppe annähernd gleichbleibt.

6.2 Visualisierung des Forschungsdesigns

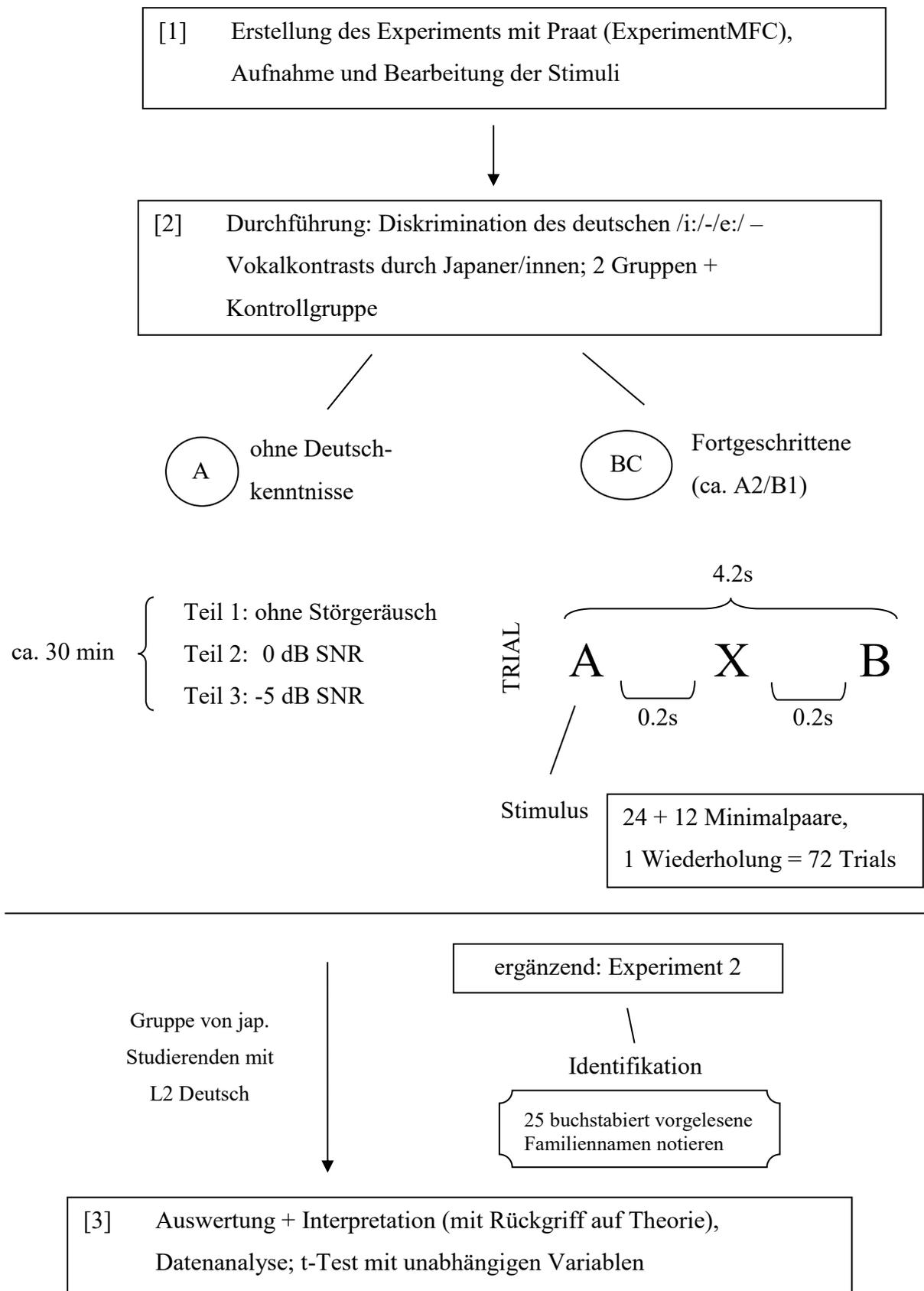


Abbildung 12: Visualisierung des Forschungsdesigns

6.3 Maßnahmen zur Sicherstellung der Gütekriterien

Gütekriterien sind abstrakte Konstrukte, welche die Qualität und Wissenschaftlichkeit empirischer Forschung gewährleisten sollen (vgl. Schmelter 2013: 35). Zentral für die Bewertung einer quantitativen Studie sind vor allem die aus der Testtheorie stammenden Gütekriterien Objektivität, Reliabilität und Validität (vgl. Schmelter 2013: 38).

Validität bewertet, ob das gemessen wird, was gemessen werden soll. Objektivität bezeichnet die Unabhängigkeit der Forschung vom Forschenden, während Reliabilität die Zuverlässigkeit überprüft, d.h. es wird kontrolliert, ob bei der Wiederholung eines Tests unter gleichen Bedingungen dasselbe Ergebnis herauskommt (vgl. Albert/Marx 2017: 14). „Dabei baut Validität insofern auf den anderen beiden Gütekriterien auf, als ohne Objektivität keine Reliabilität und ohne Reliabilität keine Validität gegeben sein kann“ (Schmelter 2013: 38).

Zur Wahrung der Gütekriterien wurden folgende Maßnahmen getroffen: Im Sinne der Transparenz werden Ablauf, Proband/inn/en, Stimuli und die gewonnenen Daten genau beschrieben. Die verwendeten Skripts bzw. Konfigurationsdateien sind größtenteils selbsterklärend und liegen dem Anhang bei. Alle Teilnehmenden haben denselben Diskriminationstest unter den gleichen durch das Experiment vorgegebenen Bedingungen gemacht. Die für das Experiment relevanten Variablen wurden bereits im Theorieteil diskutiert, mögliche Störeffekte sollten im Sinne der Validität soweit als möglich vermieden werden, wobei die externe Validität (Übertragung auf andere Kontexte) aufgrund der Faktorenkomplexion nur bedingt erreichbar ist (vgl. Schmelter 2013: 39). In der Diskussion der Ergebnisse sollen die gewonnenen Daten interpretiert und die eigene Forschung kritisch betrachtet werden.

Um das Ergebnis nicht durch Einflüsse von außen zu verfälschen, wurde das Diskriminationsexperiment in einer möglichst ruhigen Umgebung durchgeführt (wichtig für Gütekriterien Validität und Reliabilität). Leider stand mir kein schalldichter Raum zur Verfügung, um äußere Störquellen vollständig auszuschalten. Das Experiment wurde auf meinem Laptop durchgeführt, wobei ich bei der Durchführung immer persönlich dabei war, um bei etwaigen Fragen zur Verfügung zu stehen. Die Teilnehmenden wurden vor dem Experiment mündlich (auf Deutsch oder Englisch) über den Ablauf informiert und zu ihrer Sprachbiografie befragt, wobei darauf geachtet wurde, sie nicht in irgendeiner Form zu beeinflussen. Im Sinne der Objektivität sollte der Einfluss der forschenden Person möglichst gering gehalten werden.

Die genaue Zielsetzung des Experiments wurde bei der vorhergehenden Unterweisung nicht bekannt gegeben, lediglich, dass es sich um ein Wahrnehmungsexperiment handelt. Zusätzlich waren die Handlungsanweisungen (in deutscher Sprache) auf dem Bildschirm ersichtlich.

Die Testpersonen erhielten qualitativ relativ hochwertige geschlossene Kopfhörer (Teufel Aureol Massive). Die Proband/inn/en wurden einzeln getestet, da das Experiment jeweils nur auf einem Laptop gestartet wurde. Die Sprachaufnahmen der Items erfolgten durch einen 66-jährigen männlichen Sprecher der österreichischen Varietät des Standarddeutschen. Die Aufnahmen wurden in einem professionellen Tonstudio gemacht. Um zu gewährleisten, dass die Minimalpaare des AXB-Diskriminationsexperiments im Sinne der Validität bis auf den Vokalkontrast vollkommen ident sind, habe ich mich dazu entschieden, den „i-Vokal“ in das „e-Wort“ mit dem Programm Audacity hineinzuschneiden („splicing“). Auf diese Weise sollte verhindert werden, dass die Teilnehmenden auf andere Weise als durch den betreffenden Vokalkontrast einen Unterschied erkennen können. Eine Kontrollgruppe mit Hörer/innen mit L1 Deutsch soll die Reliabilität der Ergebnisse garantieren. Eine gewisse Zufälligkeit des AXB-Modus durch die zwei Auswahlmöglichkeiten ist leider nicht gänzlich zu verhindern (für den Fall, dass Proband/inn/en raten und dabei richtig liegen). Das heißt, die tatsächliche Fehlerrate/Schwierigkeit des Vokalkontrasts wäre unter Umständen höher anzusetzen. Auch Lerneffekte sind nicht auszuschließen (vgl. Kerschhofer-Puhalo 2014: 81).

Entscheidend für die resultierenden Daten ist auch der Experimentmodus. Zwei Haupttypen sind dabei zu unterscheiden: Diskriminations- und Identifikationstests. Während bei ersteren eher die auditorische Repräsentation im Kurzzeitgedächtnis angesprochen wird, wird für die Fähigkeit Laute der L2 identifizieren zu können, die spezifische Lautrepräsentation im Langzeitgedächtnis benötigt (vgl. Kerschhofer-Puhalo 2014: 78). Wichtig ist hierfür unter anderem das Interstimulus Intervall (ISI): Je kürzer dieses ist, desto eher wird die Information im Kurzzeitgedächtnis verwendet und desto einfacher ist die Unterscheidung für die Versuchsteilnehmer/innen (vgl. Kerschhofer-Puhalo 2014: 78f, Mora 2007: 2). Mit anderen Worten: Abhängig vom Experimentdesign ist unter Umständen auch der verwendete Wahrnehmungsmodus ein anderer.

Beim zweiten Experiment wurden die Versuchsteilnehmer/innen gleichzeitig getestet. Allen Proband/inn/en des Experiments 2 wurden die österreichischen Familiennamen in derselben Reihenfolge über einen Lautsprecher präsentiert. Informationen über den Ablauf erfolgten zuvor auf Deutsch und Japanisch.

6.4 Experiment 1 – *lieben oder leben* – Diskrimination

6.4.1 Proband/inn/en

Für das Experiment wurden insgesamt 34 Studierende der Dokkyo Universität zwischen 18 und 22 Jahren getestet (27 weiblich, 7 männlich, Durchschnittsalter 19,2 Jahre). Die Teilnehmer/innen sind allesamt Freiwillige ohne Erfahrung mit linguistischen Experimenten und hinsichtlich ihrer Sprachbiografie relativ homogen: alle haben mindestens 7 Jahre Englisch Schule gelernt, vereinzelt wurden (von insgesamt 6 Proband/inn/en) auch andere Fremdsprachen wie Französisch, Spanisch, Chinesisch und Koreanisch gelernt. Keine/r der Teilnehmer/innen hat jemals außerhalb Japans gelebt. Zwei Proband/inn/en wurden aus dem Pool aussortiert, da sie nicht den Vorgaben bezüglich Erstsprache und Sprachniveau in der L2 Deutsch entsprachen.

Ursprünglich waren für die Untersuchung drei Gruppen geplant, welche sich hinsichtlich ihrer Deutschkenntnisse voneinander unterscheiden: Japaner/innen ohne jeglicher sprachlicher Erfahrung in der L1 Deutsch (Gruppe A), Anfänger/innen (Gruppe B) und fortgeschrittene Lernende (Gruppe C). Da allerdings zahlreiche Student/inn/en der Gruppe B bereits in der Oberschule Deutschunterricht hatten, konnte diese Aufteilung nicht aufrechterhalten werden. Aus diesem Grund wurden die Proband/inn/en in zwei Gruppen geteilt: die ursprüngliche *Gruppe A* (10 Proband/inn/en, 9 weiblich, 1 männlich), worin sich ausschließlich Sprecher/innen mit L1 Japanisch befinden, die nie Deutsch gelernt haben und *Gruppe BC* (17 Proband/inn/en, 13 weiblich, 4 männlich), welche aus der gesamten Gruppe C und jenem Teil der zweiten Gruppe besteht, welche bereits seit mindestens zwei Jahren Deutsch lernen. Das Sprachniveau dieser kombinierten zweiten Gruppe in der L2 Deutsch bewegt sich etwa auf A2/B1. Die Proband/inn/en aus Gruppe A kommen aus verschiedenen Studienrichtungen, Gruppe B und C studieren alle Deutsch. Die Versuchsteilnehmer/innen der Gruppe C stammen zudem alle aus derselben Klasse. Zusätzlich zu den zwei Gruppen gibt es eine Kontrollgruppe (KG), welche sich aus 12 Student/inn/en (2 weiblich, 10 männlich, Durchschnittsalter 23,4 Jahre) mit L1 Deutsch an österreichischen Universitäten zusammensetzt. Die Teilnehmenden der Kontrollgruppe stammen aus dem deutschsprachigen Raum, kommen aus verschiedenen Studienrichtungen und sind auch hinsichtlich der gelernten Fremdsprachen relativ heterogen.

6.4.2 Stimuli

Das Experiment verwendet ein sogenanntes AXB- bzw. BXA-Design. A und B stehen dabei für die beiden kontrastierenden Stimuli, das X steht stellvertretend für eines der beiden Stimuli. Mögliche Varianten sind daher AAB, ABB, BAA oder BBA. Pro Trial werden folglich 3 Stimuli randomisiert abgespielt. Das Interstimulusintervall (ISI) beträgt 0,2 Sekunden. Insgesamt 36 Minimalpaare mit Vokalkontrasten wurden von einem 66-jährigen männlichen Sprecher der österreichischen Varietät des Standarddeutschen (Raum Niederösterreich) gesprochen. 24 davon unterschieden sich lediglich hinsichtlich des /i:/-/e:/-Vokalkontrasts und waren für die Auswertung relevant. Verwendet wurden dazu im Deutschen vorkommende ein- bis dreisilbige Wörter (z.B. *lieben-leben*). Ich habe mich entschieden, real existierende Wörter (und keine Nonsenswörter) einzusetzen, um natürlicher Kommunikation etwas näher zu kommen. Die beiden zu unterscheidenden Vokale sind ungerundet, lang und gespannt (siehe Kap. 4.5) und im Japanischen nicht kontrastiv. Der Vokalkontrast findet sich im An- (z.B. *Igel-Egel*) In- (z.B. *lieben-leben*) und Auslaut (z.B. *sie-See*). Mit dem Phonetikprogramm Praat (Boersma/Weenink: 2017) durchgeführte akustische Messungen der Zielvokale am Beispiel der gesprochenen Stimuli *lieben* und *leben* sind in Tabelle 3 ersichtlich. Die ersten drei Formanten und die Grundfrequenz gelten als wesentliche Kenngrößen für den /e:/- und /i:/-Vokal. Es werden hier die gemittelten Werte angegeben.

Stimuli	F0 (Hz)	Dauer (ms)	F1 (Hz)	F2 (Hz)	F3 (Hz)
['li:bŋ]	149	282	284	2096	2800
['le:bŋ]	124	284	298	2093	2594

Tabelle 3: Akustische Messungen der beiden Zielvokale am Beispiel eines Minimalpaares

Diese exemplarischen akustischen Messungen verdeutlichen nochmals, wie nahe die Formanten der zwei Vokale beieinanderliegen. Daher verwundert es auch nicht, wenn beide Vokale von Sprecher/innen mit L1 Japanisch meist als /i:/ gehört wurden (vgl. Mazuka et al. 2013: 196). Die Stimuli wurden in einem professionellen Tonstudio aufgenommen und normalisiert (Auflösung 44,1 kHz, 16bit mono wav-Datei). Folgendes Equipment wurde dazu verwendet: Mikrofon: Schoeps MK4, Vorverstärker: RME Micstasy, DAW: Pro Tools.

Die Erzeugung synthetischer Vokale wurde aufgrund nicht zufriedenstellender Ergebnisse wieder verworfen. Gegen die Verwendung synthetischer Stimuli spricht auch, dass diese möglicherweise einen unvorhergesehenen Einfluss auf die Sprachwahrnehmung haben (vgl.

Kerschhofer-Puhalo 2014: 76). Stattdessen wurden – wie bereits erwähnt – natürliche Stimuli verwendet und der „i-Vokal“ in das „e-Wort“ hineingeschnitten (siehe Abb. 13 am Beispiel der Stimuli *lieben* und *leben*). Durch diese Manipulation ist gewährleistet, dass die Silbenränder ident sind und die Diskrimination nicht ungewollterweise vereinfacht wird.

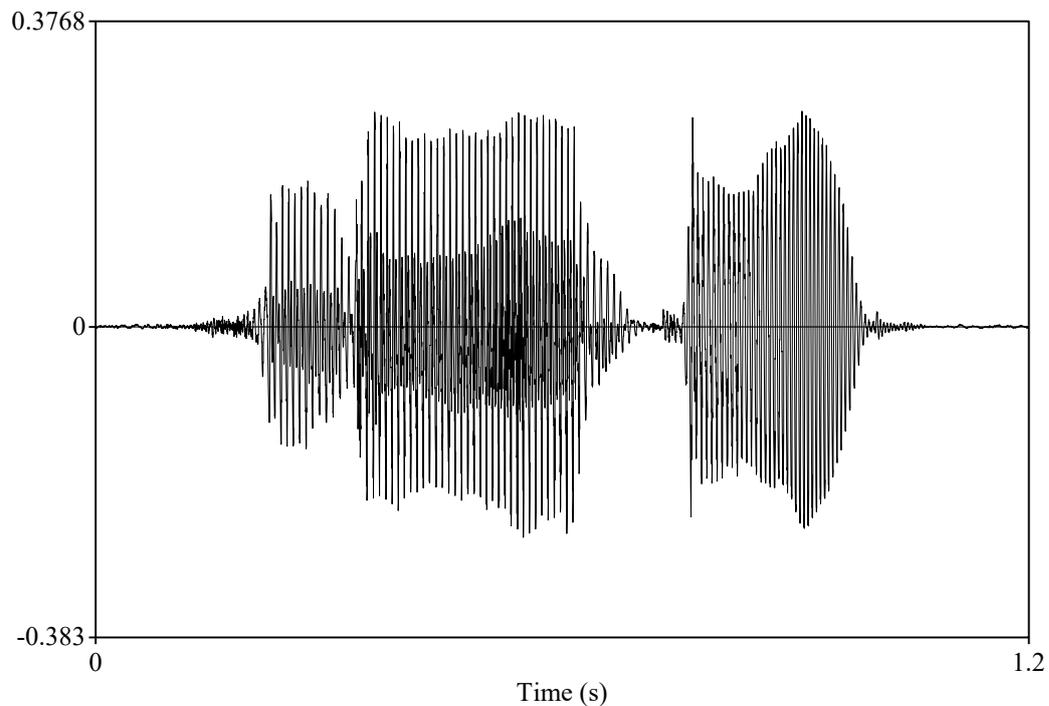
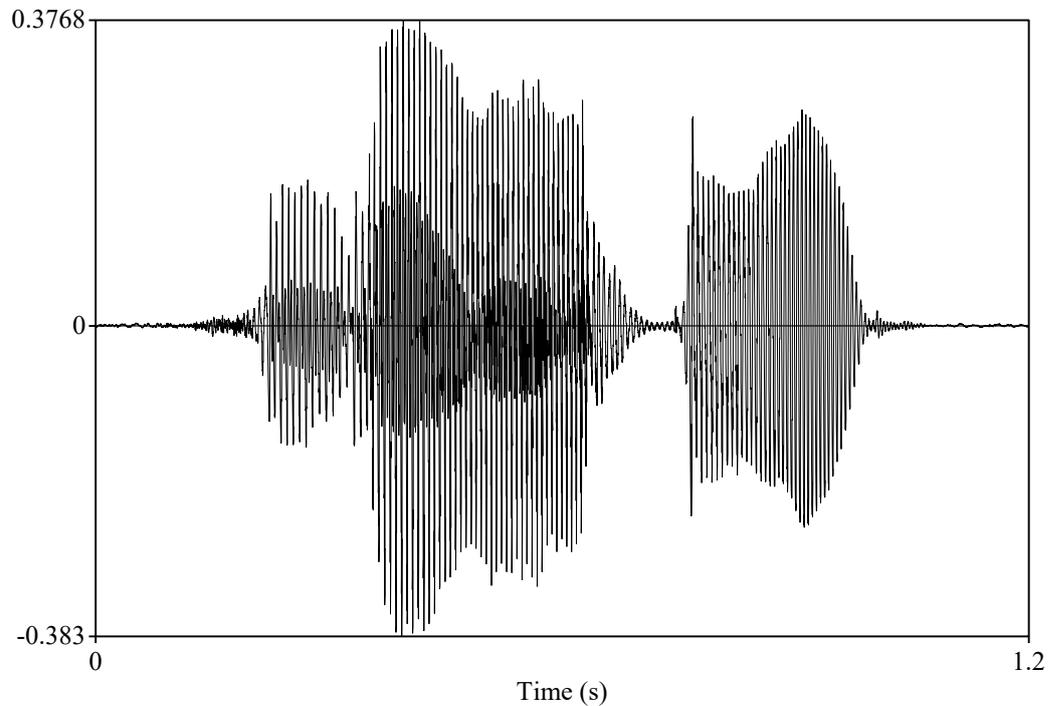


Abbildung 13: Oszillogramm der Stimuli [li:bɐ̃] (oben) und [le:bɐ̃] (unten)

Für Teil 2 und Teil 3 des Experiments wurde mithilfe eines Praatskriptes von Matt Winn (siehe Anhang C) ein Speech Shaped Noise (SSN) aus einem Stimmengewirr in einer Bar erzeugt (Ntounis 2012), welches mit den Stimuli auf 0dB und -5dB SNR mithilfe eines weiteren Skriptes von Daniel McCloy gemischt (siehe Anhang D) wurde.

6.4.3 Ablauf

Vor der Durchführung des Experiments wurden die Teilnehmer/innen über den Ablauf auf Deutsch oder Englisch informiert. Nach der Startfolie (hier nicht abgebildet) startet das Experiment direkt und das erste von insgesamt 72 Trials (pro Teil) wird abgespielt. Abbildung 14 zeigt den Bildschirminhalt während des Experiments. Die Proband/inn/en hören über Kopfhörer jeweils drei Wörter. Das mittlere ist dabei entweder gleich wie der erste oder der dritte Stimulus: Wenn die ersten beiden Wörter gleich sind, d.h. $X = A$ ist, soll auf die Schaltfläche „A“ geklickt werden (\rightarrow AAB oder BBA). Sind hingegen Stimulus 2 und 3 ident, d.h. $X = B$, soll der Proband/ die Probandin die Schaltfläche „B“ betätigen (\rightarrow ABB oder BAA). Mit einem Klick auf den Button „nochmal?“ kann ein Trial ein weiteres Mal abgespielt werden, falls die Teilnehmer/innen es aus irgendeinem Grund beim ersten Mal nicht gehört haben oder sich unsicher sind.

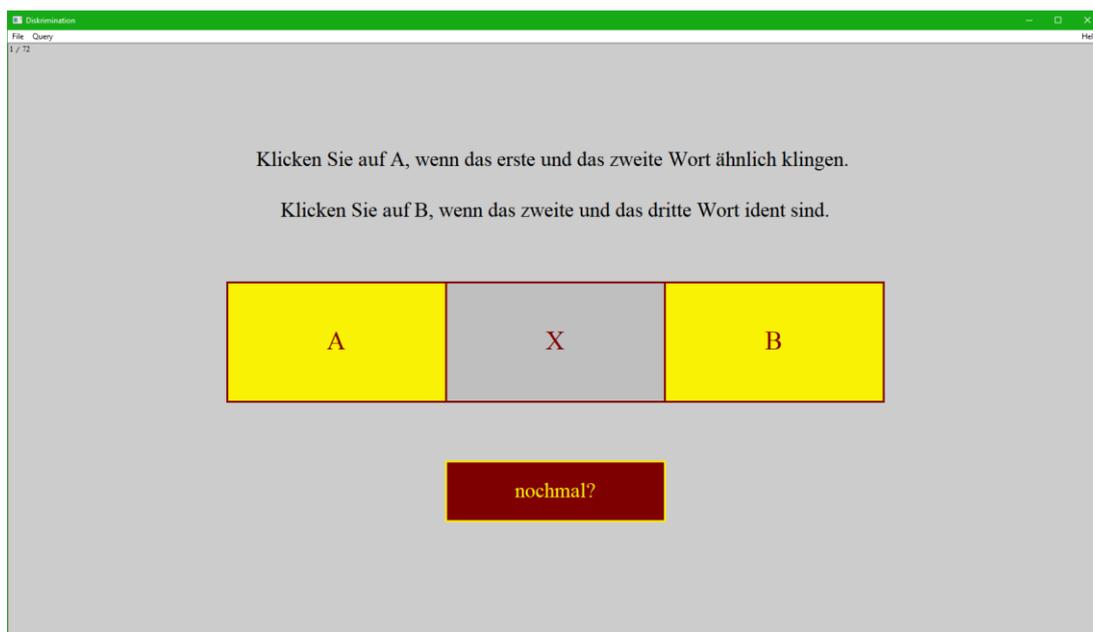


Abbildung 14: ExperimentMFC-Folie während des Diskriminationsexperiments

Es ist allerdings nur eine Wiederholung möglich. Die Reihenfolge der Trials ist randomisiert. Ob X A oder B ist, wurde mit einem digitalen zweiseitigen Würfel ausgewürfelt (= Pseudorandomisierung). Nach der Hälfte der Trials wird eine Pausenfolie angezeigt. Mit einem Klick auf das Fenster kann das Experiment fortgesetzt werden.

Nach 72 Trials folgt der zweite von insgesamt 3 Teilen. Dieselben Stimuli werden, kombiniert mit einem Speech-Shaped Noise, welches die Wahrnehmung erschweren soll, wiederum randomisiert abgespielt (mit SNR 0dB bei Teil 2 und SNR -5dB bei Teil 3). Das Experiment dauert insgesamt etwa 30 Minuten.¹⁶

6.4.4 Ergebnisse

In diesem Kapitel werden die erhobenen Daten dargestellt und die Hypothesen geprüft. Die Interpretation der Resultate erfolgt im nächsten Kapitel. Der Ablauf ist dabei wie folgt: Zuerst wird jeweils für jeden Experimentteil (ohne SSN, 0dB SNR, -5dB SNR) und für jede Gruppe (A und BC) getestet, ob die Daten normalverteilt sind, danach werden die Maße der zentralen Tendenz (Modalwert, Median und Mittelwert) und die Standardabweichung ermittelt. Anschließend werden die Daten in einem Boxplot visualisiert und Beziehungen zwischen den Daten beschrieben. Die Ergebnisse sind in Prozent angegeben. Mithilfe eines t-Tests für unabhängige Gruppen (Welch-Test, vgl. Settineri 2016: 332) werden abschließend die metrisch skalierten Daten der unterschiedlichen Gruppen miteinander verglichen. Die Berechnungen wurden – bis auf die Effektstärke – mit dem Statistikprogramm R durchgeführt. Gruppe A besteht aus 10 Personen, Gruppe BC aus 17 Personen.

Für den Experimentteil ohne SSN ergab der Shapiro-Wilk-Test für $p < 0.05$, dass die Daten von Gruppe A ($W = 0.95039$, $p = 0.6731$) und Gruppe BC ($W = 0.90976$, $p = 0.09911$) normalverteilt sind.¹⁷ Tabelle 4 zeigt die Maße der zentralen Tendenz und die Standardabweichung:

¹⁶ Die tatsächliche Dauer des Experiments ist davon abhängig, wie schnell die Proband/inn/en weiterklicken und wie oft sie ein Trial wiederholen.

¹⁷ Die Signifikanz ist größer als 0.05, das heißt die Nullhypothese, die besagt, dass Normalverteilung vorliegt, wird nicht verworfen. Ist die Signifikanz p kleiner als 0.05 liegt **keine** Normalverteilung vor.

Gruppe	A	BC	KG
Modalwert	79.17, 85.42	95.83	100
Median	85.42	95.83	100
Mittelwert	80.21	96.45	98.96
Standardabweichung (sd)	11.02	3.02	1.59

Tabelle 4: Maße der zentralen Tendenz und Standardabweichung für Teil 1 (ohne SSN)

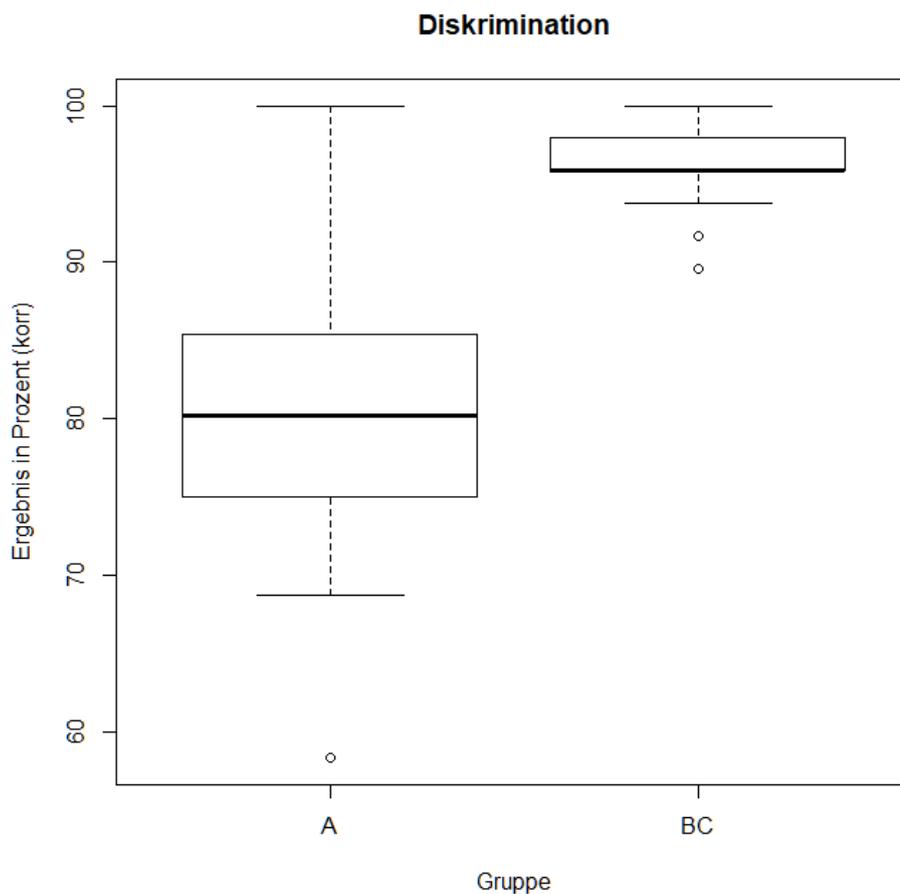


Abbildung 15: Diskriminationsrate für Teil 1 in Prozent korrekt (ohne SSN)

Mit Blick auf die Daten sollen nun erstmal die ersten beiden aus den phonetischen Modellen entwickelten Hypothesen geprüft werden:

H1: Hörer/innen mit L1 Japanisch haben leichte Schwierigkeiten den deutschen /i:/-/e:/-Vokalkontrast zu diskriminieren. Die Erkennungsrate für die Gruppe der japanischen Student/inn/en ohne Spracherfahrung in der L2 Deutsch ist voraussichtlich gut, aber nicht ausgezeichnet (< 90%, gemäß CG des PAM).

Wie in Tabelle 4 und im Boxplotdiagramm (Abb. 15) ersichtlich, liegt die Unterscheidungsrate für Gruppe A deutlich unter 90%, (Modalwert 79.17, 85.42, Median 85.42, Mittelwert 80.21), obgleich die Standardabweichung mit 11.02 relativ hoch ist. Das liegt unter anderem daran, dass eine Person alle Kontraste ohne Probleme unterscheiden konnte, wohingegen der Großteil der Teilnehmenden eindeutig Schwierigkeiten hatte, den Vokalkontrast zu diskriminieren. Die Kreise symbolisieren leichte Ausreißer. Die erste Hypothese kann somit bestätigt werden.

H2: Die Diskrimination der Gruppe der japanischen Student/inn/en mit Spracherfahrung ist signifikant besser als jene ohne sprachlicher Erfahrung in der L2 Deutsch und voraussichtlich annähernd auf erstsprachlichem Niveau.

In Gruppe BC liegen die Werte viel näher beieinander, die Schwankungsbreite ist nur gering ($sd = 3.02$). Die Erkennungsrate ist wie erwartet sehr gut und annähernd auf erstsprachlichem Niveau (Modalwert 95.83, Median 95.83, Mittelwert 96.45). Die Kontrollgruppe hat im Vergleich dazu durchschnittlich 98.96% erreicht (Median 100%, $sd = 1.59$). Die Ergebnisse beider Gruppen des ersten Experimentteils weisen signifikante Differenzen auf, wie der t-Test für unabhängige Gruppen für $p < 0.05$ zeigt: $t = -4.737$, $df = 9.8008$, $p = 0.000841$. Cohen's d (online berechnet, <https://statistikguru.de/rechner/cohens-d.html>) beträgt -2.010 , die Effektstärke $r = -0.709$. Es handelt sich demnach um einen großen Effekt (vgl. Settineri 2016: 336). Die zweite Hypothese kann folglich ebenfalls bestätigt werden. Gruppe A hat den /i:/- /e:/-Vokalkontrast signifikant schlechter diskriminiert als Gruppe BC.

Bevor wir uns der dritten Hypothese widmen, sollen Teil 2 (0dB SNR) und Teil 3 (-5dB SNR) des Experiments gesondert betrachtet und die beiden Gruppen (A und BC) jeweils miteinander verglichen werden. Für den Experimentteil mit 0dB SNR ergab der Shapiro-Wilk-Test für $p < 0.05$, dass die Daten von Gruppe A ($W = 0.90236$, $p = 0.2325$) normalverteilt und jene der Gruppe BC ($W = 0.86378$, $p = 0.01753$) nicht normalverteilt sind. Die Daten müssen für den t-Test allerdings nicht normalverteilt sein, da dieser robust genug ist (vgl. Albert/Marx 2016: 146). Tabelle 5 stellt die Maße der zentralen Tendenz und die Standardabweichung dar:

Gruppe	A	BC	KG
Modalwert	89.58	95.83	100
Median	87.5	95.83	100
Mittelwert	84.37	95.96	99.65
Standardabweichung (sd)	8.74	3.72	0.78

Tabelle 5: Maße der zentralen Tendenz und Standardabweichung für Teil 2 (0dB SNR)

Auch für diesen Teil liegt die Diskriminationsrate von Gruppe A unter 90%, wenn auch nur knapp (Modalwert 89.58, Median 87.5, Mittelwert 84.37). Die Standardabweichung ist mit 8.74 wiederum relativ hoch (siehe Abb. 16). Die Ergebnisse der Gruppe BC sind annähernd gleich wie im ersten Teil. Der Vergleich beider Gruppen zeigt für $p < 0.05$ wiederum einen signifikanten Unterschied ($t = -3.9828$, $df = 10.946$, $p = 0.002169$). Cohen's d beträgt -1.726 , die Effektstärke $r = -0.653$, d.h. es liegt ein großer Effekt vor.

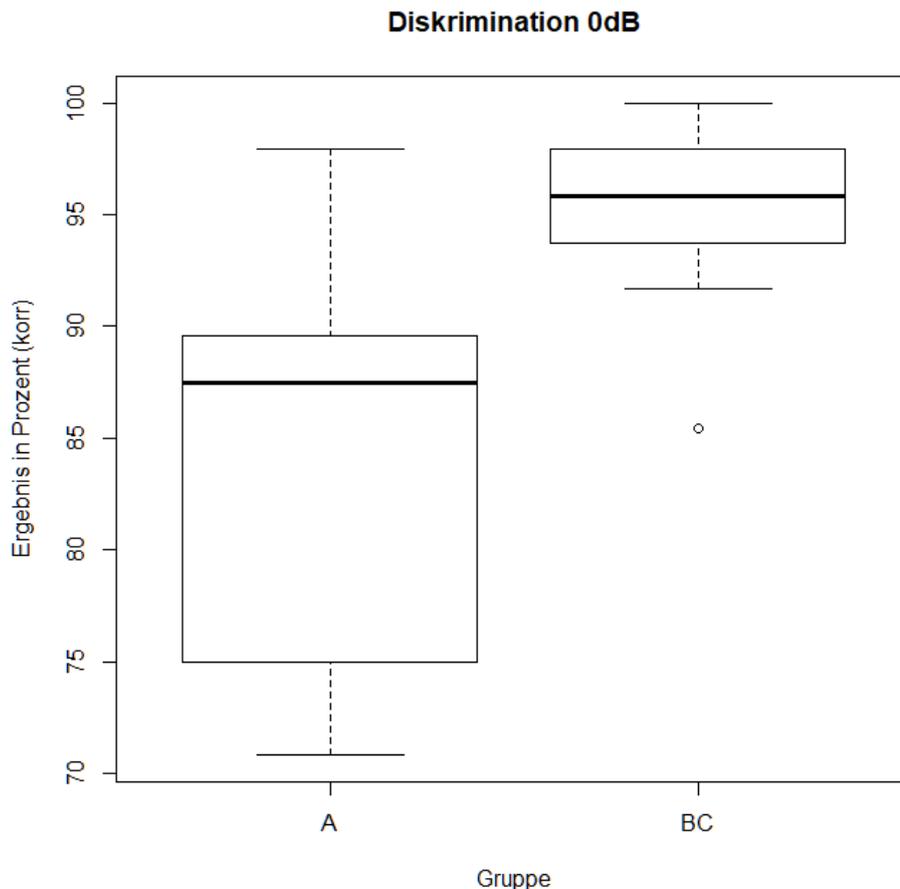


Abbildung 16: Diskriminationsrate für Teil 2 in Prozent korrekt (SNR 0dB)

Auch im dritten Teil mit -5dB SNR sind ähnliche Ergebnisse zu verzeichnen. Weder bei Gruppe A ($W = 0.74216$, $p = 0.002864$) noch bei Gruppe BC ($W = 0.85447$, $p = 0.01255$) sind die Daten normalverteilt. Die Ergebnisse der Gruppe A liegen im Mittel unter 90% (Median 87.5, Mittelwert 84.58), während die Werte der Gruppe BC nahe beisammen liegen (siehe Boxplot Abb. 17). Die Kontrollgruppe hatte erwartungsgemäß ein nahezu perfektes Ergebnis (Modalwert 100, Median 100, Mittelwert 99.48, $sd=1.24$).

Gruppe	A	BC	KG
Modalwert	83.33, 89.58, 91.67	95.83	100
Median	87.5	95.83	100
Mittelwert	84.58	95.34	99.48
Standardabweichung (sd)	11.63	3.26	1.24

Tabelle 6: Maße der zentralen Tendenz und Standardabweichung für Teil 3 (-5dB SNR)

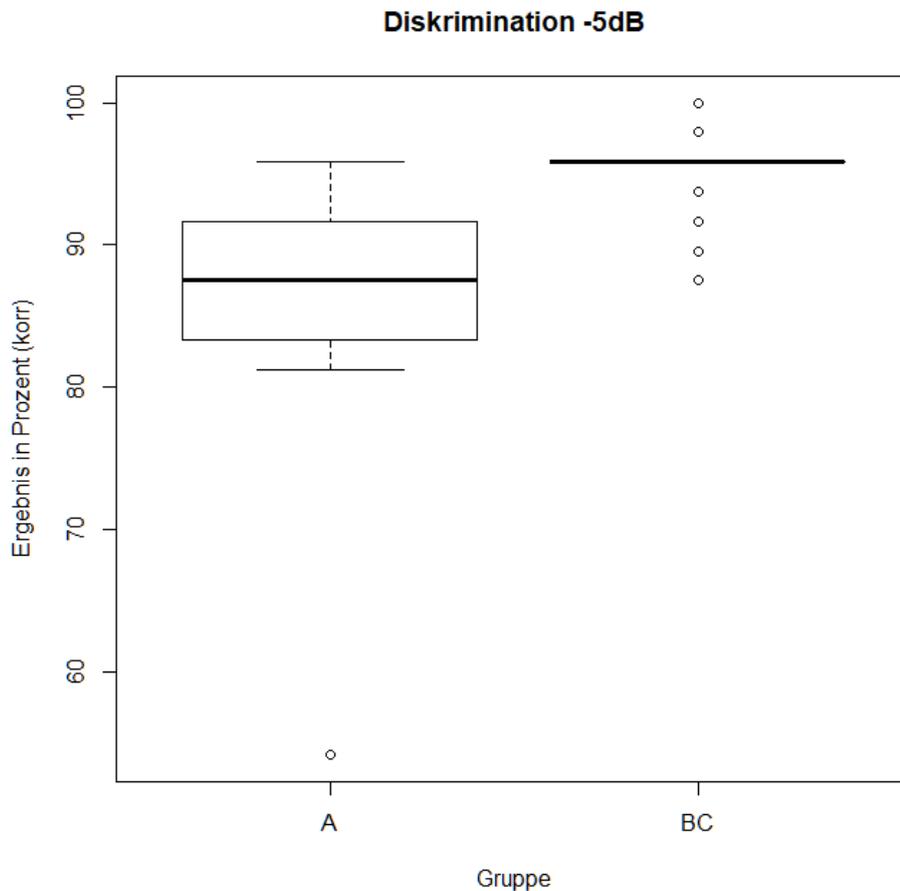


Abbildung 17: Diskriminationsrate für Teil 3 in Prozent korrekt (SNR -5dB)

Auch bei einer Störstärke von -5dB kann für $p < 0.05$ ein signifikanter Unterschied zwischen den beiden Gruppen festgestellt werden ($t = -2.8608$, $df = 9.8373$, $p = 0.0172$). Wiederum ist ein starker Effekt zu verzeichnen (Cohen's $d = -1.260$, Effektstärke $r = -0.533$). Es konnte also für alle drei Experimenteile ein signifikanter Unterschied zwischen den beiden Gruppen ermittelt werden. Nachfolgend sollen die Ergebnisse der drei Teile für jede Gruppe verglichen werden, um auf die Auswirkungen der Störstärke schließen und somit die dritte Hypothese entweder bestätigen oder verwerfen zu können.

H3: Beide Gruppen schneiden signifikant schlechter unter erschwerten Bedingungen (SNR 0dB/-5dB) ab, während die Unterscheidungsrate der Kontrollgruppe annähernd gleichbleibt.

Die beiden Boxplotdiagramme (Abb. 18 und 19) zeigen ein unerwartetes Bild. Mit steigender Störstärke wurde das Ergebnis nicht wie erwartet schlechter, sondern mitunter sogar besser. Bei Gruppe BC sind die Resultate annähernd gleichbleibend. Die Werte rücken lediglich näher zusammen (deshalb die leichten Ausreißer bei -5dB).

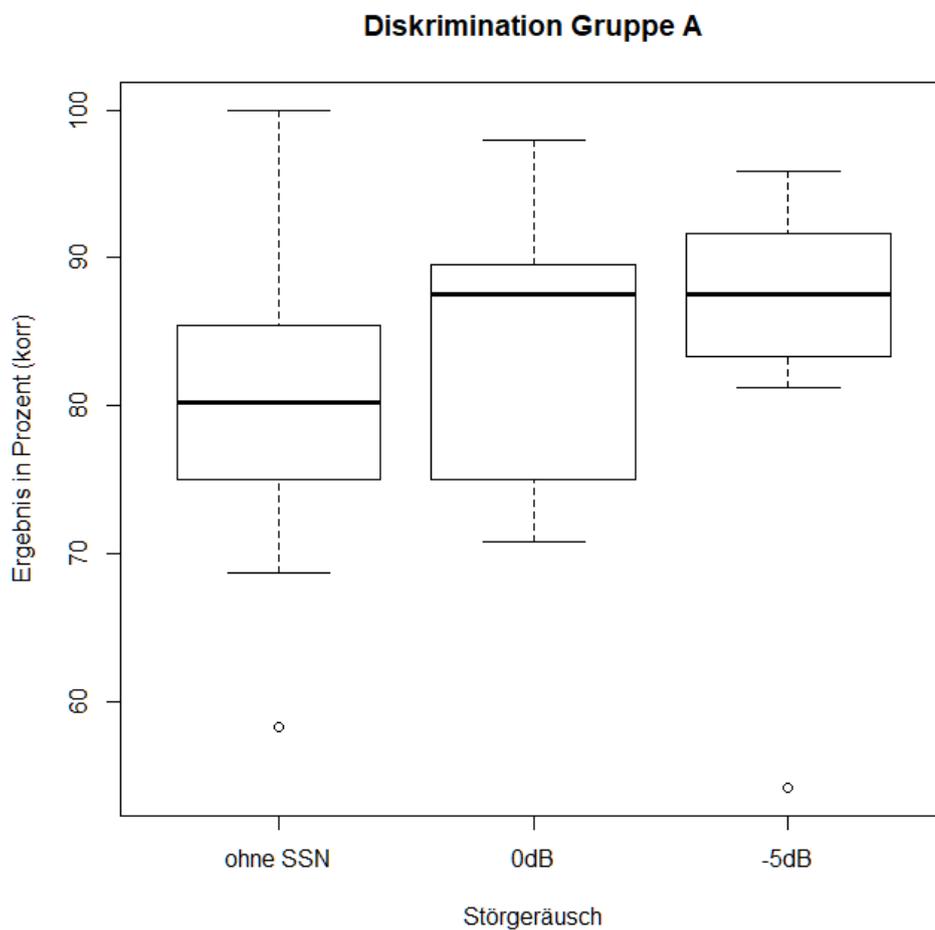


Abbildung 18: Vergleich der drei Experimentteile für Gruppe A

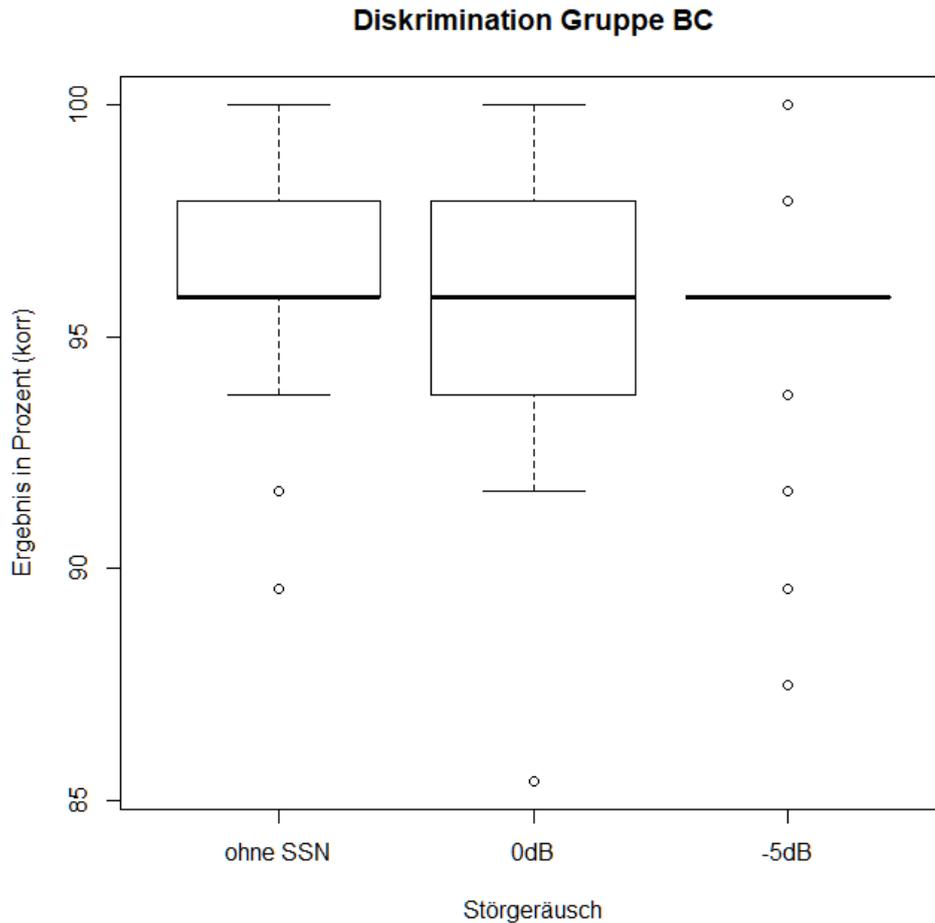


Abbildung 19: Vergleich der drei Experimenteile für Gruppe BC

Es ist daher anzunehmen, dass ein Lerneffekt stattgefunden hat, weshalb die dritte Hypothese weder bestätigt noch widerlegt werden kann. Mit zunehmendem Fortschritt erfolgte voraussichtlich eine Gewöhnung an den Experimentmodus und an die Stimuli, weshalb insgesamt ein besseres Ergebnis verzeichnet werden konnte. Eine andere bzw. zusätzliche Möglichkeit, weshalb die Ergebnisse sich mit zunehmenden Störgeräusch nicht verschlechtern, ist, dass die Störstärke auf die Unterscheidung des Vokalkontrasts nicht so große Auswirkungen wie erwartet hatte. Was die Kontrollgruppe betrifft, war das Ergebnis in allen drei Teilen wie erwartet, sehr gut (Mittelwert 98.96 / 99.65 / 99.48, Modalwert und Median jeweils 100 Prozent).

Bevor wir uns dem Identifikationsexperiment zuwenden, wollen wir die Diskrimination der Minimalpaare im Detail betrachten, um feststellen zu können, welche Minimalpaare am meisten Schwierigkeiten bereitet haben. Auf diese Weise kann indirekt auf den Einfluss der Koartikulation bzw. des Wortkontexts geschlossen werden. Es wird davon ausgegangen, dass die Diskrimination asymmetrisch, d.h. nicht alle Wortkontexte gleich schwierig sind. Abbildung 20 und 21 zeigt die Verteilung der falsch diskriminierten Minimalpaare in Prozent für die jeweilige Gruppe. Die Fehler der einzelnen Experimenteile wurden dazu addiert (für eine genauere Aufschlüsselung siehe Tabelle im Anhang G, die Liste mit den 24 Minimalpaaren ist im Anhang E zu finden).

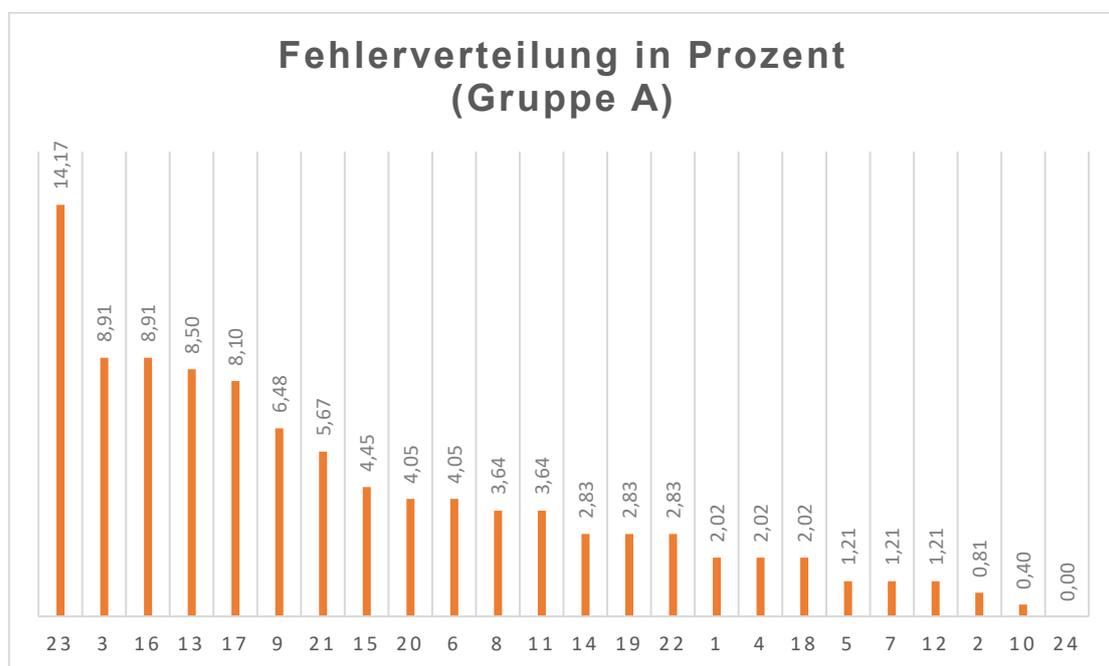


Abbildung 20: Fehlerverteilung Gruppe A

Im Vergleich der Fehlerverteilung für die einzelnen Minimalpaare innerhalb jeder Gruppe sind klare Unterschiede zu erkennen. Es zeigt sich insgesamt die leichte Tendenz, dass dreisilbige Wörter (wie *vorliegen* – *vorlegen*, *belieben* – *beleben*, *Liebende* – *Lebende*) schwieriger zu unterscheiden sind als ein- oder zweisilbige (z.B. *sie* – *See*, *stieg* – *Steg*). Anscheinend bereiten vor allem die <IVg> und <IVb>-Kontexte Probleme (z.B. *liegen* – *legen*, *liebte* – *lebte*). Im Vergleich der beiden Gruppen fällt auf, dass das Minimalpaar 13 (*Stiele* – *stehle*) in Gruppe A im Vergleich zur „erfahrenen“ Gruppe relativ oft nicht diskriminiert werden konnte (in Summe 21 Mal), während es in Gruppe BC kein einziges Mal falsch unterschieden wurde.

Im Gegensatz dazu war das Minimalpaar 1 (*lieben – leben*) in Gruppe BC scheinbar geringfügig problematischer als in Gruppe A (5 vs. 8 Mal in absoluten Werten) was aber vernachlässigt werden kann.

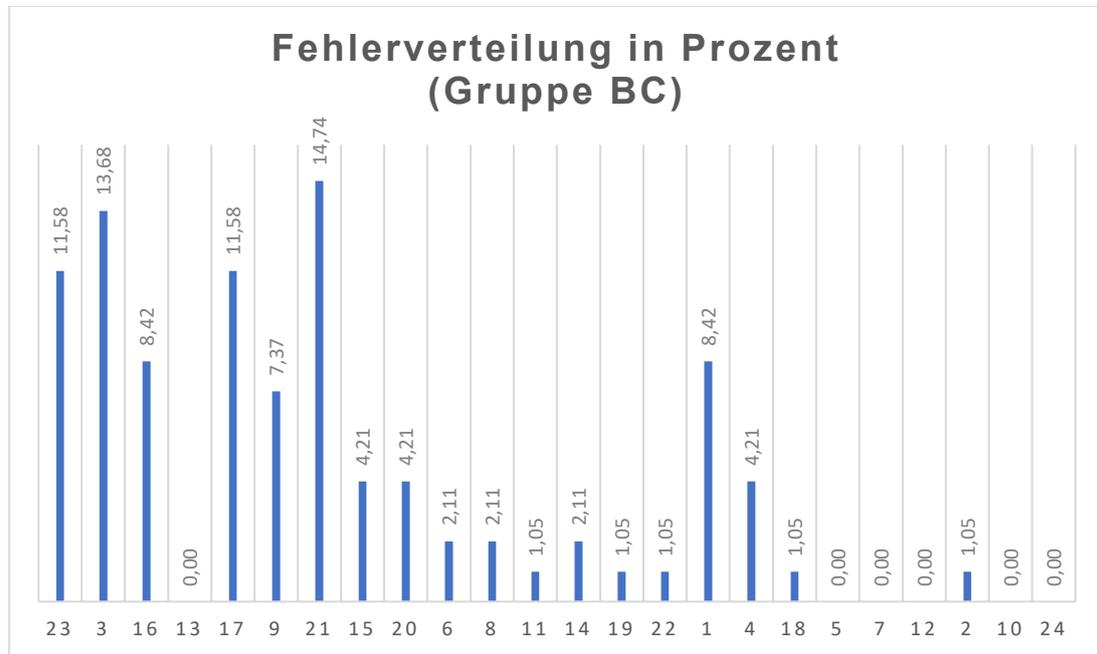


Abbildung 21: Fehlerverteilung Gruppe BC

Angemerkt werden soll an dieser Stelle, dass die Ergebnisse durch den verwendeten Experimentmodus nicht in Stein gemeißelt sind, sondern als relative Tendenzen verstanden werden sollen. Die tatsächliche Fehlerquote ist voraussichtlich höher. Außerdem täuscht die Verteilung in Prozent über die Tatsache hinweg, dass die Teilnehmenden von Gruppe A wesentlich mehr Fehler als jene von Gruppe BC gemacht haben und die Gruppengrößen auch nicht gleich groß sind (10 vs. 17 Teilnehmer/innen). Die beiden Gruppen können deshalb nur bedingt miteinander verglichen werden. Für eindeutige Resultate sollten die einzelnen Kontexte nochmal und für sich mit mehreren Wiederholungen getestet werden, um gesicherte Ergebnisse für die tatsächliche Schwierigkeit des jeweiligen Kontextes zu erlangen. Dies war aber in diesem Experiment nicht das Ziel und auch nicht möglich, da es sonst zu lange gedauert hätte und damit einhergehend unweigerlich Ermüdungseffekte eingetreten wären.

6.5 Experiment 2 – ABC der Familiennamen – Identifikation

Ergänzend zum Diskriminationsexperiment soll in einem weiteren Experiment die Identifikation der beiden Zielvokale untersucht werden, da aufgrund eigener Beobachtung im Gespräch mit japanischen Deutschlernenden und den bereits mehrfach erwähnten Testergebnissen von Ito (2010) vermutet wird, dass diese mehr Probleme als die bloße Unterscheidung bereitet (vgl. Mazuka 2014: 195). Anders als in diesem Test konnten aufgrund des Experimentdesigns allerdings nur japanische Deutschlernende (und keine naiven Hörer/innen) getestet werden, da die Kenntnis der Aussprache des deutschen Alphabets vorausgesetzt wird. Es wird davon ausgegangen, dass im Experiment das deutsche /e:/ vielfach als /i:/ gehört und notiert wird.

6.5.1 Proband/inn/en

Insgesamt wurden 16 Studierende (7 weiblich, 9 männlich) der Keio-Universität getestet, die Jus und Politikwissenschaften studieren und zusätzlich an der Universität Deutsch lernen. Es handelt sich um Deutschlernende mit ungefähr 90-270 Stunden Deutschlernerfahrung, die in drei Niveau-Gruppen geteilt wurden (1-3) von Gruppe 1 (Grundstufe mit ca. 90 Stunden Unterricht) bis Gruppe 2 und 3 (Mittelstufe und Oberstufe ca. 270 Stunden Unterricht). Die Teilnehmenden nahmen an einem Ferienintensivkurs teil und waren alle von demselben Lehrer unterrichtet worden. Sie wurden im Zuge des Kurses gemeinsam getestet.

6.5.2 Stimuli

Als Stimuli wurde vom selben Sprecher wie beim Diskriminationsexperiment das deutsche Alphabet gesprochen. Das heißt, die Stimuli bzw. einzelnen Laute werden isoliert präsentiert. Diese Buchstabenlaute wurden dann im Experiment zu insgesamt 25 deutschen bzw. österreichischen Familiennamen zusammengesetzt, welche aus einem online Telefonbuch stammen (<https://www.telefonabc.at/>). Bei der Auswahl wurde darauf geachtet, eher ungewöhnliche Namen zu verwenden, wo davon ausgegangen werden kann, dass die Proband/inn/en sie nicht kennen und die sowohl mit <e> als mit <i> „sinnvoll“ klingen (z.B.

Kerner – Kirner). Folgende Nachnamen wurden verwendet (in alphabetischer Reihenfolge): Allesch, Binder, Birnstingl, Beganovic, Drechsel, Eberl, Gebhart, Girstmair, Grasböck, Hemetsberger, Jedlicka, Kerner, Linzner, Maritsch, Metzler, Nemecek, Olbrich, Ölinger, Pelzmann, Perner, Rettensteiner, Steger, Trenker, Zehrfuchs und Zechner. Das Interstimulus-Intervall (ISI) beträgt 0,4 Sekunden.

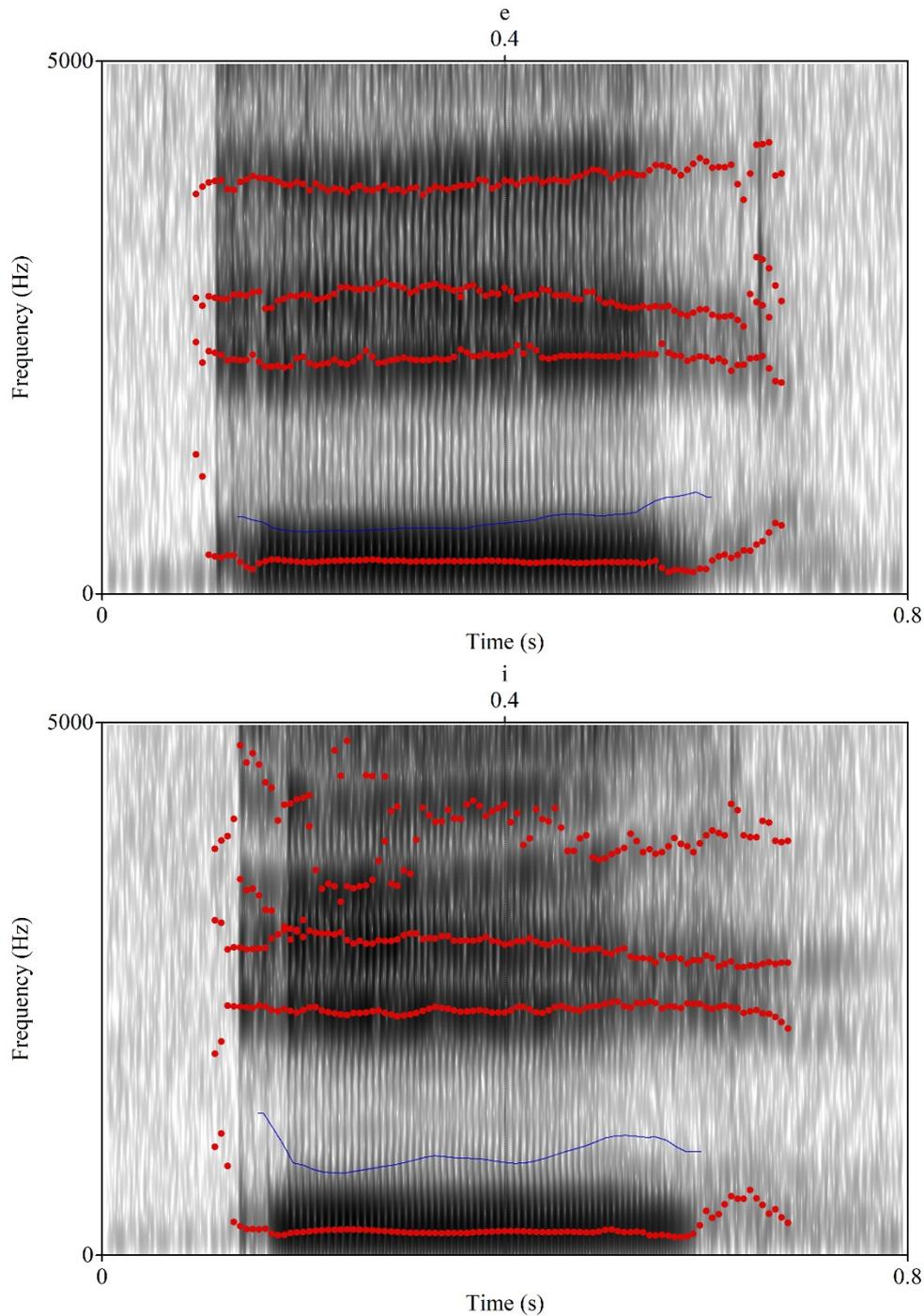


Abbildung 22: Spektrogramm von /e:/ (oben) und /i:/ (unten) in Isolation

Der direkte grafische Vergleich des isolierten /e:/- und /i:/-Vokals des eingesprochenen Alphabets zeigt (siehe Abb. 22), dass vor allem die Grundfrequenz F0 (durchgezogene Linie), also die Tonhöhe (auf engl. pitch) und der vierte Formant im Spektrogramm einen anderen Verlauf nimmt, während sich bei Betrachtung der ersten drei Formanten (Punkte) kein offensichtlicher Unterschied offenbart. Dabei handelt es sich um ein sprachuniversales Phänomen, das in der Literatur auch als *Intrinsic Pitch* bekannt ist. Demnach zeigen sich abhängig von der Vokalqualität bzw. von der Vokalhöhe unterschiedliche Grundfrequenzen (vgl. Kerschhofer-Puhalo 2014: 190). In diesem Fall wird ein /i:/ tendenziell „höher“ gesprochen als ein /e:/ (in obigen Aufnahmen durchschnittlich 155 Hz vs. 134 Hz). Eventuell spielt für die Identifikation des Vokalkontrasts also auch die Wahrnehmung der Tonhöhe als Unterscheidungsmerkmal eine nicht unwesentliche Rolle.

6.5.3 Ablauf

Die Teilnehmenden werden vorab auf Deutsch und Japanisch über den Ablauf informiert. 25 buchstabiert vorgelesene österreichische Familiennamen sollen (handschriftlich) notiert werden. Die Namen werden jeweils einmal wiederholt und in zufälliger Reihenfolge abgespielt. Alle Teilnehmenden werden gleichzeitig getestet. Das Experiment wurde auf einem Laptop gestartet, über einen CD-Player abgespielt und von mir gesteuert. Sobald auf „OK“ geklickt wird, startet automatisch der nächste Stimulus. Abbildung 23 zeigt eine ExperimentMFC-Folie während des Experiments.

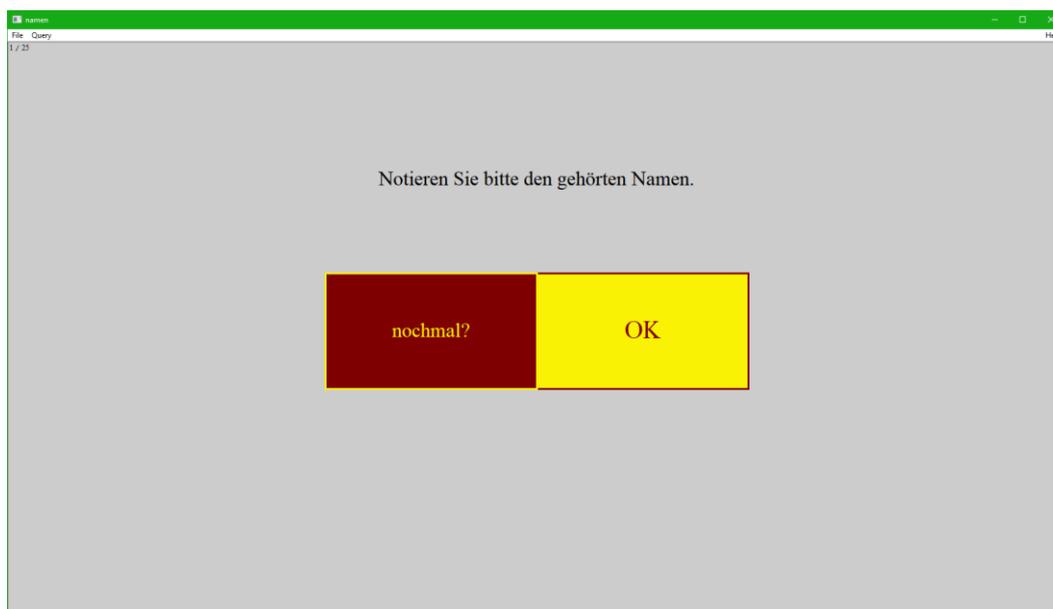


Abbildung 23: ExperimentMFC-Folie während des Identifikationsexperiments

6.5.4 Ergebnisse

Ausgewertet wurden folgende Substituierungsprozesse: <e> als <i>, <i> als <e> und <e> als <a> wahrgenommen bzw. notiert (siehe Tabelle 7). Erhoben wurde die Anzahl der richtig geschriebenen Familiennamen (hinsichtlich der darin vorkommenden Vokale). Die Anzahl der maximal erreichbaren Punkte ist 25. Es wurden nur klare Fehler als solche gewertet. Auslassungen wurden beispielsweise als richtig gewertet.

	<i>e_als_i</i>	<i>i_als_e</i>	<i>e_als_a</i>
<i>Gruppe 1</i>	25	25	23
	25	17	14
	25	22	23
	25	24	23
	25	22	25
<i>Gruppe 2</i>	25	24	25
	25	23	25
	25	25	22
	25	21	19
	24	19	24
<i>Gruppe 3</i>	25	16	24
	25	25	25
	25	24	25
	25	24	25
	25	24	25
<i>Mittelwert</i>	24,94	22,50	23,19
<i>Median</i>	25	24	24
<i>MW_Prozent</i>	99,75	90	92,75

Tabelle 7: Ergebnisse des Identifikationsexperiments (in absoluten Werten)

Wider Erwarten wurde der Laut /e/:

/ von allen drei Gruppen lediglich einmal als <i> verschriftlicht (durchschnittlich zu 99.75%), während umgekehrt ein /i:/ relativ oft als <e> notiert wurde (im Durchschnitt 90%), was vermutlich auf den Einfluss des Englischen als erstgelernte Fremdsprache zurückzuführen ist (ebenso die Verschriftlichung von /e:/ als <a> zu 92.75%).

7 Zusammenfassung und Diskussion der Ergebnisse

In diesem Kapitel sollen die Ergebnisse beider Experimente interpretiert und mit der Theorie in Beziehung gesetzt werden. Gleichzeitig handelt es sich um eine Zusammenfassung der im Theorieteil gewonnenen Erkenntnisse (Kap. 2-5). Auf Basis der gesammelten Daten sollen Schlüsse für die fremdsprachliche Vokalwahrnehmung gezogen, die Forschungsfragen soweit als möglich beantwortet und die eigene Forschung kritisch betrachtet werden. Besonderes Augenmerk soll außerdem der Frage geschenkt werden, warum die Ergebnisse des Experiments sich erheblich von den in „natürlicher“ Umgebung gemachten eigenen Beobachtungen unterscheiden. Die Unterschiede beider Kontexte (Experiment vs. natürlicher Umgebung) sollen dazu herausgearbeitet werden. Durch die Verknüpfung von Theorie und Empirie soll ein Erkenntnisgewinn erzielt werden.

Einmal mehr sollen dazu die zuvor aufgestellten Hypothesen angeführt werden:

H1: Hörer/innen mit L1 Japanisch haben leichte Schwierigkeiten den deutschen /i:/-/e:/-Vokalkontrast zu diskriminieren. Die Erkennungsrate für die Gruppe der japanischen Student/inn/en ohne Spracherfahrung in der L2 Deutsch ist voraussichtlich gut, aber nicht ausgezeichnet (< 90%, gemäß CG des PAM).

Die erste Hypothese, welche auf Basis der Vorhersage des PAM entwickelt wurde, konnte im Diskriminationstest bestätigt werden. Die Gruppe der japanischen Studierenden ohne sprachlicher Erfahrung in der L2 Deutsch erreichte im ersten Teil des Experiments durchschnittlich 80.21 %, die hohe Standardabweichung lässt allerdings auf eine heterogene Verteilung der individuellen Diskriminationsraten schließen.

Dieses Ergebnis widerspricht dem Resultat der Studie von Mazuka et al. (2013), worin für den /bi:k/-/be:k/-Kontrast bei monolingualen japanischen Erwachsenen eine sehr gute Diskriminationsrate von 92% erzielt worden ist. Die Autor/inn/en haben für diesen Kontrast den Entwicklungspfad *improvement without exposure* festgestellt (also Verbesserung der Diskrimination ohne Kontakt mit der Zielsprache). In der vorliegenden Untersuchung sind jedoch verschiedene Kontexte gemischt, weshalb die beiden Experimente nur bedingt miteinander verglichen werden können. Die Inkongruenz der Ergebnisse könnte im Einfluss der Koartikulation begründet sein, der etwas später noch näher besprochen werden soll.

Ein Experiment von Polka (1995) hat gezeigt, dass die Diskrimination von deutschen Vokalkontrasten bei „naiven“ Hörer/inn/en ohne Deutschkenntnisse nicht immer und für alle Kontrastpaare gleich gut (oder schlecht) ist. Untersucht wurde die Unterscheidung des deutschen /y/-/u/ und /ɔ/-/Y/-Vokalkontrastes in einem dVt-Kontext durch monolinguale Englischsprechende. Die gespannten Vokale (/y/-/u/) konnten von den Hörer/inne/n mit L1 Englisch sehr gut unterschieden werden, während der ungespannte Vokalkontrast /ɔ/-/Y/) durchaus Probleme bereitet hat (vgl. Polka/Werker 1995: 1292).

Grundsätzlich kann davon ausgegangen werden, dass ein Kontrast umso schwieriger ist, je ähnlicher ein Laut der L2 zu einem Laut der L1 ist. Wird ein zweitsprachlicher Laut nicht als verschieden zu einem Laut der Erstsprache erkannt, kann es zu Substitutionen kommen, wie das etwa bei der Analyse der Konfusionsdaten von Kerschhofer-Puhalo (2014: vgl. 592) der Fall war, wo das /e:/ vielfach durch ein /i:/ ersetzt worden ist. Auch die intralinguale Ähnlichkeit innerhalb der Erst- bzw. Zweitsprache spielt mitunter eine nicht zu unterschätzende Rolle. Mit Blick auf die im Vokaltrapez veranschaulichte Nähe im Vokalraum (siehe Abb. 24) und die vorangegangenen vergleichenden artikulatorischen Beschreibungen der Zielvokale im Theorierteil kann für den /i:/-/e:/-Vokalkontrast eine große Ähnlichkeit angenommen werden.

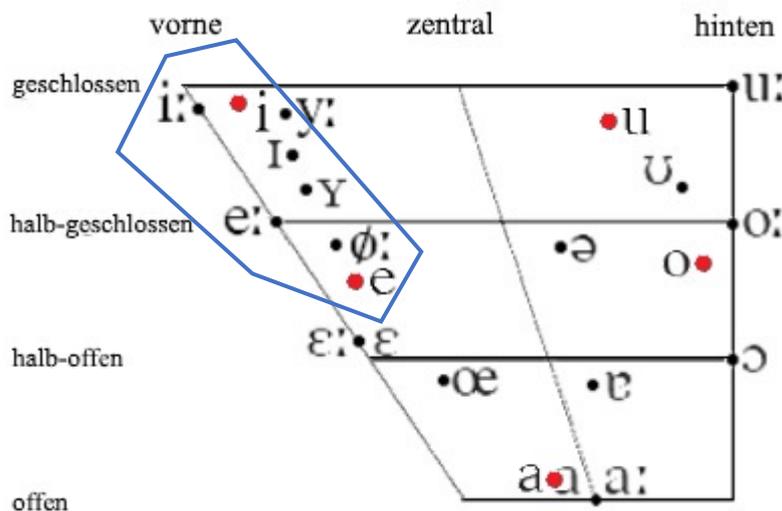


Abbildung 24: Vokaltrapez Deutsch (deutsche Vokale sind in schwarzer, japanische in roter Farbe)

Für die Gruppe der japanischen Studierenden ohne Deutschkenntnisse wurde aufgrund der inter- und intralingualen Ähnlichkeit der zwei Zielvokale in beiden Sprachen und der Resultate von vorhergehenden Studien (vgl. Mazuka et al. 2013) auf einen Category Goodness Difference (CG) geschlossen, wonach die Diskrimination durch naive Hörer/innen relativ gut sein sollte.

Aufgrund der Vorhersagen des SLM bzw. PAM-L2 wurde weiters angenommen, dass die Diskrimination anfangs aufgrund der Äquivalenzqualifikation schlechter sein sollte, mit zunehmender sprachlicher Erfahrung sollte die phonetische erstsprachliche Kategorie für den unähnlicheren Laut aber soweit modifiziert werden, dass er problemlos oder zumindest besser differenziert werden kann. Die zweite Hypothese lautete demnach:

H2: Die Diskrimination der Gruppe der japanischen Student/inn/en mit Spracherfahrung ist signifikant besser als jene ohne sprachlicher Erfahrung in der L2 Deutsch und voraussichtlich annähernd auf erstsprachlichem Niveau.

Gruppe BC schnitt in allen drei Teilen signifikant besser als Gruppe A ab. Also konnte auch diese Vorhersage bestätigt werden. Beide Gruppen unterscheiden sich vor allem hinsichtlich der sprachlichen Erfahrung in der L2 Deutsch. Bis auf diesen Punkt waren die Teilnehmenden beider Gruppen in Bezug auf die Sprachbiografie relativ homogen. Alle Proband/inn/en haben für mindestens sechs Jahre Englisch in der Schule gelernt, vereinzelt waren auch andere Fremdsprachen wie Französisch, Spanisch, Italienisch und Chinesisch, Koreanisch in der Schule oder an der Universität gelernt worden, der überwiegende Teil der Proband/inn/en hatte aber nur Englisch- und (bei Gruppe BC) Deutschunterricht. Es kann also davon ausgegangen werden, dass der Faktor Spracherfahrung für diesen Kontrast eine entscheidende Rolle spielt.

Um auf den Einfluss des Störgeräusches schließen zu können, wurden die drei Teile gesondert für jede Gruppe verglichen. Es wurde aufgrund der Studie von Cutler (2004) davon ausgegangen, dass beide Gruppen im Vergleich zur Kontrollgruppe unter erschwerten Bedingungen ein signifikant schlechteres Ergebnis erreichen:

H3: Beide Gruppen schneiden signifikant schlechter unter erschwerten Bedingungen (SNR 0dB/-5dB) ab, während die Unterscheidungsrate der Kontrollgruppe annähernd gleichbleibt.

Dies war allerdings nicht der Fall. Die Diskrimination wurde mit steigender Störstärke mitunter sogar besser, anstatt schlechter zu werden, weshalb vermutet wird, dass durch die Wiederholung der Stimuli und der Einstellung auf die Hörsituation ein Lerneffekt stattgefunden hat. Die Teilnehmenden sowohl von Gruppe A als auch von BC haben sich vermutlich nach einer Weile auf das Hören des Vokalunterschieds spezialisiert, wodurch die Unterscheidung auch mit Störgeräusch keine größeren Schwierigkeiten bereitet hat. Der Einfluss der Störstärke war dabei offensichtlich nicht so stark wie angenommen. Die Unterscheidungsrate der Kontrollgruppe war wie erwartet durchgehend ausgezeichnet, die Unterscheidung stellte in

keinem der drei Teile ein Problem dar. Die dritte Hypothese konnte somit also weder bewiesen noch widerlegt werden.

Womit wir ohne weitere Umschweife und um an die Einleitung anzuschließen zur Beantwortung der Forschungsfragen kommen wollen, welche an dieser Stelle nochmals in Erinnerung gerufen werden sollen:

1. Können japanische L1-Hörer/innen (mit und ohne DaF-Erfahrung) den deutschen /i:/- /e:/-Vokalkontrast – unter den durch das Experiment vorgegebenen Bedingungen – unterscheiden? Wenn ja, wie gut?
2. Welche Faktoren/Variablen haben Einfluss auf die fremdsprachliche Wahrnehmung der beiden Vokale? Spielt die sprachliche Erfahrung in der Zweitsprache eine Rolle?
3. (Inwiefern) verändert sich das Ergebnis unter erschwerten Bedingungen (SNR 0dB/-5dB) im Vergleich zu deutschen L1-Hörer/innen?

Die erste Frage kann für Gruppe BC mit einem klaren Ja beantwortet werden. Die Ergebnisse der „erfahrenen“ Gruppe sind im Vergleich zur Kontrollgruppe mit L1 Deutsch nur geringfügig schlechter. Die Proband/inn/en hatten in Übereinstimmung mit den Vorhersagen der phonetischen Modelle keinerlei Probleme die beiden Vokale zu unterscheiden. Auch die Resultate der Gruppe A sind mit durchschnittlich >80% in allen drei Teilen relativ gut, wobei für diese Gruppe zumindest kleinere Schwierigkeiten erkennbar waren. Die Gruppe mit DaF-Erfahrung hat ein signifikant besseres Ergebnis erzielt als die Gruppe ohne Deutschkenntnisse.

Damit unterscheiden sich die Resultate eindeutig von den in „natürlicher Umgebung“ gemachten Beobachtungen. Selbst im Gespräch mit fortgeschrittenen japanischen Deutschlernenden konnten vielfach Verwechslungen von /e:/ mit /i:/ festgestellt werden, während dieselben Personen im Diskriminationsexperiment ein sehr gutes bis ausgezeichnetes Ergebnis erreicht haben. Warum sich Experiment und Beobachtung so gravierend unterscheiden, soll am Ende dieses Kapitels und im Resümee noch diskutiert werden. Zuvor soll aber auf die zweite Forschungsfrage eingegangen werden und wesentliche Variablen für die fremdsprachliche Wahrnehmung von Vokalen identifiziert und diskutiert werden.

Beginnen wir zuerst mit der Ja-/Nein-Frage „Spielt die sprachliche Erfahrung in der Zweitsprache eine Rolle?“. Diese Frage kann mit Blick auf die erhobenen Daten ebenfalls wie die erste Frage mit einem klaren „Ja“ beantwortet werden. Das AXB-Diskriminationsexperiment hat gezeigt, dass zwischen den Ergebnissen beider Gruppen japanischer L1-Hörer/inne/n ein signifikanter Unterschied besteht – und zwar für alle drei Teile bzw. Störstärken. Somit kann für die Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts durch Hörer/innen mit L1 Japanisch auf einen entscheidenden Einfluss sprachlicher Erfahrung geschlossen werden – in dieser Arbeit repräsentiert durch mehrjährigen Kontakt mit der L2 bzw. durch Unterricht in der Fremdsprache Deutsch im universitären Kontext in Japan. Die sprachliche Kompetenz der Gruppe BC in der L2 Deutsch spiegelt sich indirekt in den universitären „Niveaunklassen“ wider, worin die Lernenden eingeteilt worden sind.

Das Resultat des Diskriminationsexperiments reiht sich in eine lange Reihe von Untersuchungen ein, die den Einfluss sprachlicher Erfahrung auf die fremdsprachliche Wahrnehmung aufzeigen und belegen (vgl. beispielsweise Ingram/Park 1997, Bohn 1998, Flege 2007, Levy/Strange 2008), wobei einmal mehr angemerkt werden soll und muss, dass der Faktor Spracherfahrung nur schwer quantifizierbar ist, weshalb Studien zu dieser Thematik oft nur bedingt vergleichbar sind. Eine möglichst genaue Definition, die deutlich macht, was unter Spracherfahrung verstanden wird, ist deshalb umso wichtiger.

Wie bereits im theoretischen Teil angesprochen, nimmt die Variable Alter mitunter indirekt Einfluss, sie ist aber sicherlich nicht der alles entscheidende Faktor. Vielmehr kommt es vor allem auf die Qualität und Quantität des sprachlichen Inputs an, weshalb das Konzept sprachliche Erfahrung in dieser Arbeit sowohl eine bestimmte Lern- bzw. Kontaktdauer mit und eine gewisse Sprachkompetenz in der L2 Deutsch einschließt: Die Proband/inn/en der Gruppe BC haben für mindestens zwei Jahre ausschließlich im Herkunftsland Japan an der Universität bzw. davor in der Schule Deutsch gelernt. Der Unterricht erfolgte zu einem großen Teil durch erstsprachliche Lehrende, womit zumindest zu einem gewissen Grad sowohl die Quantität und Qualität des sprachlichen Inputs gewährleistet sein sollte.

Phonetische Modelle, die sprachliche Erfahrung berücksichtigen, sind Fleges SLM und das PAM-L2 von Best und Tyler. Ob für einen L2-Laut eine neue phonetische Kategorie gebildet wird oder nicht, ist nach dem SLM von der Äquivalenzklassifikation und somit vom bereits angesprochenen Faktor Ähnlichkeit abhängig. Diese Variable hängt wiederum zusammen mit dem nativen Phoneminventar, genauer gesagt mit dem wechselseitigen Verhältnis zwischen den Phonen der Erstsprache und jenen der Zweitsprache (→ CLI). Das SLM geht davon aus, dass anfängliche fremdsprachliche Wahrnehmungsschwierigkeiten abhängig vom jeweiligen

L2-Lautkontrast und der L1 überwunden werden können. In unserer Untersuchung sind wir gemäß dem PAM-L2 davon ausgegangen, dass für den CG Difference mit zunehmender sprachlicher Erfahrung einer der beiden Laute ([i:]) einer L1-Kategorie zugeordnet wird, während für den anderen ([e:]) voraussichtlich eine neue phonetische L2-Kategorie gebildet wird.

Damit sind wir schon bei der zentralen zweiten Frage angelangt, welche Variablen wesentlichen Einfluss auf die fremdsprachliche Vokalwahrnehmung nehmen. Neben der gerade thematisierten Wirkung des Cross-linguistic Influence und der inter- und intralingualen Ähnlichkeit, welche indirekt den Einfluss des nativen Phoneminventars, der Zweitsprache und vokalimmanente Merkmale des betreffenden Lautkontrasts (*acoustic cues*) widerspiegeln, wurden als weitere bestimmende Faktoren bereits weiter oben angesprochen individuelle Lernendenvariablen wie Spracherfahrung und Alter und methodologische Faktoren wie etwa die Aufmerksamkeit angeführt, auf die an dieser Stelle nochmal genauer eingegangen werden soll.

Aufmerksamkeit steuert die Wahrnehmung von Sprache, welche von Winifred Strange als sinnsuchende Aktivität beschrieben wird (vgl. Strange 2011: 457). Dem Automatic Selective Perception (ASP) Modell (vgl. Strange 2011) zufolge kann zwischen zwei Wahrnehmungsmodi unterschieden werden – einem phonologischen und einem phonetischen Modus. Zur Erinnerung: Der phonologische Modus kommt vor allem bei der Wahrnehmung erstsprachlicher Äußerungen zum Einsatz, während der phonetische Modus bei unbekanntem Dialekten oder in der Fremd- bzw. Zweitsprache greift. Er zeichnet sich durch den Fokus auf phonetische und akustische Merkmale aus und sollte dem Modell zufolge bei Diskriminationsexperimenten mit kurzem Interstimulusintervall (ISI, 0-750ms) – wie bei unserer Untersuchung – für gute Unterscheidungsraten sorgen (vgl. Kerschhofer-Puhalo 2014: 78f). Ursache für das gute Ergebnis sind „Spuren“ (traces) im Kurzzeitgedächtnis, welche für die Diskrimination nutzbar gemacht werden können. Der phonologische Modus ist im Gegensatz dazu weitgehend automatisiert und durch perzeptuelle Routinen (selective perception routines, SPR) gekennzeichnet. Bei längerer ISI werden für die Diskrimination hingegen die Repräsentationen lautlicher Kategorien im Langzeitgedächtnis benötigt (vgl. Kerschhofer-Puhalo 2014: 114). Die Unterscheidung zwischen phonologischen und phonetischen Modus kann als erster Anhaltspunkt für die Diskrepanz zwischen Experiment und Wirklichkeit dienen.

Als weitere wesentliche Einflussgröße im Diskriminationsexperiment konnte zudem die Koartikulation (vgl. Höhle 2010: 43) bestimmt werden. Der den Vokal umgebende Kontext hatte Einfluss auf die Schwierigkeit des jeweiligen Minimalpaares wie die Analyse der Fehlerverteilung gezeigt hat. Die Reihung nach Schwierigkeit war dabei in weiten Teilen bei beiden Gruppen übereinstimmend, wobei es auch „Ausreißer“ wie Minimalpaar 13 (*Stiele-stehle*) und 21 (*Liebende-Lebende*) gegeben hat, die von Gruppe zu Gruppe unterschiedlich gut unterschieden wurde. Längere Wörter (mit bis zu drei Silben) waren insgesamt schwieriger zu unterscheiden als einsilbige. Am schwierigsten war etwa das Minimalpaar 23 (*vorliegen-vorlegen*), während Minimalpaar 24 (*stieg-steg*) kein einziges Mal falsch diskriminiert worden ist. Besonders schwierig waren scheinbar die <IVg> und <IVb>-Kontexte (z.B. *liegen – legen, liebte – lebte*).

Die dritte Forschungsfrage „(Inwiefern) verändert sich das Ergebnis unter erschwerten Bedingungen (SNR 0dB/-5dB) im Vergleich zu deutschen L1-Hörer/innen?“ kann aufgrund des bereits thematisierten Lerneffekts leider nicht zufriedenstellend beantwortet werden (siehe weiter oben). Stattdessen muss die Frage gestellt werden, wie das Experiment gestaltet werden müsste, um die aufgetretenen Lerneffekte zu vermeiden (mehr dazu in der Schlussbetrachtung). Die Ausführungen zur Aufmerksamkeit haben deutlich gemacht, dass der verwendete Experimentmodus selbst Einfluss auf die Wahrnehmung und somit auf die erhobenen Daten nimmt, wie auch Nadja Kerschhofer-Puhalo betont:

Usually, results from discrimination tasks reflect the auditory component rather than or in addition to long-term memory representations of L2 categories. Typically, vowels are inherently more discriminable than consonants [...]. Of course, the discriminability (related to auditory short term memory) of phones might be better than the ability to correctly identify L2 sounds (related to representations in long term memory) (Kerschhofer-Puhalo 2014: 78).

Im Theorieteil (Kap. 2.5) wurde die Wahrnehmung von Vokalen als „eher kontinuierlich“ bestimmt, hauptsächlich deshalb, weil Selbstlaute innerkategorial meist besser unterschieden werden als Konsonanten. Jedoch werden Vokale abhängig vom Experimentdesign mitunter auch kategorial wahrgenommen – dies ist etwa bei Identifikationsexperimenten der Fall:

While in perceptual discrimination tasks [...] the *continuous mode* is activated, identification tasks [...] rather involve the *categorical mode* of perception to enable listeners to “label” input stimuli as belonging to a specific category (Kerschhofer-Puhalo 2014: 97).

Die Identifikation von Vokalen erfordert im Gegensatz zur Diskrimination eine aktive Kategorisierung des Gehörten (vgl. Kerschhofer-Puhalo 2014: 79).

Aus diesem Grund wurde ergänzend zum Diskriminationsexperiment die Identifikation des /i:/- /e:/-Vokalkontrasts getestet. Allerdings mussten aufgrund des Experimentdesigns zwei Einschränkungen in Kauf genommen werden: Zum einen wurden die Vokale in Isolation und nicht in einem Wort- oder Satzkontext präsentiert (was vor allem mit Blick auf die unterschiedlichen Normalisierungstheorien einen wesentlichen Unterschied macht, vgl. Kap. 2.5), zum anderen konnten nur japanische Deutschlernende getestet werden, da die Kenntnis der Aussprache des deutschen Alphabets vorausgesetzt wurde. Dafür musste nicht für die Benennung bzw. Kategorisierung auf eine möglicherweise problematische Verschriftung der im Experiment vorkommenden Stimuli zurückgegriffen werden (vgl. Kerschhofer-Puhalo 2014: 81). Das Ergebnis des Identifikationsexperiments deckt sich nicht mit den Erwartungen und dem Test von Ito (2010, siehe Abb. 25), der in Mazuka et al. (2013) angeführt ist und in dem japanische Erwachsene deutsche Wörter mit japanischen Kana verschriftlichen sollten. /e:/ wurde darin zu 94% als /i:/ wahrgenommen und verschriftlicht. Leider wurde in der Studie nicht angegeben, welche Wörter und welche Kontexte im Identifikationstest von den Hörer/innen mit L1 Japanisch notiert werden sollten.

Table 1. Acoustic Measurements (and Standard Deviations) of the Stimuli in Experiment 1–3, and the Relevant Results of Ito (2010), in Which Japanese Adults Were Asked to Write Down How Various German Vowels Sound Like in Japanese Scripts

Stimuli	F0 (Hz)	Duration (ms)	F1 (Hz)	F2 (Hz)	F3 (Hz)	Japanese Adults Heard As
/bu:k/	245 (3.3)	660 (37.5)	390 (32.3)	720 (43.3)	2,990 (229.0)	/u:/ 57% /o:/ 32%
/by:k/	244 (4.3)	630 (41.9)	350 (18.7)	1,870 (76.6)	2,210 (214.6)	/u:/ 72%
/bi:k/	224 (4.1)	655 (28.2)	270 (17.7)	2,620 (55.7)	3,610 (180.2)	/i:/ 98%
/be:k/	225 (3.9)	625 (68.7)	360 (32.6)	2,550 (73.2)	3,340 (32.2)	/i:/ 94%
/bo:k/	245 (2.8)	656 (34.0)	470 (11.9)	750 (22.9)	3,000 (137.2)	/o:/ 96%

Abbildung 25: Ergebnisse von Ito (rot eingekreist) (aus Mazuka et al. 2013: 196).

Überraschenderweise wurde /e:/ in meinem Identifikationstest mit japanischen Studierenden nur ein einziges Mal als „i“ verschriftlicht (durchschnittlich 0.25%), während die Substitution in die andere Richtung (/i:/ als „e“) mit durchschnittlich 10% wesentlich häufiger aufgetreten ist. Da auch /e:/ zu 7.25% als „a“ notiert wurde, kann das Ergebnis vermutlich auf einen Einfluss des Englischen zurückgeführt werden. Englisch stellt in Japan in der Regel die erstgelernte Fremdsprache dar und wird auch noch an der Universität gelehrt. Somit ist die Aussprache des englischen Alphabets vermutlich geläufiger als jene des deutschen. Aufgrund der oben angegebenen Einschränkungen sind die Resultate leider nur bedingt aussagekräftig. Wie und ob die Vokale in einem Wort- oder Satzkontext und von naiven Hörer/innen identifiziert werden können, konnte in diesem Experiment nicht ermittelt werden.

Abschließend sollen einige wesentliche Unterschiede zwischen Diskriminationsexperiment und beobachteter Wirklichkeit diskutiert und die eigene Forschung kritisch betrachtet werden. Die zentrale Frage, um das eigentliche Forschungsinteresse und die gewonnenen Ergebnisse zusammenzubringen, lautet: Warum unterscheiden sich die Ergebnisse des Experiments von der (eigenen) Beobachtung? Drei Unterschiede zwischen Experiment vs. „Realität“ sollen als mögliche Erklärung angeführt werden:

1. Diskrimination ist ungleich Identifikation
2. Natürliche Sprache ist durch einen kontinuierlichen Sprachstrom gekennzeichnet und unterscheidet sich von den im Experiment präsentierten Stimuli
3. Deshalb ist bei natürlicher Kommunikation die Aufmerksamkeit anders geregelt und damit der Wahrnehmungsmodus ein anderer als bei einem Experiment

Der Unterschied zwischen Diskrimination und Identifikation wurde bereits mehrmals angesprochen. Einen Unterschied zwischen zwei Minimalpaaren zu erkennen, fällt wesentlich leichter als einen Vokal in einem Wort oder einem Satz zu identifizieren, da das Gehörte bei dem AXB-Diskriminationsexperiment für eine Weile im Kurzzeitgedächtnis verbleibt und (bei kleinem ISI) für die Unterscheidung nutzbar gemacht werden kann.

In natürlicher Kommunikation will der Hörer/die Hörerin hingegen aus einem kontinuierlichen Sprachstrom Informationen herausfiltern, der Zweck ist dabei ein gänzlich anderer als bei einem Experiment, wo die Stimuli mehr oder minder isoliert präsentiert werden. Während natürliche Sprachwahrnehmung in der Regel der Informationsentnahme dient und als top-down-Prozess strukturiert ist, legt ein Wahrnehmungsexperiment den Fokus automatisch auf die sprachliche Form und ist als bottom-up-Prozess organisiert:

Speech perception can be characterized as an active process of purposeful information-seeking. Natural language processing is a process that proceeds mainly in a *top-down*-direction, i.e. listeners are driven by their expectations about the content of the linguistic message and the speaker's intentions, seeking sufficient information from the continuous speech signal to identify the string of words in an utterance and to decode the message encoded by the speaker. In perceptual experiments, however, listeners are mainly involved in *bottom-up*-procedures that differ from processes of online processing of continuous speech input. Therefore, experimental results and their relevance for "real-world" speech processing in L2 have to be (re)viewed very critically (Kerschhofer-Puhalo 2014: 109).

Aus diesem Grund ist auch die Aufmerksamkeit bzw. der Fokus auf das Gehörte anders strukturiert. Wörter sind bei mündlicher Sprache normalerweise nicht durch Pausen getrennt und durch die Performanz des Sprechens sind die lautlichen Realisierungen jedes Mal anders. Bei phonetischen Experimenten hingegen sind die Stimuli meist abgegrenzt und der Kontext

kontrolliert. Die Fülle der Einflussfaktoren wird im Experiment bei weitem nicht erreicht. Die Relevanz von Wahrnehmungsexperimenten für die Beschreibung natürlicher Sprachverarbeitung muss daher kritisch betrachtet werden.

Im Laufe des Experiments bemerkt man meistens, was man besser oder zumindest anders machen hätte können. Im Sinne einer kritischen und selbstreflexiven Arbeitsweise sollen diese Punkte in diesem Abschnitt angesprochen werden. In dem Diskriminationsexperiment wurden tatsächlich existierende deutsche Wörter als Stimuli verwendet, um natürlicher Kommunikation ein klein wenig näher zu kommen. Allerdings können semantische und morphologische Beziehungen zwischen den Stimuli und eine ungleiche Häufigkeit der kontrastierenden Minimalpaare die Wahrnehmung unter Umständen beeinflussen (vgl. Kerschhofer-Puhalo 2014: 77), was für die Gruppe ohne Deutschkenntnisse zwar keinen Unterschied gemacht hätte, jedoch gegebenenfalls das Ergebnis für die Gruppe der japanischen Deutschlernenden geringfügig verändert hat. Außerdem war die Menge der Minimalpaare dadurch eingeschränkt. Aus diesen Gründen hätten vielleicht doch lieber Nonsenswörter bevorzugt werden sollen. Durch eine größere Menge von Minimalpaaren hätte man zudem die Lerneffekte eventuell verringern, aber wahrscheinlich nicht gänzlich ausschalten können. Mehr Abwechslung im Experiment (andere Aufgaben) zwischen den einzelnen Teilen hätten dem ebenfalls entgegenwirken können, wobei man immer bedenken muss, das Experiment nicht zu lang werden zu lassen, da sonst unweigerlich Ermüdungseffekte eintreten.

Weiters hätte man Trägersätze (sog. *carrier sentences*) einbauen können, um die Diskrimination etwas komplexer zu gestalten. Anstatt bis auf den Vokalkontrast vollkommen idente Wörter zu verwenden, würde ich heute dieselben Wörter mehrmals einsprechen lassen, sodass das X im AXB-Diskriminationsexperiment nicht physikalisch ident zu Stimuli A bzw. B und somit mehr Redundanz vorhanden ist. Die Unterscheidung sollte dadurch wesentlich erschwert und auch näher an der Realität sein. Außerdem wäre zu überlegen, ob man die Aufmerksamkeit in der Aufgabenbeschreibung nicht auf andere Elemente lenken sollte, um natürlicher Kommunikation im phonologischen Modus näherzukommen. Eventuell wäre auch die Herstellung eines Zeitdrucks (ohne Wiederholung der Stimuli) für die Abbildung einer realistischeren Gesprächssituation von Vorteil gewesen. Die Frage, wie man natürliche Wahrnehmung testen kann, soll in der Schlussbetrachtung noch etwas weitergesponnen werden.

8 Resümee und Ausblick

Auf die Gefahr hin, mich zu wiederholen, sollen für den kursorischen Leser/die kursorische Leserin abschließend die wesentlichen Erkenntnisse dieser Arbeit nochmal zusammengefasst werden. Weiters sollen offene gebliebene und neu entstandene Fragen angesprochen werden.

Für die Gruppe der japanischen Studierenden ohne Deutschkenntnisse konnte gezeigt werden, dass die Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts leichte Schwierigkeiten bereitet hat. Im Vergleich dazu hatte die „erfahrene“ Gruppe japanischer Deutschlerner mit der Unterscheidung nahezu keine Probleme, womit der Einfluss sprachlicher Erfahrung aufgezeigt werden konnte. Die Ergebnisse waren beinahe auf dem Level der erstsprachlichen Kontrollgruppe. Der Unterschied zwischen beiden Gruppen war signifikant und die Vorhersagen des SLM, PAM und PAM-L2 konnten bestätigt werden. Aufgrund von unerwünschten Lerneffekten konnte die Wirkung des Störgeräusches auf die Resultate nicht mit Sicherheit bestimmt werden.

Die Ergebnisse des Identifikationsexperimentes entsprachen ebenfalls nicht den durch die Literatur und die eigene Beobachtung geschürten Erwartungen. Die Identifikation des deutschen /e:/- und /i:/-Vokals (in Isolation) wurde von den japanischen Deutschlernenden problemlos bewältigt. Lediglich ein Einfluss des Englischen als erstgelernte Fremdsprache konnte im Buchstabiertest aufgezeigt werden.

Als wesentliche Einflussfaktoren auf die fremdsprachliche Wahrnehmung wurden folgende Variablen identifiziert: der wechselseitige sprachliche Einfluss zwischen L1 und L2 (Cross-linguistic Influence, CLI) – darunter das native Phoneminventar als starke Einflussquelle, die Ähnlichkeit als Maß für die Schwierigkeit eines Vokalkontrasts, die Spracherfahrung in bzw. der Sprachkontakt mit der L2 und damit zusammenhängend die Quantität und Qualität des sprachlichen Inputs während des Zweitsprachenerwerbs, die Koartikulation, die den Einfluss des umgebenden Kontextes beschreibt und nicht zuletzt die Aufmerksamkeit, welche die Wahrnehmung von Sprache maßgeblich steuert. Dazu kommt als entscheidender methodologischer Faktor der verwendete Experimentmodus selbst, der indirekt ebenfalls Einfluss auf die Wahrnehmung nimmt. Es versteht sich von selbst, dass es unmöglich ist, hier alle Einflussgrößen auf die sprachliche Wahrnehmung anzuführen, weshalb für diese Untersuchung eine subjektive Auswahl der vermutlich bedeutendsten Variablen getroffen werden musste.

In jedem Fall hat diese Masterarbeit lediglich an der Oberfläche gekratzt und das beobachtete Phänomen ist bei weitem nicht vollständig geklärt. Es wurde lediglich ein erster Grundstein gesetzt, auf den weitere Forschung gegebenenfalls aufbauen kann. Um diese Studie zu bestätigen oder zu widerlegen, sollten unbedingt noch weitere Wortkontexte des betreffenden Kontrastes für sich betrachtet und mit unterschiedlichen Altersgruppen und Kompetenzniveaus in der L2 Deutsch getestet werden. Weiters müssen gegebenenfalls neue Experimentmodi gefunden werden, um die Grenzen des Experiments zu erweitern. Vor allem hinsichtlich der Identifikation des /i:/-/e:/-Vokalkontrasts besteht noch Forschungsbedarf. Das recht einfach gestrickte Identifikationsexperiment in dieser Arbeit diente lediglich als kleine Ergänzung, um den Erkenntnishorizont etwas zu erweitern.

Die Diskrepanz zwischen den Ergebnissen und der individuellen Beobachtung hat die Grenzen phonetischer Experimente deutlich gemacht. Das ursprüngliche Forschungsinteresse, welches sich aus realer Beobachtung ergeben hat, konnte durch das Experiment nicht erklärt werden. Stattdessen wurden neue Fragen aufgeworfen. Es wird davon ausgegangen, dass sich im beobachteten Widerspruch der Gegensatz zwischen natürlicher Sprachwahrnehmung und den künstlichen Bedingungen des Experiments gezeigt hat – zwei sehr unterschiedliche Kontexte, die offenbar nicht so einfach zusammengebracht werden können. Es stellt sich die Frage, wie und ob natürliche Wahrnehmung am besten nachgebildet und getestet werden kann. Es handelt sich dabei also im Grunde vorrangig um ein methodisches Problem.

Um die Frage beantworten zu können, müssten als erster Schritt zuerst einmal alle Einflussfaktoren auf die Prozessierung von gesprochener Sprache identifiziert werden und/oder andere Wege gefunden werden wie man die Wahrnehmung testen kann. Wie am Beginn des theoretischen Teils erwähnt, ist bis heute unklar, was genau passiert, wenn ein sprachlicher Laut das menschliche Hörorgan erreicht. So kann bis heute niemand gesichert sagen, wie erst- oder zweitsprachliche Hörer/innen es schaffen, Vokale aus dem kontinuierlichen Sprachstrom herauszufiltern und trotz mangelnder Invarianz und Nicht-Linearität des sprachlichen Signals zu identifizieren. Eventuell kann das angeführte methodische Problem erst dann gelöst werden, wenn diese Frage eindeutig geklärt ist.

9 Literaturverzeichnis

Albert, Ruth & Marx, Nicole (2017), *Empirisches Arbeiten in Linguistik und Sprachlehrforschung*. Tübingen: Narr Francke Attempto.

Barkowski, Hans; Grommes Patrick u.a. (2014), *Deutsch als fremde Sprache (Deutsch lehren lernen 3)*. München: Klett-Langenscheidt.

Bent, Tessa & Kewley-Port Diane; Ferguson, Sarah Hargus (2010), Across-talker effects on non-native listeners' vowel perception in noise. In: *The Journal of the Acoustical Society of America* 128 (2010), 3142-3151.

Best, Catherine T.; McRoberts, Gerald W. & Goodell, Elizabeth (2001), Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. In: *Journal of the Acoustic Society of America*, Vol. 109. No. 2, February 2001, 775-794.

Best, Catherine T. & McRoberts, Gerald W. (2003), Infant Perception of Non-Native Consonant Contrasts that Adults Assimilate in Different Ways. In: *Language and Speech* 46, 183-216.

Best, Catherine T. & Tyler, Michael D. (2007), Nonnative and second-language speech perception: Commonalities and complementarities. In: Munro, Murray J. & Bohn, Ocke-Schwen (Hrsg.), *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins, 13-34.

Boersma, Paul & Weenink, David (2017), Praat: doing phonetics by computer [Computer program]. Version 6.0.30, retrieved 22 July 2017 from <http://www.praat.org/>

Bohn, Ocke-Schwen (1998), Wahrnehmung fremdsprachlicher Laute: Wo ist das Problem? In: Wegener, Heide (1998), *Eine zweite Sprache lernen. Empirische Untersuchungen zum Zweitspracherwerb*. Tübingen: Narr, 1-20.

Cooper, Nicole & Cutler, Anne (2004), Perception of Non-native Phonemes in Noise. online unter http://pubman.mpg.de/pubman/item/escidoc:59961:3/component/escidoc:59962/cooper_2004_perception.pdf [letzter Zugriff: 14.08.2017].

Cutler, Anne; Smits, Roel & Cooper, Nicole (2005), Vowel perception: Effects of non-native language vs. non-native dialect. In: *Speech Communication* 47 (2005), 32-42.

Dieling, Helga (1992), *Phonetik im Fremdsprachenunterricht Deutsch* (Fremdsprachenunterricht in Theorie und Praxis). Berlin, München u.a.: Langenscheidt.

Ernst, Peter (2004), *Germanistische Sprachwissenschaft. Eine Einführung in die synchrone Sprachwissenschaft des Deutschen* (korrigierter Nachdruck 2008). Wien: WUV.

Flege, James E. & Munro, Murray J.; Fox, Robert A. (1994), Auditory and categorical effects on cross-language vowel perception. In: *The Journal of the Acoustical Society of America* 95 (1994), 3623-3641.

Flege, James E. (2007), Language contact in bilingualism: Phonetic system interactions. In: Cole, Jennifer & Hualde, José Ignacio (Hrsg.) (2007), *Laboratory Phonology 9*. Berlin/New York: Mouton de Gruyter, 353-381.

Goethe, Johann Wolfgang von (1998), *Dichtung und Wahrheit*. 3. und 4. Teil (Werkausgabe in zehn Bänden, Band 9), Köln: Könnemann.

Guion, Susan G. & Pederson, Eric (2007), Investigating the role of attention in phonetic learning. In: Bohn, Ocke-Schwen & Flege, James Emil (2007), *Language experience in second language speech learning: in honor of James Emil Flege*. Amsterdam u.a.: Benjamins, 57-77.

Hazan, Valerie & Simpson, Andrew (1996), Cue-Enhancement Strategies for Natural VCV and Sentence Materials Presented in Noise. online unter <http://www.phon.ucl.ac.uk/home/sh19/val/hazan.htm> [letzter Zugriff: 14.08.2017].

Hirschfeld, Ursula (1995), Phonetische Merkmale in der Aussprache Deutschlernender und deren Relevanz für deutsche Hörer. In: Deutsch als Fremdsprache 32 (1995), Heft 3, 177-183.

Höhle, Barbara (Hrsg.) (2010), Psycholinguistik. München: Oldenbourg Akademieverlag.

Huang, Tasan (2011), Why are Vowels and Tones Perceived Less Categorically than Stop Consonants? – Categorical Perception and Speech Sound Types Revisited. online unter <http://www.ling.ohio-state.edu/LehisteSymposium/abstracts/Huang.pdf> [letzter Zugriff 09.05.2018].

Ingram, John C. L., Park See-Gyoon (1997), Cross-language vowel perception and production by Japanese and Korean learners of English. In: Journal of Phonetics (1997) 25, 343-370.

Jin, Su-Hyun & Liu, Chang (2014), English vowel identification in quiet and noise: effects of listeners' native language background. In: *frontiers in Neuroscience* Volume 8, Article 305 (September 2014), 1-8.

Kerschhofer-Puhalo, Nadja (2014), Similarity, Cross-linguistic Influence and Preferences in Non-Native Vowel Perception. An experimental cross-language comparison of German vowel identification by non-native listeners. Dissertation Universität Wien.

Kaneko, Tohru & Neyer, Franz-Anton: Vergleich der Lautstrukturen des Deutschen und Japanischen. In: Kaneko, Tohru/Stickel, Gerhard (1987), Japanische Schrift. Lautstrukturen. Wortbildung (Deutsch und Japanisch im Kontrast, Band 1). Heidelberg: Julius Groos Verlag, 64-107.

Kaneko, Yukiko (2005), Phonetisch-phonologischer Vergleich der Vokalsysteme der deutschen und der japanischen Sprache. Zur Produktion deutscher Vokale bei japanischen Muttersprachlern. Frankfurt am Main: Peter Lang.

Kohler, Klaus J. (1995), Einführung in die Phonetik des Deutschen. Berlin: Erich Schmidt Verlag.

Kubuzono, Haruo (2015), Introduction to Japanese phonetics and phonology. In: Kubuzono, Haruo (2015), Handbook of Japanese Phonetics and Phonology. Boston: De Gruyter Mouton, 38-61.

Levy, Erika S. & Strange Winifred (2008), Perception of French vowels by American English adults with and without French language experience. In: Journal of Phonetics 36 (2008), 141-157.

Mazuka, Reiko; Hasegawa, Mihoko & Tsuji, Sho (2013), Development of Non-Native Vowel Discrimination: Improvement Without Exposure. In: Developmental Psychobiology, February 2014 Vol. 56 (2), 192-209.

McCloy, Daniel, Praat Script "Mix Speech with Noise" (modified by Chun Chan), Online unter:

http://groups.linguistics.northwestern.edu/speech_comm_group/documents/praat%20scripts/MixSpeechNoise.praat [letzter Zugriff: 12.05.2018].

Mora, Joan C. (2007), Methodological issues in assessing L2 perceptual phonological competence. online unter

https://www.researchgate.net/publication/237626298_Methodological_issues_in_assessing_L2_perceptual_phonological_competence [letzter Zugriff: 10.05.2018].

Mokari, Payam Ghaffarvand & Werner, Stefan (2017), Perceptual assimilation predicts acquisition of foreign language sounds: The case of Azerbaijani learners' production and perception of Standard Southern British English vowels. In: Lingua 185 (2017), 81-95.

Ntounis, Leandros (2012), Crowd in a bar (LCR recording). online unter

<https://freesound.org/people/Leandros.Ntounis/sounds/163995/> [letzter Zugriff: 11.05.2018].

Pisoni, David B. (1973), Auditory and phonetic memory codes in the discrimination of consonants and vowels. online unter

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3515632/pdf/nihms418818.pdf> [letzter Zugriff: 09.05.2018].

Pisoni, David B. (1975), Auditory short-term memory and vowel perception. online unter <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3831373/pdf/nihms418810.pdf> [letzter Zugriff: 15.08.2017].

Polka, Linda & Werker, Janet F. (1994), Developmental Changes in Perception of Nonnative Vowel Contrasts. In: *Journal of Experimental Psychology: Human Perception and Performance* 1994, Vol. 20, No. 2, 421-435.

Polka, Linda (1995), Linguistic influences in adult perception of non-native vowel contrasts. In: *Journal of Acoustic Society of America* 97 (2), February 1995, 1286-1296.

Raphael, Lawrence J. (2005), Acoustic Cues to the Perception of Segmental Phonemes. In: Pisoni, David B. & Remez, Robert E. (2011), *The Handbook of Speech Perception*. Oxford: Blackwell Pub., 182-206.

Rausch, Rudolf & Rausch, Ilka (2002), *Deutsche Phonetik für Ausländer*. Leipzig, Berlin u.a.: Langenscheidt/Verlag Enzyklopädie.

Reetz, Henning & Jongman, Allard (2009), *Phonetics. Transcription, Production, Acoustics, and Perception*. Hoboken u.a.: Wiley-Blackwell.

Schmelter, Lars (2014), Gütekriterien. In: Settinieri, Julia u.a. (2014): *Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache*. Ferdinand Schönigh: Paderborn, 33-44.

Sendlmeier, Walter F. & Seebode, Julia (2007), *Formantkarten des deutschen Vokalsystems*. TU Berlin, Institut für Sprache und Kommunikation. online unter https://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf [letzter Zugriff: 09.05.2018].

Settinieri, Julia (2016), Deskriptiv- und Inferenzstatistik. In: Caspari, Daniela u.a. (2016), *Forschungsmethoden in der Fremdsprachendidaktik. Ein Handbuch*. Tübingen: Narr, 324-341.

Vance, Timothy J. (1987), *An Introduction to Japanese Phonology*. Albany: State University of New York Press.

Wiese, Richard (2011), *Phonetik und Phonologie* (UTB 3354). Paderborn: Wilhelm Fink.

Winn, Matt, Make SSN from LTAS selected sounds (Praatskript). online unter http://www.mattwinn.com/praat/Make_SSN_from_LTAS_selected_sounds.txt [letzter Zugriff: 11.05.2018].

eLAUT – Phonetische Beschreibung einzelner Sprachen: Deutsch. online unter http://www.coli.uni-saarland.de/elaut/Languages_Sites/sampaDeutsch.htm [letzter Zugriff: 10.05.2018].

eLAUT – Phonetische Beschreibung einzelner Sprachen: Japanisch. online unter http://www.coli.uni-saarland.de/elaut/Languages_Sites/sampaJapanisch.htm [letzter Zugriff: 10.05.2018].

10 Anhang

A Abstract

Lieben oder leben? Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts durch Hörer/innen mit L1 Japanisch

Die Wahrnehmung neuer Phonemkontraste stellt mitunter ein Problem für L2-Lernende dar. Ziel dieser Untersuchung ist es, herauszufinden, ob und wie gut erwachsene japanische DaF-Lernende den /i:/-/e:/-Vokalkontrast unterscheiden können, welcher im Japanischen nicht kontrastiv ist. Dazu werden zwei Gruppen japanischer Studierender getestet, die sich hinsichtlich Deutschkenntnissen bzw. Kontaktdauer mit der L2 Deutsch unterscheiden. Durch den Vergleich beider Gruppen soll die Relevanz des nur schwer quantifizierbaren Faktors Spracherfahrung für diesen Kontrast bestimmt werden. Die Vorhersagen zweier phonetischer Modelle – des Perceptual Assimilation Models (PAM bzw. die Weiterentwicklung PAM-L2) von Best und Tyler und des Speech Learning Models (SLM) von Flege – sollen dabei überprüft werden. Außerdem sollen weitere wesentliche Einflussgrößen auf die fremdsprachliche Vokalwahrnehmung identifiziert werden. Aufgrund der inter- und intralingualen Ähnlichkeit des Vokalkontrasts und der Ergebnisse vorangegangener Studien wurde von einem Category-Goodness Difference ausgegangen, wonach Hörer/innen mit L1 Japanisch voraussichtlich leichte Schwierigkeiten haben, die beiden Vokale zu diskriminieren (Diskriminationsrate < 90%). Dem PAM-L2 zufolge ist die Diskrimination der Gruppe der japanischen Student/inn/en mit Spracherfahrung voraussichtlich signifikant besser als jene ohne sprachlicher Erfahrung in der L2 Deutsch und voraussichtlich annähernd „native-like“. Die Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts durch Sprecher/innen mit L1 Japanisch wird in einem AXB-Diskriminationsexperiment untersucht. Dazu werden zwei Gruppen japanischer Studierender zwischen 18 und 22 Jahren der Dokkyo Universität in Saitama getestet. Gruppe A besteht aus 10 Studierenden ohne Deutschkenntnisse, Gruppe B aus 17 japanischen Deutschstudent/inn/en, die seit mindestens 2 Jahren Deutsch lernen. Als Stimuli fungieren 24 Minimalpaare (gemischt mit anderen irrelevanten Vokalkontrasten), die bis auf den betreffenden Vokal identisch sind (z.B. *lieben* - *leben*). Um die Gleichheit der Silbenränder zu gewährleisten, wurde der „i-Vokal“ in das „e-Wort“ hineingeschnitten (sog. splicing). Zur Untersuchung der Fragen wird ein MFC-Wahrnehmungsexperiment in Praat konzipiert. Außerdem werden insgesamt drei Schwierigkeitsstufen (mit und ohne Störgeräusch) eingefügt, um die Bedingungen des Hörens

zu erschweren und möglicherweise größere Unterschiede herausarbeiten zu können. Als Störgeräusch wird ein Speech Shaped Noise (SSN) in unterschiedlicher Stärke (0dB SNR und -5dB SNR) verwendet, das mithilfe eines Praatskripts aus einer vielstimmigen Aufnahme eines Stimmengewirrs in einer Bar erzeugt wurde. In einem weiteren Experiment wird die Identifikation getestet. Die Studie versteht sich als Beitrag zum Verständnis der Vokalwahrnehmung in der Zweit- bzw. Fremdsprache.

B Diskriminationsexperiment ExperimentMFC-

Konfigurationsdatei

C Praat-Skript zur Erzeugung des SSN

D Praat-Skript zum Mischen des SSN mit den Stimuli

E Liste mit Minimalpaaren (Experiment 1)

F Ergebnisse Diskriminationsexperiment

G Fehlerverteilung nach Minimalpaaren

H Liste der Familiennamen (für Experiment 2)

I Identifikationsexperiment ExperimentMFC-

Konfigurationsdatei

J Ergebnisse Identifikationsexperiment