

Zeitschrift für Interkulturellen Fremdsprachenunterricht

Didaktik und Methodik im Bereich Deutsch als Fremdsprache

ISSN 1205-6545 Jahrgang 16, Nummer 2 (Oktober 2011)

Soziale Akzeptanz unterschiedlicher Normabweichungen in der L2-Aussprache Deutsch

Julia Settinieri

Deutsch als Fremd- und Zweitsprache
Fakultät für Linguistik und Literaturwissenschaft
Universität Bielefeld
Postfach 10 01 31
Bielefeld
Tel.: 0049 521 106 3634
Fax: 0049 521 106 2996
Email: julia.settinieri@uni-bielefeld.de

Abstract. Während die Auswirkungen unterschiedlicher phonetischer Abweichungen von der Zielsprachennorm auf die Verständlichkeit und Akzentuiertheit von L2-Äußerungen bereits relativ gut erforscht sind, herrscht hinsichtlich der Konsequenzen für die soziale Akzeptanz devianter Aussprachemerkmale noch großer Forschungsbedarf. Erste empirische Studien zur Reaktion von Muttersprachlern auf den Akzent von Nichtmuttersprachlern weisen jedoch bereits auf das Problem der negativen Sanktionierung bestimmter Akzentmerkmale in der L2 hin. Im Rahmen der hier beschriebenen quantitativen Querschnittsstudie werden in Anlehnung an die *matched guise technique* (Lambert, Hodgeson, Gardner & Fillenbaum 1960) von muttersprachlichen Sprechern eingesprochene Kurzsätze, die jeweils genau eine von neun verschiedenen, für L2-Sprecher typischen segmentalen Abweichungen enthalten, sowie ihre fehlerfreien Dubletten 100 muttersprachlichen Ratern zur Bewertung vorgelegt. Diese evaluieren die randomisiert eingespielten Äußerungen mittels Likert-Skalen hinsichtlich ihrer sozialen Akzeptabilität. Mit Hilfe von Varianzanalysen wird untersucht, ob die verschiedenen Fehlertypen signifikant unterschiedlich bewertet werden. Anschließend werden Einstellungen der Rater zu unterschiedlichen Akzenten und Akzentmerkmalen erhoben. Die Ergebnisse zeigen, dass einzelne phonetische Abweichungen signifikant positiver bzw. negativer bewertet werden als andere. Darüber hinaus korrelieren die Bewertungen für L1-Sprecher des Französischen bzw. des Russischen typischer Fehler mit der allgemeinen Bewertung eines französischen bzw. russischen Akzents. Die abschließende Diskussion der Ergebnisse wirft methodologische Fragestellungen, insbesondere das Spannungsverhältnis zwischen interner und externer Validität von Studien zur sozialen Akzeptanz betreffend, auf.¹

While the effects of various phonetic deviations from the target norm on the intelligibility and accentedness of L2 utterances have already been relatively well researched, there is still a need for more research in the field of the effects of deviant pronunciation features on their social acceptability. Initial empirical studies on the effects of foreign sounding accents, however, point to the problem that particular characteristics of an L2 accent often invoke negative reactions from native speakers. This paper will describe a quantitative cross-sectional study which uses a variant of the so called *matched guise technique* (Lambert; Hodgeson; Gardner & Fillenbaum 1960). Native-speakers of German utter several short sentences. Each one of these sentences contains one of nine different segmental deviations from the target norm that are typical for L2 speakers. These sentences and their error-free equivalents are recorded and then presented to 100 native raters. These raters evaluate the randomised utterances on a Likert scale with regard to their social acceptability. An analysis of variance investigates whether the evaluations of the various types of deviations show statistically significant differences. Afterwards, the raters' attitudes towards non-native accents and accent features are surveyed. The findings indicate that some deviations are rated significantly higher or lower than others. Furthermore, the ratings for deviations which are typical for speakers with L1 French or L1 Russian correlate with the overall ratings of a French or Russian accent respectively. The final discussion of the results raises methodological questions, particularly concerning the conflict between internal and external validity of studies on social acceptability of L2 accents.

Schlagwörter: Ausspracheerwerb, Akzentbewertung, soziale Akzeptanz, Validität, *matched guise technique*

1. Einleitung

L2-Aussprache ist lange Zeit primär unter einer (zudem stark normorientierten) Produktionsperspektive betrachtet worden. Entscheidend für das Gelingen von Kommunikation ist jedoch nicht nur die Produktions-, sondern letztendlich vor allem die Rezeptionsseite, da nicht alle Abweichungen von der Zielnorm in gleichem Maße störend wirken. Diesen Faktor bezieht der Sprecher in der Regel auch bereits in seine Äußerungen ein. Grundsätzlich ist davon auszugehen, dass die Ausspracheformen, die ein Sprecher produziert, von drei Wirkgrößen beeinflusst werden: Einerseits bemüht sich der Sprecher um Verständlichkeit, wobei er gleichzeitig versucht, seinen Artikulationsaufwand so gering wie möglich zu halten. Als dritte Wirkgröße kommt die soziale Akzeptabilität unterschiedlicher Formen ins Spiel, die beeinflusst, ob sich ein Sprecher in einer gegebenen Sprechsituation eher um Klarheit oder eher um Nähesprachlichkeit (vgl. Koch & Oesterreicher 1985) bemüht. Kohler (2001: 39f) fasst zusammen: „[P]honic variability is considered to be controlled by a complex interplay of articulatory economy in speech production, of perceptual distinctiveness for speech recognition and of social acceptance in speech communication.“

Während der Aspekt der Verständlichkeit und Akzentuiertheit² nichtmuttersprachlicher Äußerungen, insbesondere auch international, bereits recht ausführlich erforscht worden ist (vgl. z.B. Anderson-Hsieh, Johnson & Kohler 1992; Derwing & Munro 1997; Hirschfeld 1994, 1995; Jilka 2000; Magen 1998; Munro & Derwing 1995) und auch bereits einige Studien zu Reduktionsformen in der L2-Aussprache vorliegen (vgl. Gut 2003; Kaltenbacher 1998; Richter 2008; Wenk 1985; Young-Scholten 1993), ist der Aspekt der sozialen Akzeptanz meiner Einschätzung nach bislang am wenigsten erforscht worden. Gleichzeitig stellen die sozialen Implikationen nonnormativer Aussprache einen zentralen Faktor für einen erfolgreich verlaufenden Ausspracheerwerb dar, da mangelnde Akzeptabilität bestimmter phonetischer Merkmale der Lernersprache zu Sanktionierungen seitens der Hörer führen kann, was wiederum zu Sprechangst und damit zur Hemmung des gesamten Spracherwerbsprozesses führen kann (vgl. Grotjahn 1998: 40f; Hirschfeld 2001: 872). Daher widmet sich die hier beschriebene Studie dem Problem der Erforschung der sozialen Akzeptabilität unterschiedlicher Ausspracheabweichungen in der L2 Deutsch.

2. Forschungsüberblick

Dass auch das Prestige einer (Aus-)Sprache Einfluss auf die Bewertung einzelner Sprecher haben kann, zeigt die bahnbrechende Studie von Lambert, Hodgeson, Gardner & Fillenbaum (1960). Kanadische Sprecher in Montreal mit L1 Englisch bzw. Französisch bewerten vier Sprecher hinsichtlich ihres Charakters. Sie werden nicht darüber aufgeklärt, dass sie vier bilinguale Sprecher einmal auf Englisch und einmal auf Französisch hören werden, sondern bekommen quasi acht Sprecher, die zur Hälfte Englisch und zur Hälfte Französisch sprechen, vorgespielt. Dabei zeigen die französischen Rater eine Art Minderheitenreaktion, indem sie die Sprecher der eigenen L1-Gruppe im Durchschnitt schlechter bewerten als dies die anglophonen tun. Dies wird mit dem geringeren sozialen Prestige der frankophonen Kanadier im Verhältnis zu den anglophonen erklärt. Ähnliches konnte wiederholt für Sprecher unterschiedlicher Dialekte gezeigt werden (vgl. den Forschungsüberblick bei Hellwig-Fábián 2007: 30f).

Schließlich liegen auch erste empirische Studien zur Reaktion von Muttersprachlern auf den Akzent von Nichtmuttersprachlern vor, die auf das Problem der negativen Sanktionierung bestimmter Akzentmerkmale in der L2 fokussieren und von denen hier einige exemplarisch referiert werden sollen. Dretzke (1987) untersucht die Wirkung muttersprachlich deutscher Sprecher in der L2 Englisch. Eine bilinguale Sprecherin spricht Texte mit je 12-14 Fehlern für insgesamt 24 Fehlertypen sowie fehlerfreie Kontrolltexte (mit und ohne Pausen) ein. 225 Rater bewerten diese mittels siebenstufiger Likert-Skala hinsichtlich *pleasantness*, *intelligibility* und *social prestige*. Besonders negativ in der Gesamtbewertung schneiden [p] statt [b], [k] statt [g], [ŋ] statt [ŋg] und [s] statt [θ] ab. Dretzke (1987: 513) leitet aus den Ergebnissen anschließend ein System zur Bewertung der Fehlerschwere unterschiedlicher Abweichungen ab.

Cunningham-Andersson (1997) untersucht die Bewertung von L2-Akzenten durch Muttersprachler des Schwedischen. 98 Rater (80 Technik-Studierende, 18 SaF-Lehrer) bewerten insgesamt 11 Sprecher, die die schwedische Version von *The Northwind and the Sun* eingelesen haben. Die fünf zu bewertenden Fehlertypen sind in einer den Ratern vorliegenden schriftlichen Version des Textes unterstrichen und werden auf einer 5-stufigen Skala nach folgenden Kriterien evaluiert: „a) how important they felt it was that the speakers should try to eliminate each of the

Julia Settineri (2011), Soziale Akzeptanz unterschiedlicher Normabweichungen in der L2-Aussprache Deutsch. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 16: 2, 66-80. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-16-2/beitrag/Settineri.pdf>.

deviant pronunciations from their speech, b) how friendly the speakers sounded, c) how intelligent the speakers sounded” (Cunningham-Andersson 1997: 136). Die Ergebnisse zeigen, dass die untersuchten phonetischen Abweichungen unterschiedlich negative Bewertungen hinsichtlich Freundlichkeit und Intelligenz des Sprechers nach sich ziehen. Ein auf den velaren Nasal folgender Plosiv, regressive Assimilation an die Stimmhaftigkeit und Wortakzentfehler schneiden dabei schlechter ab als Schwa-Epenthese und mangelnde Unterscheidung von Vokalqualitäten. Unterschiedliche Angaben zur Herkunft der Sprecher (ebenfalls in Form eines *matched-guise*-Designs) haben hingegen keinen Einfluss auf die Bewertung, was allerdings auch daran liegen könnte, dass den Ratern die doppelte Zuordnung teilweise aufgefallen war. Insgesamt bewerten die Lehrer milder als die Technik-Studierenden. Cunningham-Andersson (1997: 142) schlussfolgert: „It would clearly be worthwhile for immigrants to learn to avoid the non-native pronunciations that are stigmatized.”

Cohrs (2007) lässt eine laut Expertenurteil im Deutschen als L2 phonetisch sehr gute, eine mittlere und eine schwache Sprecherin mit Ungarisch als L1 von muttersprachlichen Hörern hinsichtlich sprachlicher Gestaltung, Sprechwirkung, Sprechleistung und Sprecherbeurteilung bewerten. Die Ungarinnen werden interviewt, lesen ein Märchen und ein Gedicht auf Band. Die Aufnahmen werden von den Ratern, 60 DaF-Studierenden im zweiten Semester, durchgängig als weniger „interessant“, „angenehm“ und „sympathisch“ eingestuft, je niedriger das phonetische Niveau ist. Zu bedenken ist allerdings, dass lediglich drei Sprecherinnen evaluiert werden, die sich über ihr Ausspracheniveau im Deutschen hinaus auch hinsichtlich anderer Merkmale, wie z.B. Stimmqualität, Sprechtempo usw. voneinander unterscheiden, so dass die Ergebnisse kaum zu generalisieren sind.

Köhne (2009) untersucht in ihrer Masterarbeit Zusammenhänge zwischen dem Grad segmentaler und suprasegmentaler Ausspracheabweichungen und der soziophonetischen Bewertung türkischer Deutschlerner durch deutsche Muttersprachler. Da insgesamt nur zehn Sprecher bewertet werden, interpretiert Köhne die Ergebnisse ihrer Studie sehr vorsichtig. Deutlich wird jedoch, dass objektive und subjektive Beurteilung des Ausspracheniveaus starke Übereinstimmungen aufweisen und dass auch Gesamtbeurteilungen der Sprecher im Aussprache- und Persönlichkeitsbereich zahlreiche Analogien zum Ausspracheniveau zeigen. Darüber hinaus zeigen sich Zusammenhänge zur Bewertung von Verständlichkeit und Akzeptabilität der anhand eines *Map Tasks* elizitierten Äußerungen. Ferner scheint es, dass die Sprecher tendenziell als weniger sympathisch und weniger gebildet eingeschätzt werden, wenn sie als Türken erkannt werden als wenn sie nicht eingeordnet werden können oder für Westeuropäer gehalten werden.

Zusammenfassend zeichnet sich ab, dass unterschiedliche Abweichungen von der Zielnorm soziophonetisch unterschiedlich bewertet werden und dass Zusammenhänge zwischen der Verständlichkeit, der Akzentuiertheit bzw. dem Ausspracheniveau, der angenommenen Herkunft und der soziophonetischen Bewertung bestehen.

3. Empirische Studie

3.1 Fragestellungen, Hypothesen und Zielsetzungen

Im Spannungsfeld dieser Faktoren siedelt sich auch die im Folgenden beschriebene empirische Querschnittsstudie an. Zielsetzung der Untersuchung ist zum einen aufzuzeigen, dass unterschiedliche typische phonetische Zielsprachenabweichungen in der L2 Deutsch von L1-Ratern signifikant unterschiedlich positiv bzw. negativ bewertet werden. Dies wäre als Beleg für die Relevanz der sozialen Akzeptabilität unterschiedlicher phonetischer Abweichungen von der Zielnorm zu werten. Zum anderen sollen erste Hinweise darauf gewonnen werden, welche Herkunftssprachen und welche Fehlertypen besonders positiv bzw. negativ sanktioniert werden. Schließlich sollen auch mögliche Datenerhebungsmethoden auf ihre Gegenstandsadäquatheit überprüft werden. Diese Erkenntnisse könnten für das Design tiefer gehender Folgestudien hilfreich sein. Im Einzelnen verfolgt die Studie die folgenden Fragestellungen:

- Werden unterschiedliche segmentale bzw. phonotaktische Abweichungen von der deutschen Standardaussprache von L1-Hörern signifikant unterschiedlich positiv bzw. negativ bewertet? (→ vgl. Hypothese 1, weiter unten)
- Lassen sich Unterschiede in der Bewertung männlicher vs. weiblicher Rater messen? (→ Hypothese 2)
- Welche Akzente werden im Deutschen besonders positiv bewertet?
- Welche Akzente werden im Deutschen besonders negativ bewertet?

Julia Settineri (2011), Soziale Akzeptanz unterschiedlicher Normabweichungen in der L2-Aussprache Deutsch. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 16: 2, 66-80. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-16-2/beitrag/Settineri.pdf>.

- Werden ein russischer und ein französischer Akzent im Deutschen als L2 signifikant unterschiedlich positiv bzw. negativ bewertet? (→ Hypothese 3)
- Lässt sich ein Zusammenhang zwischen der Bewertung eines russischen bzw. französischen Akzents und der Bewertung phonetischer Abweichungen, die nur im Russischen, nur im Französischen oder in beiden Sprachen vorkommen, feststellen? (→ Hypothese 4)
- Welche Akzentmerkmale werden im Deutschen besonders positiv bewertet?
- Welche Akzentmerkmale werden im Deutschen besonders negativ bewertet?

Ein Teil der Studie ist demnach hypothesentestend, ein anderer hypothesengenerierend angelegt. Die zu testenden Hypothesen lauten:

- Hypothese 1: Unterschiedliche segmentale bzw. phonotaktische Abweichungen von der deutschen Standardaussprache werden von L1-Hörern signifikant unterschiedlich positiv bzw. negativ bewertet.
- Hypothese 2: Weibliche L1-Rater bewerten die Abweichungen insgesamt negativer als männliche.³
- Hypothese 3: Ein russischer Akzent im Deutschen als L2 wird signifikant negativer bewertet als ein französischer.⁴
- Hypothese 4: Rater, die einen russischen bzw. französischen Akzent positiv bzw. negativ bewerten, bewerten auch phonetische Abweichungen, die dem Russischen bzw. Französischen zuzurechnen sind, signifikant positiver bzw. negativer als die jeweils andere Gruppe.

3.2 Design

Zunächst werden Muttersprachlern des Deutschen Teststimuli zur soziophonetischen Bewertung vorgelegt. Bei den Items handelt es sich um Einzelsätze, die von vier Muttersprachlern (zwei Frauen, zwei Männern) eingesprochen und instrumentalphonetisch so verändert werden, dass sie jeweils einmal in ihrer standardsprachlichen Form und einmal mit genau einer segmentalen Ausspracheabweichung vorliegen.⁵ Durch den Einsatz muttersprachlicher Sprecherinnen und Sprecher wird sichergestellt, dass die Stimuli wirklich immer nur genau ein nichtmuttersprachliches Merkmal enthalten. Dieses Verfahren stellt eine Weiterentwicklung der in vorangehenden Studien häufig verwendeten *matched guise technique* (Lambert et al. 1960) dar.

Die vier Sprecher sind Studierende, die von zwei phonetisch geschulten Muttersprachlern aus mehreren Stimmen ausgewählt werden. Die Sprecher müssen zum einen in der Lage sein, alle erforderlichen Fehler zu produzieren. Zum anderen sollen sie abgelesene Sätze möglichst natürlich intonieren können und eine angenehme, neutrale Stimme haben.

Als segmentale Abweichungen werden exemplarisch neun im Deutschen als L2 häufig vorkommende Fehlertypen in die Untersuchung aufgenommen, wie sie u.a. von russischen und/oder französischen L2-Sprechern des Deutschen produziert werden (s. Tabelle 1).⁶ Pro Fehlertyp werden zwei Sätze eingesprochen. Dabei wird auf größtmögliche Strukturähnlichkeit der Sätze geachtet: Es handelt sich durchgehend um Aussagesätze in der 3. Person Singular, die zur Hälfte mit *er*, zur Hälfte mit *sie* beginnen und vier bis sechs Silben lang sind. Der Satzakkzent liegt auf der Silbe, die den Aussprachefehler enthält, sofern dies silbenstrukturell möglich ist (Ausnahmen: Nr. 7 und 8). Die Sätze sind lexikalisch und grammatisch simpel gehalten, enthalten keine Namen und auch keine schwierig zu artikulierenden phonetischen Strukturen, z.B. keine komplexen Konsonantencluster. Sie sind emotional neutral und so gebaut, dass sie trotz der enthaltenen Fehler verständlich bleiben.

Insgesamt entstehen so 9 (Fehlertypen) x 2 (Sätze pro Fehlertyp) x 4 (Sprecher) x 2 (mit und ohne Fehler) Sätze, von denen alle fehlerhaften, jedoch nur die Hälfte der fehlerfreien, also insgesamt 108 Sätze, den Ratern vorgespielt werden. Pretests hatten gezeigt, dass der Einschluss aller fehlerfreien Sätze die Konzentrationsspanne der Rater überfordert hätte. Da der Einschluss der fehlerfreien Sätze ohnehin nur dazu dient, einen Eindruck davon zu bekommen, ob die Rater fehlerfreie Sätze auch als solche erkennen und nicht etwa als nichtmuttersprachlich bewerten, erscheinen 36 fehlerfreie Sätze eine ausreichend große Bewertungsbasis zu sein.

Julia Settineri (2011), Soziale Akzeptanz unterschiedlicher Normabweichungen in der L2-Aussprache Deutsch. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 16: 2, 66-80. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-16-2/beitrag/Settineri.pdf>.

Nr.	Fehlerbeschreibung und Testitems	L1 Russ.	L1 Franz.
1	regressive Assimilation an Stimmhaftigkeit statt progressive Assimilation an Stimmlosigkeit <i>Er spielt Fu<u>ß</u>ball./Sie findet ihn nied<u>l</u>ich.</i>	X	X
2	/i/ statt /ɪ/ <i>Er malt ein Bi<u>l</u>d./Er hat Ki<u>n</u>der.</i>	X	X
3	/ŋg/ statt /ŋ/ <i>Er überlegt lan<u>g</u>e./Sie mag viele Di<u>n</u>ge.</i>	X	X
4	/z/ statt /t͡s/ <i>Sie geht in den <u>Z</u>oo./Sie nimmt den <u>Z</u>ug.</i>		X
5	/ʃ/ statt /ç/ <i>Er holt die Mi<u>l</u>ch./Er hat eine Ni<u>ch</u>te.</i>		X
6	/ʀ/ statt /h/ <i>Er ruft seinen <u>H</u>und./Sie ist bereits <u>h</u>ier.</i>		X
7	/ʁ/ statt /v/ <i>Sie trifft ihre Mu<u>t</u>ter./Sie trinkt ein Bi<u>e</u>r.</i>		X
8	/r/ statt /v/ <i>Sie nimmt die Bu<u>t</u>ter./Sie sucht seine Uh<u>r</u>.</i>	X	
9	/x/ statt /h/ <i>Er kauft ein <u>H</u>aus./Sie ist sehr <u>h</u>übsch.</i>	X	

Tabelle 1: Testitems

Die Testsätze werden insgesamt 100 per Aushang rekrutierten Muttersprachlern des Deutschen (50 Frauen, 50 Männer; alles Studierende der Universität Bielefeld, die in Bielefeld oder maximal 100 Kilometer von der Stadt entfernt aufgewachsen sind, zwischen 20 und 30 Jahre alt [mean = 24,45, SD = 2,86]) in randomisierter Reihenfolge vorgespielt und mittels Likert-Skalen hinsichtlich ihrer sozialen Akzeptabilität bewertet. Die Befragung erfolgt schriftlich am Computer, wobei die Testitems über Kopfhörer in einer den Probanden angenehmen Lautstärke eingespielt werden.⁷ Im Anschluss an einleitende Instruktionen lautet der Stimulus für die Testsätze: *Wie bewerten Sie diesen Akzent?* Die endpunktbenannte 7-stufige Likert-Skala ist 1 = *negativ* und 7 = *positiv* gepolt sowie mit der Antwortmöglichkeit *Ich höre keinen Akzent*. versehen.⁸ Intendiert ist also eine Globalbewertung hinsichtlich der sozialen Akzeptabilität einer Äußerung, da bei einer Erhebung unterschiedlicher Bewertungskriterien ungeklärt bliebe, zu welchen Anteilen diese tatsächlich in die Gesamtwirkung eingingen.

Im Anschluss an die Bewertung der Testsätze werden zur Generierung weiterer im Rahmen einer Folgestudie zu testender Hypothesen Fragen zu Einstellungen bzgl. unterschiedlicher Akzenten gestellt, und zwar welche Akzente im Deutschen für die Befragten besonders schön bzw. nicht so schön klingen, welche Aussprachemerkmale ihnen an einem fremden Akzent besonders gut bzw. nicht so gut gefallen und schließlich wie sie einen russischen bzw. französischen Akzent im Deutschen als L2 bewerten (endpunktbenannte 7-stufige Likert-Skala mit 1 = *negativ* und 7 = *positiv*). Abschließend werden biographische Daten der Rater (Geschlecht, Alter, Ort, an dem die Rater aufgewachsen sind, Muttersprachen, Nationalitäten, Studiengänge, Studienfächer, Fremdsprachen) erhoben. Die Befragung endet mit Platz für Anmerkungen, einem Benachrichtigungsangebot über Ergebnisse der Studie und einem Dankeschön.

Julia Settineri (2011), Soziale Akzeptanz unterschiedlicher Normabweichungen in der L2-Aussprache Deutsch. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 16: 2, 66-80. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-16-2/beitrag/Settineri.pdf>.

3.3 Ergebnisse

Im Ergebnis ist festzustellen, dass die H1 zutrifft: Während die Bewertungen der einzelnen Ausspracheabweichungen, die jeweils auf 800 Urteilen beruhen, einander in einem relativ breiten Mittelbereich gleichen, werden Nummer 1 und 3 signifikant über-, Nummer 4 signifikant unterdurchschnittlich bewertet.⁹ Bonferroni-korrigierte t-Tests¹⁰ für verbundene Stichproben zeigen, dass die folgenden Unterschiede signifikant sind:

Paarung	t	df	p (2-seitig)
Nr. 1 - Nr. 2	5,332	51	,000
Nr. 1 - Nr. 4	8,785	50	,000
Nr. 1 - Nr. 6	3,802	47	,000
Nr. 1 - Nr. 7	4,149	49	,000
Nr. 1 - Nr. 9	5,050	51	,000
Nr. 3 - Nr. 4	8,038	69	,000
Nr. 3 - Nr. 9	4,497	71	,000
Nr. 4 - Nr. 2	-6,089	92	,000
Nr. 4 - Nr. 8	-5,229	97	,000

Tabelle 2: t-Test-Ergebnisse für Mittelwertunterschiede zwischen den Bewertungen der einzelnen Fehlertypen

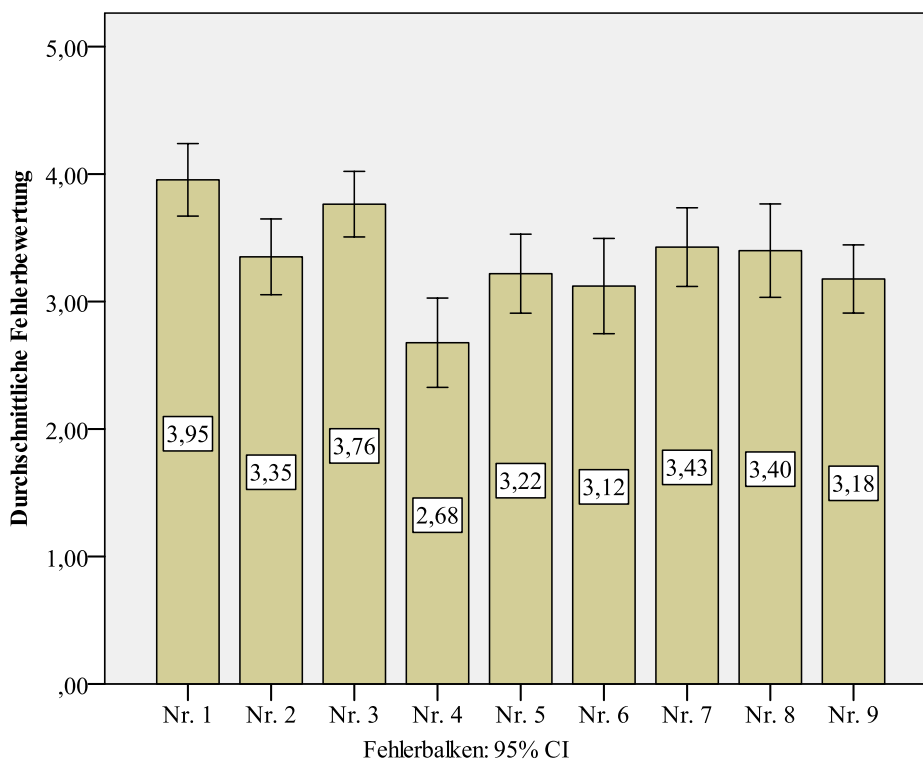


Abbildung 1: Fehlerbewertungen im Vergleich

Auffällig ist außerdem, dass alle Mittelwerte unter 4 liegen, was der Mittelwert der 7-stufigen Likert-Skala ist, dass mithin alle Abweichungen von der Zielnorm negativ bewertet werden.

H2 hingegen muss zurückgewiesen werden. Es können keine signifikanten Bewertungsunterschiede zwischen den Geschlechtern festgestellt werden. Für H3 und H4 gibt es wiederum starke positive Evidenz. Wie erwartet, wird ein russischer Akzent im Deutschen als L2 signifikant negativer bewertet als ein französischer ($t(99) = 6,23, p = .000$; vgl. Abbildung 2).

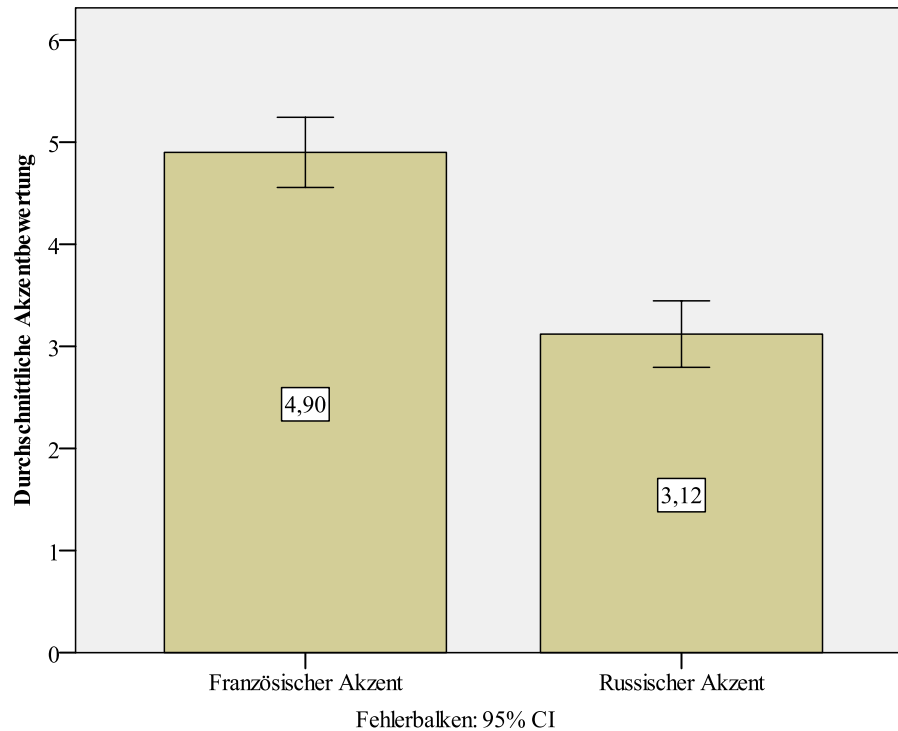


Abbildung 2: Globalbewertung eines französischen vs. eines russischen Akzents im Deutschen als L2

Außerdem zeigt sich, dass Rater, die einen russischen bzw. französischen Akzent positiv bzw. negativ bewerten, auch phonetische Abweichungen, die dem Russischen bzw. Französischen zuzurechnen sind, signifikant positiver bzw. negativer bewerten (s. Tabelle 3). Dies trifft vor allem auf die Bewertung des Russischen zu. Korreliert wird die Globalbewertung eines französischen bzw. russischen Akzents jeweils mit der durchschnittlichen Fehlerbewertung von Fehlern, die sowohl der einen als auch entweder der einen oder der anderen Sprache zugeordnet werden können (*Fehler Französisch* bzw. *Fehler Russisch*), und zum anderen mit dem Durchschnittswert der Fehler, die ausschließlich der einen bzw. der anderen Sprache zugeordnet werden können (*Fehler nur Französisch* bzw. *Fehler nur Russisch*).

	Akzentbewertung Französisch	Akzentbewertung Russisch
Fehler Französisch (Fehler 1-7)	$r = .193, p = .260, \text{n.s.}$	$r = .190, p = .267, \text{n.s.}$
Fehler Russisch (Fehler 1-3, 8-9)	$r = .179, p = .264, \text{n.s.}$	$r = .581, p = .000$
Fehler nur Französisch (Fehler 4-7)	$r = .278, p = .010$	$r = .028, p = .798, \text{n.s.}$
Fehler nur Russisch (Fehler 8-9)	$r = -.156, p = .122, \text{n.s.}$	$r = .631, p = .000$

Tabelle 3: Korrelationen zwischen globalen Akzentbewertungen und dem Französischen bzw. dem Russischen als L1 zuzuordnenden Fehlerbewertungen

Ferner korrelieren die Bewertung eines französischen und eines russischen Akzents negativ miteinander ($r = -.433$, $p = .000$).

Die 100 Befragten sprechen durchschnittlich 2,45 (SD = 1,02) Fremdsprachen. Darunter befindet sich bei 57 der Befragten auch Französisch, bei nur 5 der Befragten Russisch. Die Anzahl erlernter Fremdsprachen korreliert jedoch nicht mit der Globalbewertung eines französischen oder russischen Akzents. Auch spielt es für die Bewertung keine Rolle, ob die Rater selbst Französisch bzw. Russisch sprechen können.

Des Weiteren sind die Studierenden gebeten worden anzugeben, welche Akzente für sie im Deutschen besonders schön, welche weniger schön klingen (offene Frage ohne vorgegebene Antwortoptionen). Tabelle 4 enthält alle Antworten, die mindestens dreimal gegeben wurden, und alle Antworten, die sich auf Varianten des Englischen beziehen. Fettgedruckt sind jeweils die drei häufigsten Nennungen.

„Sprache“	Positiv	Negativ
Arabisch	1	6
Chinesisch	-	7
Englisch	25	5
Englisch, amerikanisches	4	6
Englisch, britisches	3	-
Englisch, Oxford	-	1
<i>Französisch</i>	54	20
Italienisch	16	5
Niederländisch	13	7
Polnisch	5	24
<i>Russisch</i>	15	56
Spanisch	26	10
Türkisch	-	24

Tabelle 4: Beliebte und unbeliebte Akzente im Deutschen

Französisch und Russisch sind kursiv gesetzt, da bei der Interpretation der Ergebnisse im Hinterkopf behalten werden muss, dass die Probanden, wenn sie diese Frage beantworten, bereits 108 Sätze, die Ausspracheabweichungen, wie sie von französischen und/oder russischen Deutsch-als-L2-Sprechern häufig gemacht werden, gehört haben. Die häufige Nennung dieser beiden Sprachen ist also nur bedingt aussagekräftig, da sie möglicherweise auf einen Reihenfolgeeffekt der Befragung zurückzuführen ist.

Schließlich antworten die 100 Studierenden auf die Fragen: *Welche Aussprachemerkmale gefallen Ihnen an einem fremden Akzent besonders/nicht so gut? Versuchen Sie bitte, diese möglichst genau zu beschreiben.* Hier zeigt sich, wie anzunehmen war, dass es phonetischen Laien schwerfällt, diese Auffälligkeiten in Worte zu fassen. Exemplarisch können die folgenden Antworten¹¹ einen Eindruck vermitteln:

- *Mir gefällt eine bestimmte Wortmelodie, die ich nicht genau beschreiben kann.*
- *Wenn durch die Betonung nicht das Wort verschlissen bzw. verschmiert klingt, sondern die Worte noch ein bisschen Schwung dazu gewinnen.*
- *Besonders gut gefällt mir, wenn die Aussprache der Buchstabenfolge so wie in der eigentlichen Muttersprache gesprochen werden.*

Julia Settineri (2011), Soziale Akzeptanz unterschiedlicher Normabweichungen in der L2-Aussprache Deutsch. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 16: 2, 66-80. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-16-2/beitrag/Settineri.pdf>.

Auch diese Beobachtung sollte bei der Konstruktion von Folgestudien berücksichtigt werden, da sie deutlich zeigt, dass direkte Befragungen von Probanden methodisch nicht sinnvoll erscheinen.

Relevant für die Interpretation der referierten Ergebnisse ist außerdem die Betrachtung der fälschlich als fehlerhaft bewerteten sowie der nicht als fehlerhaft erkannten Sätze. Insgesamt werden nur 72,43% der „richtigen“ Sätze auch als solche erkannt, 2,25% der einen Fehler enthaltenden Sätze werden anders herum als akzentfrei bewertet. Der erste Wert verweist darauf, dass die Rater andere Aussprachemerkmale neben den intendierten Abweichungen von der Zielsprachennorm in ihre Bewertung einschließen. Der zweite Wert zeigt, dass nicht alle Ausspracheabweichungen vom Hörer überhaupt wahrgenommen werden.

Fehler 1 und 3, die in den Bewertungen der Rater am besten abschneiden, werden gleichzeitig mit Abstand am häufigsten gar nicht als Fehler erkannt. Nur 52 der 100 Hörer erkennen Fehler 1 in allen acht Items als Fehler, 72 von 100 gelingt dies bzgl. Fehler 3. Alle anderen Werte liegen über 90. In die Fehlerbewertung gehen allerdings, wie oben erläutert, nur die Rater ein, die den jeweiligen Fehler auch als solchen erkannt haben.

Auffällig ist ferner, dass die vier Sprecher signifikant unterschiedlich bewertet werden. Bonferroni-korrigierte t-Tests für verbundene Stichproben ($p = .05/6 = .008$) zeigen, dass alle Unterschiede in der durchschnittlichen Bewertung aller fehlerhaften Sätze zwischen den vier Sprechern signifikant sind.

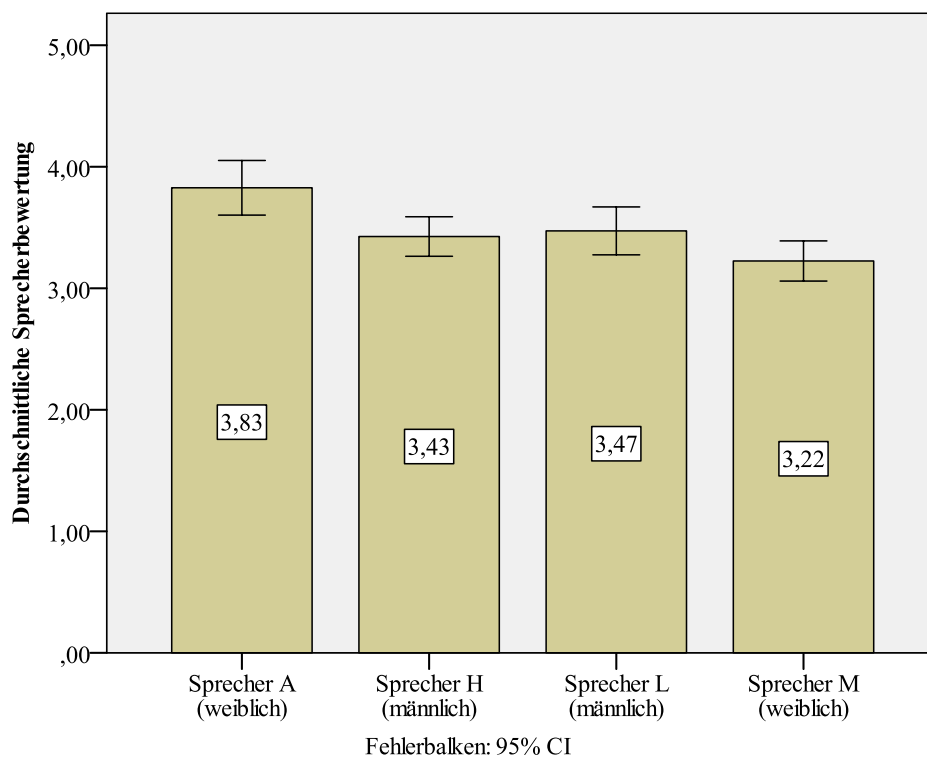


Abbildung 3: Durchschnittliche Bewertung aller Fehler enthaltenden Sätze über die vier Sprecherinnen und Sprecher. Diese Tatsache spiegelt sich interessanterweise auch in den freiwilligen Anmerkungen der Probanden zur Studie:

- *es kommt auch oft auf die Stimmfarbe an*
- *Ich hatte den Eindruck, dass die verschiedenen Stimmen und Sätze einen größeren Einfluss auf meine Bewertung ausgeübt haben als die jeweiligen Akzente.*

3.4 Diskussion

Im Folgenden sollen die dargestellten Ergebnisse daher einer kritischen Methodendiskussion unterzogen werden. So erscheint es bzgl. des vorgestellten Datensatzes methodisch problematisch, dass die Rater offensichtlich nicht nur die eingebauten Ausspracheabweichungen negativ sanktionieren, sondern darüber hinaus auch andere Äußerungsmerkmale, obwohl durch die Muttersprachlichkeit der Sprecher genau dieses Problem umgangen werden sollte. Eine mögliche Erklärung wäre die vergleichsweise Kürze der Stimuli. Während längere Äußerungen eine relativ klare Unterscheidung zwischen muttersprachlichen und nichtmuttersprachlichen Sprechern ermöglichen (vgl. Major 2007), scheint dies bei kürzeren Äußerungen schwieriger zu sein. Dies wird teilweise auch von den Probanden angemerkt:

- *Die Sätze der Hörbeispiele waren zu kurz.*
- *längere Satzbildung hätte manchmal bei der Bewertung für Erleichterung gesorgt.*

Darüber hinaus ist der Effekt der Sprecherstimmen auf die Bewertung zu nennen. Der Umstand, dass die vier Sprecher signifikant unterschiedlich bewertet worden sind, zeigt deutlich, dass die Stimmqualität einen Einfluss auf die Bewertung hat. Der Auswahl der Sprecher im Vorfeld der eigentlichen Untersuchung sollte also prinzipiell eine hohe Bedeutung beigemessen werden. Da die Sprecherstimmen in der vorliegenden Untersuchung jeweils zu gleichen Anteilen in die Bewertung der einzelnen phonetischen Abweichungen von der Standardaussprache eingehen, stellt dieser Umstand keinen Störfaktor bzgl. der Unterschiede zwischen den Ausspracheabweichungen dar. Auf die durchschnittliche Höhe der Bewertungen insgesamt könnten die Stimmqualitäten jedoch durchaus Einfluss genommen haben.

Bei der Interpretation der Ergebnisse zu bedenken sind auch mögliche Reihenfolgeeffekte der Befragung. Die Tatsache, dass die Rater zunächst Fehler bewerten, die häufig von L2-Deutsch-Sprechern mit französischer und/oder russischer L1 gemacht werden, und direkt im Anschluss nach ihren Einstellungen zu einem französischen bzw. russischen Akzent im Deutschen befragt werden, kann zu Extrembewertungen geführt haben, was sowohl einen Einfluss auf die häufige Nennung des Französischen und Russischen bei den besonders un-/beliebten Sprechern gehabt als auch zu einer Erhöhung der Korrelation zwischen Globalbewertung und Fehlerbewertungen geführt haben könnte. Eine Umkehrung der Fragenreihenfolge hätte das Problem jedoch noch verstärkt. Möglicherweise hätte das Einfügen von Distraktoren, die weder dem Französischen noch dem Russischen als L1 zugeordnet werden können, hier eine Lösung bieten können. Diese Vorgehensweise hätte den Test allerdings wiederum länger gemacht.

Ein grundsätzliches methodisches Problem aller Studien zur sozialen Akzeptanz unterschiedlicher Ausspracheformen besteht darin, dass es praktisch unmöglich ist, zwischen der Bewertung einer phonetischen Form und der Bewertung der kulturellen Zugehörigkeit eines Sprechers zu unterscheiden. Häufig werden bestimmte Abweichungen eben gerade deshalb negativ evaluiert, weil sie das Schibboleth einer weniger angesehenen Bevölkerungsgruppe sind. Andererseits könnte man diese Vermischung zweier Variablen (vgl. auch das *process model of native reactions to nonnative speech* von Doeleman (1998: 9) und Lindemann (2005)) auch als irrelevant betrachten, da es auf dasselbe hinausläuft, ob jemand, der auf einen Akzent negativ reagiert, dies tut, weil er dem Kulturkreis des Sprechers gegenüber negativ eingestellt ist oder weil er tatsächlich die Lautung nicht mag. Aufgrund der in jedem Fall negativen Konsequenzen für den Sprecher wäre diese Ausspracheabweichung sinnvoll zu beheben.

Mit dieser Vermischung in Zusammenhang steht sicherlich auch die Feststellung, dass in der vorliegenden Studie ganz andere Laute positiv bzw. negativ evaluiert werden als in anderen, z. T. im Forschungsüberblick zitierten Studien, in denen andere Sprachenpaarungen untersucht wurden. So schneidet /ŋg/ statt /ŋ/ in der vorliegenden Studie besonders positiv ab, die umgekehrte Ersetzung bei Dretzke (1987), der die L2 Englisch untersucht, hingegen besonders negativ. Dasselbe gilt für die regressive Assimilation an die Stimmhaftigkeit im Deutschen vs. Schwedischen als L2. Aus der Abhängigkeit der sozialen Bewertung devianter Aussprachemerkmale von bestimmten Gruppenzuschreibungen ergibt sich, dass empirische Studien im Bereich der Akzeptabilität unterschiedlicher Ausspracheformen nur Aussagekraft für die untersuchte Sprachenkonstellation haben. Während im Bereich der Verständlichkeit oder der Ökonomie in vielen Bereichen auch sprachenübergreifende Schlussfolgerungen gezogen werden können, ist dies hinsichtlich der sozialen Akzeptanz kaum möglich.

Solche, die Untersuchungsergebnisse beeinflussenden Einstellungen der Rater können zwischen unterschiedlichen Ratergruppen wiederum stark variieren. In der hier beschriebenen Studie wurden die untersuchten Ausspracheabweichungen von Bielefelder Studierenden im Alter zwischen 20 und 30 Jahren bewertet. Ratergruppen anderen Alters, anderer regionaler Herkunft oder auch anderen sozioökonomischen Statuses wären hier eventuell zu ganz anderen Einschätzungen gekommen.

Das besonders positive Abschneiden der Fehlertypen 1 und 3, die gleichzeitig am häufigsten gar nicht als Fehler erkannt werden, weist ferner darauf hin, dass auch die unterschiedliche perzeptuelle Auffälligkeit der phonetischen Abweichungen bei der Bewertung der sozialen Akzeptabilität eine Rolle spielen könnte, dass beide Variablen also konfundiert sein könnten. Fraglich ist somit, ob der gewählte Stimulus wirklich soziale Akzeptanz oder eine Mischung mehrerer Beurteilungskriterien, die z.B. auch Verständlichkeit und/oder Akzentuiertheit umfassen könnten (vgl. auch Llorca (2000) zum engen Zusammenhang zwischen Ausspracheniveau, Verständlichkeit und soziophonetischer Bewertung), misst. Da Wert auf eine globale Akzentbewertung gelegt wurde, wurde jedoch bewusst auf eine Aufspaltung der Abfrage in Teilaspekte der Bewertung verzichtet (vgl. 3.2). Vergleichsstudien mit alternativen Stimuli könnten hier genaueren Aufschluss bieten.

Schließlich ist die Frage der externen Validität von Studien wie der hier beschriebenen kritisch zu hinterfragen. Beim Design der vorliegenden Studie wurden verschiedene Anstrengungen unternommen, um die interne Validität der Studie zu maximieren. Durch Rückgriff auf eine neue Variante der *matched guise technique* wurde sichergestellt, dass den Ratern Stimulussätze eingespielt wurden, die von Muttersprachlern stammen und jeweils nur genau eine segmentale Abweichung enthalten. So sollte kontrolliert werden, dass nur diese Abweichungen bewertet wurden, auch wenn sich bei der Datenanalyse herausstellte, dass letztendlich auch andere Sprechercharakteristika in die Bewertung eingegangen sind. Diese Vorgehensweise, zusammen mit der Tatsache, dass Sätze vorgelesen wurden, wiederum mit dem Ziel, viermal genau dieselben Sätze zu erhalten, führt jedoch dazu, dass die Sätze künstlich werden. Die notwendige Kontrolle von Störeinflüssen steht hier im Widerstreit zur Forderung nach Natürlichkeit der Stimuli. Eine, unter Umständen relativ auffällige, Abweichung von der Zielnorm kontrastiert mit einem ansonsten vollkommen korrekten Satz, wie es in der Realität nur selten der Fall wäre. Ein Franzose, der das Wort *lange* mit französischem Akzent spräche, würde wohl in vielen Fällen nicht nur ein /g/ nach dem /ŋ/ einfügen, sondern auch das /a/ nasalieren und den Schwa-Laut runden, um nur ein Beispiel zu nennen. Diese Künstlichkeit fiel auch vielen der Rater auf, wie einige Anmerkungen zeigen:

- *da es sich um gestellte aufnahmen handelt war es schwieriger sie zu bewerten als wenn akzente einem in der realität begegenen*
- *die akzente klangen nicht echt, eher wie von deutschen extra nachgesprochen. das hat bei mir den effekt abgeschwächt, glaube ich*

Wäre stattdessen mit natürlichen Äußerungen nichtmuttersprachlicher L2-Sprecher gearbeitet worden, hätten diese jedoch ganz unterschiedliche Sätze mit einer Vielzahl parallel auftretender Abweichungen produziert, was es wiederum unmöglich gemacht hätte zu eruieren, was genau die Rater eigentlich bewertet haben. In diesem Fall wäre zwar die externe Validität der Ergebnisse gestiegen, die interne aber rapide gesunken (vgl. Riemer & Settineri (2010: 770) zum Wechselverhältnis von interner und externer Validität). Dieser methodologische Konflikt ist im Rahmen einer Einzelstudie wohl nicht vollständig zufriedenstellend zu lösen. Lediglich im Zusammenspiel methodisch ganz unterschiedlich vorgehender qualitativer und quantitativer Studien zu demselben Forschungsgegenstand können die Defizite beider Herangehensweisen einander längerfristig ausgleichen.

4. Zusammenfassung und Schlussfolgerungen

Zusammenfassend konnte gezeigt werden, dass unterschiedliche phonetische Abweichungen von der L2-Norm signifikant unterschiedlich positiv bzw. negativ bewertet werden. Dabei stechen aus einem breiten Mittelfeld einzelne Phänomene nach oben bzw. unten hervor, was es didaktisch sinnvoll erscheinen lässt, den Aspekt der sozialen Akzeptanz in die Progression einer Ausspracheschulung einzubeziehen. Besonders negativ evaluierte Abweichungen sollten dabei prioritär behandelt werden, wobei die soziale Akzeptabilität der Aussprachefehler selbstverständlich nur ein Kriterium zur Etablierung einer sinnvollen Progression neben anderen ist. Zu nennen wären in diesem

Julia Settineri (2011), Soziale Akzeptanz unterschiedlicher Normabweichungen in der L2-Aussprache Deutsch. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 16: 2, 66-80. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-16-2/beitrag/Settineri.pdf>.

Zusammenhang vor allem der Grad der Gefährdung der Verständlichkeit, die Frequenz des Fehlers in der Zielsprache, aber auch in der Lernergruppe und selbstverständlich die individuellen Lernziele der Lerner (vgl. Settineri 2008: 9), wobei letzterer Aspekt sowie das Kriterium der Verständlichkeit sicherlich auch weiterhin zentral sind.

Untermauert wird dieses Desiderat auch durch den Befund, dass die Bewertung der einzelnen Ausspracheabweichungen tatsächlich mit der globalen Bewertung eines Akzents korreliert. Zusammengefasst mit dem Tatbestand, dass die Bewertung von Ausspracheabweichungen auch damit zusammenhängt, auf welche Herkunft des Sprechers sie schließen lassen, könnte es ein Ziel von Ausspracheschulung sein, gerade diese typisierenden Ausspracheabweichungen prioritär zu beheben. Dies entspricht dem Wunsch vieler Lerner, endlich nicht mehr im ersten Satz als „aus dem Land X stammend“ identifiziert zu werden.

Dazu sind jedoch zahlreiche weitere quantitative sowie qualitative, je sprachenpaarspezifische Untersuchungen vonnöten, die genau untersuchen, welche Abweichungen stigmatisierend wirken und welche eher positiv evaluiert werden. Interessant wäre sicherlich auch, nicht nur muttersprachliche Rater, sondern auch nichtmuttersprachliche in Untersuchungen einzubeziehen sowie verschiedene Altersklassen, sozioökonomische Faktoren usw., um so möglicherweise vorhandenen universelleren Bewertungsmustern auf die Spur zu kommen. Außerdem ist eine Methodenvielfalt weiterer Studien erstrebenswert, insbesondere um unvermeidliche Validitätsprobleme einzelner Studien zu kompensieren und Methodeneffekte der Studien besser einschätzen zu können. Zum Einsatz kommen sollten dabei v.a. indirekte Befragungsmethoden, da – wie die oben genannten Antwortbeispiele zeigen – eine direkte Befragung zu phonetischen Merkmalen wenig erfolgversprechend ist. Ratinguntersuchungen mit Dubletten, wie in der vorliegenden Studie, aber auch Designs, in denen Probanden gebeten werden zu imitieren, was ihnen an einer nichtmuttersprachlichen Aussprache besonders gut oder eher weniger gefällt, so dass die reproduzierten Merkmale im Anschluss von einem phonetisch geschulten Untersuchungsleiter in phonetische Fachsprache „übersetzt“ werden können, wären beispielsweise mögliche Wege.

Britisches Englisch, Spanisch oder auch Italienisch als Erstsprache könnten einen Einstieg in die Erforschung eher positiv bewerteter Akzente im Deutschen als L2 bieten, Türkisch, Polnisch, Chinesisch oder Arabisch in die Gruppe der eher negativ bewerteten. Dass es sich beim Thema *Soziale Akzeptanz* um ein methodologisches Minenfeld handelt, sollte aufgrund der Relevanz des Themas für L2-Erwerber nicht von seiner Erforschung abschrecken, denn, wie eine der Raterinnen es auf den Punkt bringt: „Leider ist es wahr, dass ein negativer Akzent für die subjektive Beurteilung einer Person manchmal ausschlaggebend ist.“

Literatur

- Anderson-Hsieh, Janet; Johnson, Ruth & Koehler, Kenneth (1992), The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42, 529-555.
- Cohrs, Silke (2007), Prosodische Interferenzen ungarischer Deutschlerner. Wirkung und Akzeptanz bei Muttersprachlern. *Deutsch als Fremdsprache* 44, 97-102.
- Cunningham-Andersson, Una (1997), Native speaker reactions to non-native speech. In: James, Allan & Leather, Jonathan (Hrsg.) (1997), *Second-Language Speech. Structure and Process*. Berlin, New York: de Gruyter, 133-144.
- Derwing, Tracey M. & Munro, Murray J. (1997), Accent, intelligibility, and comprehensibility. Evidence from four L1s. *Studies in Second Language Acquisition* 19, 1-16.
- Doeleman, Rianne (1998), *Native Reactions to Nonnative Speech*. Tilburg: Tilburg University Press.
- Dretzke, Burkhard (1987), Beurteilung von Aussprache Fehlern im Englischen. *Die neueren Sprachen* 86, 507-517.
- Grotjahn, Rüdiger (1998), Ausspracheunterricht. Ausgewählte Befunde aus der Grundlagenforschung und didaktisch-methodische Implikationen. *Zeitschrift für Fremdsprachenforschung* 9, 35-83.

- Gut, Ulrike (2003), *Non-native speech rhythm in German. Proceedings of the XVth ICPHS Conference, Barcelona*, 2437-2440.
- Hellwig-Fabián, Inessa (2007), *Deutsch mit ausländischem Akzent. Eine empirische Studie zu Einstellungen junger Deutscher gegenüber Sprechern mit ostslavischer Muttersprache*. Frankfurt/M. u. a.: Peter Lang.
- Hirschfeld, Ursula (1994), *Untersuchungen zur phonetischen Verständlichkeit Deutschlernender*. Frankfurt/M.: Wissenschaftliche Buchhandlung Theo Hector.
- Hirschfeld, Ursula (1995), Phonetische Merkmale in der Aussprache Deutschlernender und deren Relevanz für deutsche Hörer. *Deutsch als Fremdsprache* 32, 177-183.
- Hirschfeld, Ursula (2001), Vermittlung der Phonetik. In: Helbig, Gerhard; Götze, Lutz; Henrici, Gert & Krumm, Hans-Jürgen (Hrsg.) (2001), *Deutsch als Fremdsprache. Ein internationales Handbuch*. 2. Halbband. Berlin, New York: de Gruyter, 872-879.
- Jilka, Matthias (2000), *The Contribution of Intonation to the Perception of Foreign Accent. Identifying Intonational Deviations by Means of F0 Generation and Resynthesis*. Dissertationsschrift zur Erlangung der Würde eines Doktors der Philosophie (Dr. phil.), eingereicht bei der Fakultät für Philosophie der Universität Stuttgart.
- Kaltenbacher, Erika (1998), Zum Sprachrhythmus des Deutschen und seinem Erwerb. In: Wegener, Heide (Hrsg.) (1998), *Eine zweite Sprache lernen. Empirische Untersuchungen zum Zweitsprachenerwerb*. Tübingen: Narr, 21-38.
- Koch, Peter & Oesterreicher, Wulff (1985), Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36, 15-43.
- Kohler, Klaus J. (2001), Variability of opening and closing gestures in speech communication. In: Kohler, Klaus (Hrsg.) (2001), *Sound Patterns in German Read and Spontaneous Speech. Symbolic Structures and Gestural Dynamics*. Kiel: Institut für Phonetik und digitale Sprachverarbeitung, 33-96.
- Köhne, Annik Brigitte (2009), *Phonetische Interferenzen türkischer DeutschlerInnen – Beurteilung durch deutsche MuttersprachlerInnen*. Humboldt Universität zu Berlin: Masterarbeit im Fach Deutsch als Fremdsprache.
- Lambert, Wallace E.; Hodgeson, R. C.; Gardner, Robert C. & Fillenbaum, S. (1960), Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60, 44-51.
- Lindemann, Stephanie (2005), Who speaks “broken English”? US undergraduates’ perceptions of non-native English. *International Journal of Applied Linguistics* 15, 187-212.
- Llurda, Enric (2000), Effects of intelligibility and speaking rate on judgements of non-native speakers’ personalities. *IRAL* 38, 289-299.
- Magen, Harriet S. (1998), The perception of foreign-accented speech. *Journal of Phonetics* 26, 381-400.
- Major, Roy C. (2007), Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition* 29, 539-556.
- Munro, Murray J. & Derwing, Tracey M. (1995), Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38, 289-306.
- Richter [jetzt Settineri], Julia (2008), *Phonetische Reduktion im Deutschen als L2. Eine empirische Querschnittsstudie*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Riemer, Claudia & Settineri, Julia (2010), Empirische Forschungsmethoden in der Zweit- und Fremdsprachenerwerbsforschung. In: Krumm, Hans-Jürgen; Fandrych, Christian; Hufeisen, Britta & Riemer, Claudia (Hrsg.) (2010), *Deutsch als Fremd- und Zweitsprache*. Berlin: de Gruyter, 764-781.
- Settineri, Julia (2008), Teaching pronunciation. In: Gramley, Stephan & Gramley, Vivian (Hrsg.) (2008), *Bielefeld Introduction to Applied Linguistics. A Course Book*. Bielefeld: Aisthesis Verlag, 1-12.
- Wenk, Brian J. (1985), Speech rhythms in second language acquisition. *Language and Speech* 28, 157-175.

Young-Scholten, Martha (1993), The L2 acquisition of informal speech in German. In: Kettemann, Bernhard & Wieden, Wilfried (Hrsg.) (1993), *Current Issues in European Second Language Acquisition Research*. Tübingen: Narr, 111-123.

Anmerkungen

¹ Die hier beschriebene Studie wurde aus Nachwuchsmitteln der Universität Bielefeld finanziert. Mein Dank gilt außerdem Johanna Müller für ihren Einsatz bei der Vorbereitung und Durchführung der Datenerhebung sowie Alexis Feldmeier, Nazan Gültekin-Karakoç und den anonymen Reviewern für ihre zahlreichen hilfreichen Anmerkungen zu früheren Versionen dieses Aufsatzes.

² Unterschieden werden kann zwischen der Akzentuiertheit (*degree of accent, accentedness*) und der phonetischen Verständlichkeit von Äußerungen. Verständlichkeit kann weiter objektiv (*intelligibility*), z. B. über Transkription, oder subjektiv (*comprehensibility*), z. B. über Rating-Skalen, erhoben werden (Derwing & Munro 1997: 2; Munro & Derwing 1995: 290f). Auch wenn die drei Dimensionen korrelieren können, so muss dies nicht immer der Fall sein: „[C]omprehensibility and accent are partially-independent dimensions [...]. Therefore, the notion that foreign accent ‘reduction’ automatically entails improved comprehensibility is quite incorrect.” (Munro & Derwing 1995: 302)

³ Diese Hypothese wird aufgenommen, da Frauen im Vergleich zu Männern häufig eine stärkere Normorientierung zugeschrieben wird (vgl. Richter 2008: 95f) und da es außerdem möglich wäre, dass Frauen und Männer voneinander abweichende Aussprachemerkmale präferieren.

⁴ Da für das Deutsche als L2 bislang meines Wissens noch keine großen, sprachenübergreifenden Studien zur sozialen Akzeptanz phonetischer Abweichungen von der Zielsprache vorliegen, werden diese beiden Sprachen auf Grundlage von Erfahrungen mit internationalen Studierenden in Ausspracheschulungsseminaren an der Universität Bielefeld ausgewählt. Während insbesondere russischsprachige Studierende vergleichsweise häufig äußern, unter ihrem Akzent zu leiden, geben französischsprachige Studierende häufig an, gar keine Notwendigkeit zu sehen, ihren Akzent zu reduzieren, da dieser insgesamt sehr positiv sanktioniert werde.

⁵ Die Testsätze werden von den vier Sprechern in einer Sprecherkabine mit und ohne Fehler ausgesprochen. Anschließend werden mithilfe der Softwares *PRAAT* und *Audacity* die Silben, die den jeweiligen Fehler enthalten, in die fehlerfreien Testsätze hineingeschnitten, um so die notwendigen Dubletten zu generieren.

⁶ Im Rahmen dieser Pilotstudie werden relativ willkürlich ausschließlich segmentale Abweichungen ausgewählt, da diese instrumentalphonetisch leichter zu manipulieren sind. Folgestudien sollten selbstverständlich auch Suprasegmentalia umfassen bzw. idealerweise alle möglichen Typen phonetischer Abweichungen von der Zielsprache. Weniger frequent ist Fehler Nr. 7, der in Analogie zu Fehler Nr. 8 aufgenommen wurde.

⁷ Die Online-Maske wurde mittels *Medialab* erstellt und so eingestellt, dass Weiterklicken zur nächsten Frage nicht möglich ist, ohne die vorherige beantwortet zu haben. So wurde die Entstehung fehlender Werte von vornherein verhindert. Darüber hinaus erstellt das Programm aus den Antworten der Probanden automatisch eine SPSS-Matrix, so dass Datenübertragungsfehler ebenfalls ausgeschlossen werden.

⁸ Die dem Test vorangestellten Instruktionen verweisen außerdem erläuternd darauf, dass es um die Bewertung sozialer Akzeptabilität gehen soll:

„Wir alle beurteilen, ohne es zu wollen oder uns in jedem Fall bewusst zu machen, andere Menschen nach äußeren Merkmalen, wie z.B. ihrer Kleidung, ihrer Haltung und insbesondere auch nach ihrer Art zu sprechen. Auf Sprecher mit nicht-deutscher Muttersprache trifft dies in besonderem Maße zu: Manche Akzentmerkmale wirken auf uns positiv, andere negativ.

Auch wenn das natürlich Vorurteile sind, die nichts mit den wahren Eigenschaften des Sprechers zu tun haben, bekommen Nichtmuttersprachler positive oder negative Reaktionen auf ihre Aussprache zu spüren. Ein Ziel von

Ausspracheschulung ist es daher, die deutsche Aussprache von Nichtmuttersprachlern so zu verbessern, dass sie von Muttersprachlern sozial möglichst positiv sanktioniert wird.

Dazu benötigen wir Informationen darüber, welche Akzentmerkmale zu einer eher positiven Bewertung des Sprechers und welche zu einer eher negativen Bewertung des Sprechers führen.“

⁹ Aus den acht Bewertungen, die jeder Rater pro Fehlertyp abgibt, wird ein Mittelwert gebildet. Nicht als Fehler erkannte Items werden benutzerdefiniert fehlend gesetzt, was zur Konsequenz hat, dass der entsprechende Rater aus der Bewertung dieses Fehlertyps ausgeschlossen wird.

¹⁰ Anstelle eines Signifikanzniveaus von 5% wird ein Signifikanz-Niveau von $.05/36 = .001$ angelegt.

¹¹ Alle Probandenäußerungen werden buchstabengetreu zitiert.