

Nonresponse und Interviewer-Erfolg im Telefoninterview

Empirische Untersuchungen zum Einfluss stimmlicher Eigenschaften der Interviewer

Nonresponse in CATI-Surveys

An Empirical Study on the Effects of Interviewers' Vocal Characteristics

Leander Steinkopf, Gerrit Bauer und Henning Best

Zusammenfassung

Basierend auf der Messung von objektiven Eigenschaften und subjektiven Bewertungen der Stimmen von 56 weiblichen Telefoninterviewern wird der Zusammenhang zwischen Stimmeigenschaften und dem Interviewerfolg empirisch untersucht. Theoretisch lässt sich argumentieren, dass die Interviewer-Stimmen einerseits Nutzenerwartungen der Respondenten beeinflussen können, indem sie z. B. belohnend wirken oder Seriosität vermitteln. Ausgehend von dual-process-Theorien ist außerdem zu erwarten, dass die Stimme einen situativen Hinweisreiz darstellt, der – ohne rationale Abwägung – Kooperationsnormen aktivieren kann und hierdurch die Wahrscheinlichkeit der Teilnahme am Interview erhöht. Empirisch zeigt sich, dass subjektiv eingeschätzte Stimmerkmale in keinem Zusammenhang mit der Erfolgsquote der Interviewer stehen. Objektive Merkmale der Stimme hingegen, insbesondere die Stimmhöhe, haben substanziellen Einfluss auf die Erfolgsquote. Es ist jedoch wichtig zu bemerken, dass der Zusammenhang zwischen der Stimmhöhe oder der Sprechgeschwindigkeit und dem Interviewerfolg nicht linear ist, sondern einen umgekehrt U-förmigen Verlauf hat. Wir finden empirisch, dass Interviewerinnen mit durchschnittlicher Stimmhöhe und durchschnittlicher Sprechgeschwindigkeit die höchsten Erfolgsquoten aufweisen.

Abstract

This study examines nonresponse in telephone surveys. Our analysis relates response rates to interviewer voice characteristics. Drawing on theory we argue that interviewer voices can affect the respondents' utility expectations, e.g. by indicating integrity, or by providing an intrinsic reward. Based on dual-process theories, one can additionally expect the voice to be a situative cue. We use contact data generated in a CATI survey and supplement these data with information on interviewers' voices. The survey was conducted during 2007 and 2008 at the University of Mannheim, Germany. To obtain metadata, we recorded the interviewers' voices and generated phonetic measures of vocal characteristics. This enabled us to study the determinants of interviewer effectiveness and nonresponse with a special focus on objective voice characteristics (pitch, speech rate, etc.). Additionally, we accounted for a variety of subjective voice attributes (friendliness, trustworthiness, etc.). The results show that objective vocal characteristics have greater explanatory power than subjective indicators. Furthermore, the objective voice characteristics are related to the interviewers' productivity in a nonlinear way.

1 Einleitung

Telefonische Befragungen haben sich in den Sozialwissenschaften und insbesondere der Markt- und Meinungsforschung längst zu einem Standardinstrument entwickelt. CATI-Befragungen bieten insgesamt eine Reihe von Vorteilen, etwa die im Vergleich zu persönlich-mündlichen Interviews deutlich geringeren Kosten und die einfachere Feldsteuerung. Allerdings bleiben die Ausschöpfungsquoten telefonischer Befragungen nach wie vor hinter Face-to-Face durchgeführten Studien zurück, wenngleich sich durch Telefonsurveys bessere Response-Raten erzielen lassen als durch Websurveys (vgl. z. B. Fricker 2005; Roster et al. 2004). Die Ausschöpfungsquote einer Befragung kann grundsätzlich als aggregierte Folge von Teilnahmeentscheidungen der potentiellen Befragten verstanden werden. Aus Untersuchungen zum Befragtenverhalten hat sich unterdessen ein umfangreiches Forschungsfeld entwickelt, denn Nonresponse ist – gleich welche Befragungsmethode genutzt wird – stets ein wesentliches Problem sozialwissenschaftlicher Umfrageforschung. Weil Ausfälle häufig selektiv erfolgen, können die Ausfallgründe erstens mit den die Forscher interessierenden Merkmalen korreliert sein, zweitens verringert sich die Repräsentativität einer Befragung. Somit treten mit großer Wahrscheinlichkeit Verzerrungen auf (Nonresponse bias, siehe z. B. Groves 1989; Schnell 1997). In der Praxis der Umfrageforschung besteht daher das Ziel, den Ursachen des Ausfalls ganzer Erhebungseinheiten nachzugehen und Nonresponse zu reduzieren.

Unit Nonresponse, also der vollständige Ausfall einer Erhebungseinheit (Groves et al. 2001; Schnell 1997), kann seine Ursache(n) auf verschiedenen Ebenen haben: der soziale und kulturelle Kontext, die Erhebungsorganisation und deren Auftraggeber, das Erhebungsdesign, der Interviewer und die Erhebungseinheit selbst (de Leeuw/de Heer 2001). Trotz der umfangreichen Forschungsarbeiten zum Nonresponse bleiben dennoch bestimmte, die Befragungssituation kennzeichnende Merkmale, bisher wenig beachtet: Merkmale der Stimme des Interviewers. Im deutschsprachigen Raum liegt hierzu lediglich eine Studie vor (Hüfken/Schäfer 2003), und auch international ist der Forschungsstand unbefriedigend. Dies ist überraschend, da in der Kommunikationsforschung unumstritten ist, dass im Gespräch sowohl verbale Inhalte als auch nonverbale Ausdrücke – zu denen stimmliche Merkmale üblicherweise gezählt werden – übertragen werden. Schon Goffman unterscheidet zwischen diesen beiden Informationskanälen, dem „Ausdruck, den er [der Akteur] sich selbst gibt, und der Ausdruck, den er ausstrahlt“ (Goffman 1969: 6). Unser Beitrag setzt an letzterem an und analysiert systematisch den Einfluss sowohl objektiv gemessener als auch subjektiv bewerteter Stimmmerkmale von Interviewern auf die Teilnahme an telefonischen Befragungen. Während in Face-to-Face-Surveys weitere

physische Interviewereigenschaften, etwa körperliche Attraktivität, ein gepflegtes Erscheinungsbild oder eine ansprechende Mimik und Gestik das Antwortverhalten und vor allem die Kooperationsbereitschaft des Befragten beeinflussen können, sind diese potentiellen Faktoren in Telefoninterviews ausgeschaltet. Der Befragte nimmt lediglich die Stimme des Interviewers wahr.¹

Wir werden im Folgenden zunächst die theoretischen Mechanismen beschreiben, die einen Einfluss der Stimme von Interviewern auf die Teilnahme des Respondenten nahelegen. Sodann geben wir einen Überblick über den bisherigen Forschungsstand und stellen das Design unserer Studie vor. Es folgt eine Darstellung der empirischen Ergebnisse, die wir schließlich vor dem Hintergrund der Theorien zum Befragtenverhalten diskutieren.

2 Theorie und Forschungsstand

2.1 Theorie

Eine klassische Erklärung der Teilnahme an Befragungen findet sich bei Esser (1986), der die Entscheidung als eine rational geleitete Wahlhandlung modelliert. Ein Befragter wird dann an einer Untersuchung teilnehmen, wenn sich daraus ein positiver Erwartungsnutzen ergibt. Dieser Nutzen wird von den direkten Kosten einer Befragung, den Opportunitätskosten und eventuellen Belohnungen mitbestimmt. In diesem Sinne ist etwa die Wahrscheinlichkeit, dass eine Person mit wenig Zeit an einer Befragung teilnimmt, sicherlich eher gering, da sie hohe Opportunitätskosten einkalkulieren muss (vgl. Esser 1986: 39). Für eine Person hingegen, die Sozialkontakte vermisst und Gesprächsbedarf hat, ist ein Interview demnach mit sozialem Nutzen verbunden.²

1 In telefonischen Befragungen sind selbstverständlich auch Überzeugungsstrategien der Interviewer von Bedeutung. Diese sind jedoch nicht Gegenstand des vorliegenden Beitrags.

2 Bezogen auf die Teilnahme an Befragungen hat die Rational-Choice-Theorie durch Groves und Couper (1998) sowie Groves et al. (2000) eine Erweiterung erfahren. Unter der Bezeichnung „Leverage-Salience“-Theorie wird der Bedeutung von Überzeugungsstrategien der Interviewer Rechnung getragen. Die Theorie geht also davon aus, dass bestimmte Attribute die Kosten-Nutzen-Entscheidung umso stärker beeinflussen, desto salienter sie sind. Interviewer versuchen gleich zu Beginn der Kontaktphase, die Salienz bestimmter Merkmale der Umfrage zu erhöhen. Sie betonen dabei solche Merkmale der Umfrage, von denen sie denken, dass sie vom jeweiligen Respondenten positiv bewertet werden (etwa Thema, Incentive, Auftraggeber etc.). Wird eine möglichst vollständige Erklärung von Nonresponse angestrebt, so sind die Überlegungen von Groves et al. von großer Bedeutung. Da in der vorliegenden Studie gezielt Hypothesen zum Einfluss stimmlicher Merkmale abgeleitet und überprüft werden sollen, sind Überzeugungsstrategien hier nicht Untersuchungsgegenstand.

Doch in welchem Zusammenhang steht die rationale Entscheidung zur Interviewteilnahme mit Eigenschaften der Stimme des Interviewers? Bereits Esser (1986) merkt an, dass „die Entscheidung zur Teilnahme meist aus einer Art Indifferenz heraus erfolgt“. Speziell in einer solchen Situation seien dann Einflüsse „relativ periphere[r] und zufällige[r] Faktoren“ (Esser 1986: 39) zu erwarten. Sowohl die direkten Kosten als auch der Nutzen eines Interviews sind mit stimmlichen Eigenschaften des Interviewers verknüpft. Wir halten zwei Kosten/Nutzen-Aspekte für besonders diskussionswürdig: Erstens kann ein Gespräch mit einer Person mit sympathischer, attraktiver oder freundlicher Stimme durchaus als angenehm empfunden werden und entsprechend für den Respondenten mit Nutzen verbunden sein. Im Gegensatz hierzu ist ein Gespräch mit einer unsympathischen, schrillen Stimme mitunter unangenehm und somit kostenträchtig (vgl. Feldstein et al. 2001; Scherer 1978).

Zweitens sind subjektive Risiken der Interviewteilnahme zu beachten – insbesondere Befürchtungen hinsichtlich einer Entanonymisierung der Befragungsergebnisse. Da telefonische Befragungen in aller Regel als „cold calls“, also ohne Vorankündigung, durchgeführt werden, hat der Befragte keine oder nur geringe Informationen über den Auftraggeber einer Studie, seine Vertrauenswürdigkeit und die Einhaltung von Datenschutzbestimmungen. Die Stimme des Interviewers kann in dieser Situation Seriosität vermitteln und Kompetenz ausstrahlen, Vertrauen wecken, oder eben auch unprofessionell wirken und verunsichern. Beispielsweise berichten Apple et al. (1979), dass höhere Stimmen eher als nervös empfunden werden, langsam sprechende Personen hingegen als tendenziell weniger glaubwürdig und überzeugend. Neuere Studien deuten darauf hin, dass stimmliche Merkmale des Empfängers bei der Beurteilung des Senders als Referenz herangezogen werden und so Einschätzungen von Kompetenz und Vertrauenswürdigkeit entstehen (Feldstein et al. 2001). Folglich werden durch nonverbal vermittelte Persönlichkeitsmerkmale die erwarteten Kosten einer Interviewteilnahme beeinflusst, da sich auf Seiten des Respondenten die subjektiv eingeschätzte Wahrscheinlichkeit eines Datenmissbrauchs verändert. Insgesamt kann argumentiert werden, dass die rationale Entscheidung zur Interviewteilnahme von subjektiven Eigenschaften, die der Stimme zugeschrieben werden, mit bestimmt werden kann.

In neueren Entwicklungen der Entscheidungstheorie wird jedoch argumentiert, dass viele Entscheidungen nicht kalkuliert gefällt werden, sondern vielmehr einer Routine folgen oder spontan getroffen werden – insbesondere, wenn Kosten und Nutzen einer Handlung eher gering sind. Wichtige Anstöße für die Analyse solcher Entscheidungen geben sozialpsychologische dual-process-Theorien (siehe z. B. Chaiken/Trope 1999) und das Modell der Frame Selektion von Esser (1990) und Kroneberg (2007).

Komplexe Erwägungen von Opportunitätskosten oder Seriositätseinschätzungen sind in diesen theoretischen Modellen nur bedingt zu erwarten. Das „Elaborationswahrscheinlichkeitsmodell der Einstellungsänderung“ (ELM) von Petty und Cacioppo (1984) sagt einen starken Einfluss peripherer Merkmale – so genannter Hinweisreize oder „situational cues“ – voraus, wenn die Akteure nur ein geringes Interesse am jeweiligen Handlungsgegenstand haben. Effekte stimmlicher Merkmale wurden vor diesem theoretischen Hintergrund beispielsweise im Bereich der Werbewirkungsforschung untersucht (Gelinac-Chebat/Chebat 2001). Angewandt auf den Gegenstand des Telefoninterviews ist auch in einer solchen low-involvement Situation von einem Effekt der Sprechstimme auszugehen. Im Gegensatz zum Rational-Choice-Modell des Nonresponse wird jedoch weder eine explizite Nutzenkalkulation angenommen, noch wird erwartet, dass der Akteur die situativen Reize einer aufmerksamen Prozessierung unterzieht. In telefonischen Befragungen bedeutet dies konkret, dass sich potentielle Teilnehmer in den ersten Sekunden der Kontaktphase häufig nicht intensiv mit dem Inhalt des Einleitungstextes befassen werden. Stattdessen oder zusätzlich orientieren sie sich an Hinweisreizen. Im Telefongespräch können solche Stimuli insbesondere von Stimme und Sprache des Interviewers ausgehen (Groves et al. 1992). Diese Hinweisreize können einerseits durch subjektiv zugeschriebene Eigenschaften der Stimme (kompetent, sympathisch, etc.) gesetzt werden, wie sie bereits im Abschnitt zur Rationalität der Interviewteilnahme diskutiert wurden. Darüber hinaus ist von einem Einfluss objektiver Stimmeigenschaften auszugehen (Tonhöhe, Lautstärke, etc.). Es ist anzunehmen, dass auch das Bewerten einer Stimme nach den genannten subjektiven Zuschreibungen kognitiven Aufwand erfordert. In einer peripheren Situation wird der Akteur jedoch versuchen, seinen kognitiven Aufwand zu minimieren. Der Kooperationsstimulus würde in diesem Sinne nicht von einer explizit „sympathischen“ oder „angenehmen“ Stimme ausgehen, sondern vielmehr würde die Stimme als solche einen Hinweisreiz setzen, ohne hierbei explizit bewertet zu werden. In der sozialpsychologischen Forschung wird argumentiert, dass Stereotype über Stimmatraktivität bestehen und diese in unbewussten Prozessen als Heuristik für eine schnelle Beurteilung herangezogen werden (Bartsch 2008). Da Stereotype nicht notwendigerweise eine akkurate Beschreibung einer Person darstellen, prägen sie besonders den ersten Eindruck (Zuckerman et al. 1990). Dies bedeutet, dass auch objektive stimmliche Eigenschaften ohne eine subjektive Evaluation bzw. ohne eine bewusste Zuschreibung von Eigenschaften verhaltenswirksam werden können. In diesem Sinne wären gerade in Telefoninterviews direkte Effekte z. B. der Stimmlage, der Tonhöhe oder der Sprechgeschwindigkeit zu erwarten, da die Stimme in dieser Situation der einzige Hinweisreiz ist und zunächst keine weiteren Informationen über den Interviewer vorliegen.

2.2 Forschungsstand

Obwohl Einflussfaktoren auf die Teilnahme an Interviews vielfach untersucht wurden (z. B. kürzlich in dieser Zeitschrift von Schnauber/Daschmann 2008 oder Weidmann et al. 2008), sind zur Wirkung von Interviewerstimmen auf das Verhalten der Befragten nur wenige Studien erschienen (Groves et al. 2008; Hufken/Schäfer 2003; Oksenberg/Cannell 1988; Oksenberg et al. 1986; Sharf/Lehman 1984; van der Vaart et al. 2005).

Die grundlegende Arbeit stammt von Oksenberg et al. (1986). Die Autoren messen objektive wie subjektive stimmliche Eigenschaften von insgesamt 12 weiblichen Interviewern mit besonders hohen bzw. niedrigen Verweigerungsraten und vergleichen diese Extremgruppen miteinander. Oksenberg et al. finden, dass sich erfolgreiche und erfolglose Interviewer stärker hinsichtlich objektiver stimmlicher Eigenschaften (insbes. Tonhöhe und Lautstärke) als bezüglich subjektiver Zuschreibungen (Attraktivität und positive Einstellung) unterscheiden. Für die Eigenschaften Sprachkompetenz, Sprechgeschwindigkeit, Betonung, Einschätzung als kompetent, das Alter und die soziale Schicht zeigten sich keine oder nur sehr geringe Effekte. Die Ergebnisse von Oksenberg et al. (1986) sollten jedoch aufgrund der sehr geringen Fallzahl und des Extremgruppenvergleichs mit Vorsicht betrachtet werden. In einer Follow-Up-Studie überprüfen Oksenberg und Cannell (1988) die Ergebnisse an einer etwas größeren Stichprobe und ergänzen den Extremgruppenvergleich um eine Korrelationsanalyse. Die Ergebnisse sind jedoch recht uneinheitlich und schwer zu interpretieren. So zeigt sich in der Korrelationsanalyse (n=25) ein negativer Einfluss der Stimmhöhe auf die Erfolgsquote sowie ein positiver Effekt von Sprechgeschwindigkeit, Lautstärke und der Einschätzung als vertrauenswürdig. Auch im Extremgruppenvergleich (getrennt durchgeführt an drei Samples von je 2*6 Interviewerinnen) führen lauterer und schnelleres Sprechen sowie eine Einschätzung der Interviewerin als „kompetent“ oder „vertrauenswürdig“ zu geringeren Verweigerungsraten. Im Gegensatz zur Korrelationsanalyse zeigt sich hier jedoch teilweise ein positiver Einfluss der Stimmhöhe. Freundlichkeit, Interessiertheit, Natürlichkeit, die Betonung³ oder die Einschätzung der Stimme als angenehm haben keinen Effekt.

Mit Effekten objektiv messbarer stimmlicher Eigenschaften, vor allem der Fundamentalfrequenz, der Intonation und der Länge von Sprachpausen befassten sich bereits Sharf und Lehman (1984). Ihre Analyse bezieht sich auf sechs weibliche

3 Die Betonung wurde in dieser Studie einerseits als subjektive Einschätzung erhoben, andererseits als Variation in der Fundamentalfrequenz gemessen. Die Ergebnisse waren unsystematisch (mal positiv, mal negativ, mal Null).

Interviewer, wobei positive Effekte einer höheren Stimme und einer größeren Variation der Stimmhöhe gefunden werden. Auch Interviewer mit einer flüssigeren Sprechweise mit kürzeren Sprachpausen hatten höhere Erfolgsquoten, und eine fallende Sprachmelodie erwies sich in diesem Kontext förderlich. Diese älteren Studien leiden unter der geringen Stichprobengröße, dem Extremgruppenvergleich (selection on the dependent variable) und damit einhergehenden Verzerrungen sowie der Nichtverwendung von multivariaten Verfahren.

Neuere Studien (Groves et al. 2008; Hüfken/Schäfer 2003; van der Vaart et al. 2005) versuchen, die Defizite teilweise auszugleichen. Hüfken und Schäfer (2003) verwenden eine Variante des Extremgruppenvergleichs. Aus ihrem Sample von 51 Interviewern wählen sie 10 mit hohen Erfolgsquoten, 10 mit niedrigen und 4 mit mittleren Quoten aus. Für die Stimmen dieser 24 Interviewer erheben sie subjektive Ratings zu verschiedenen stimmlichen und persönlichen Merkmalen für jeden einzelnen Interviewversuch, den die Interviewer durchführen. Hierdurch ergibt sich ein geclustertes Sample von $n=219$ (leider ignorieren die Autoren die Clustering bei ihren Analysen, behandeln die 219 Interviewversuche also als unabhängige Beobachtungen, was zu einer Unterschätzung der Standardfehler führt). Empirisch zeigen sich in den multivariaten Logit-Modellen ein signifikant negativer Einfluss der Sprechgeschwindigkeit sowie ein positiver Effekt „enthusiastischen Auftretens“. Der Koeffizient der Variation der Stimmhöhe ist negativ und in einzelnen Modellen signifikant. Die anderen Eigenschaften (Stimmhöhe, Lautstärke, Betonung, Wortfluss, angenehme Stimme, Interesse, Höflichkeit u. a.) haben keinen signifikanten Einfluss. Es muss jedoch beachtet werden, dass Hüfken und Schäfer Interviewer beiderlei Geschlechts in ihrer Stichprobe haben, dies in den Modellen jedoch nicht adäquat berücksichtigen (z. B. bei der Stimmhöhe). Zudem verwenden sie ausschließlich subjektive Ratings für Stimmeigenschaften, die technisch auch objektiv messbar wären.

Van der Vaart et al. (2005) untersuchen die Effekte objektiver und subjektiver Stimmeigenschaften von 31 weiblichen Interviewern. Es ist positiv anzumerken, dass die Autoren im Gegensatz zu den vorgenannten Studien ihr Sample nicht nach der Erfolgsquote selektieren, also keinen Extremgruppenvergleich durchführen. In ihren bivariaten Analysen haben die objektiven Messdaten (Stimmhöhe, Variation der Höhe, Sprechgeschwindigkeit und Pausen) keinen Einfluss auf den Interviewer-erfolg. Es zeigen sich jedoch positive bivariate Effekte der Lautstärke, des subjektiv eingeschätzten Redeflusses sowie der Autorität. Die Einschätzung des Interviewers als zuverlässig oder liebenswürdig hat keinen signifikanten Effekt.

In einer aktuellen Studie von Groves et al. (2008) mit ca. 60 Interviewern⁴ beiderlei Geschlechts bestätigt sich der positive Einfluss einer höheren Tonlage und einer lauterer Stimme, wenn auch die Variation im Geschlecht der Interviewer nicht adäquat berücksichtigt wird. Negativ auf die Interviewteilnahme wirken sich darüber hinaus maskuline Stimmen sowie solche mit rauchigem Klang aus. Wirkt der Einleitungstext nicht frei gesprochen sondern abgelesen, ist dies der Kooperation des Respondenten ebenfalls nicht zuträglich.

Insgesamt stellt sich der Forschungsstand eher uneinheitlich dar, und die Interpretierbarkeit der Ergebnisse einiger Studien wird durch methodische Probleme eingeschränkt. Neben der geringen Fallzahl halten wir insbesondere die systematische Selektion der Probanden für problematisch, da hierbei Effekte leicht überschätzt und durch Ausreißer verzerrt werden können. Multivariate Methoden werden tendenziell nur selten eingesetzt (sicherlich auch aufgrund der geringen Fallzahlen), so dass keine zuverlässige Aussage über die Effekte einzelner Variablen gemacht werden kann. Inhaltlich zeichnet sich dennoch ab, dass insbesondere höhere und lautere Stimmen zu verbesserten Kooperationsraten führen. Zu beachten ist, dass gerade der Einfluss der Stimmhöhe (Fundamentalfrequenz) erstens nicht unabhängig vom Geschlecht sein muss und zweitens ein nicht-linearer Verlauf des Zusammenhangs angenommen werden sollte. Beidem wurde in bisherigen Modellierungen aber nicht adäquat Rechnung getragen.

3 Datengrundlage und Methode

Die vorliegende Studie basiert auf Informationen zu 56 Interviewerinnen im Alter von 20–25 Jahren, die an der Universität Mannheim im Bachelorstudiengang Soziologie an einem Lehrforschungsprojekt teilnahmen. Wir beschränken uns hier aus praktischen Gründen auf weibliche Interviewer: Stimmliche Eigenschaften sollten unbedingt getrennt nach Geschlechtern untersucht werden, da eine Analyse der gepoolten Daten zur Verzerrungen führt. Ergebnisse der phonetischen als auch der sozialpsychologischen Forschung zeigen, dass Stimmen von Frauen und Männern nach anderen Kriterien beurteilt werden, während das Geschlecht der Rater (Empfänger) nur eine untergeordnete Rolle spielt (Addington 1968; Aronovitch 2001). Zwar wäre es wünschenswert, männliche und weibliche Interviewer miteinander vergleichen zu können. Leider war in unserem Sample die Zahl der männlichen Interviewer jedoch zu gering (s. u.), um zuverlässige Ergebnisse für Männer zu erhalten.

4 Leider ist es in dem Text von Groves et al. (2008) ausgesprochen schwierig Angaben zu finden, wie hoch die Fallzahlen der einzelnen Analysen waren.

Um den Einfluss stimmlicher Merkmale des Interviewers auf die Befragungsteilnahme zu überprüfen, müssen verschiedene Daten getrennt voneinander gemessen und miteinander verknüpft werden. So werden Angaben zur *Erfolgsquote* der Interviewer und Daten zu *objektiven Eigenschaften* und *subjektiven Bewertungen* der Interviewerstimmen benötigt.

3.1 Messung der Erfolgsquote

Die Erfolgsquoten berechnen sich aus Prozessdaten, die in zwei CATI-Befragungen anfielen. Die Befragungen wurden in den Jahren 2007 und 2008 unter kontrollierten Bedingungen im CATI-Labor der Universität Mannheim durchgeführt. Im Jahr 2007 führten 32 Interviewerinnen und 13 Interviewer eine bundesweite Telefonumfrage zum Thema „Soziale Sicherheit“ durch. Hierfür wurde eine Gabler-Häder-Stichprobe von N=4.045 Telefonnummern verwendet (siehe Gabler/Häder 1999). Abzüglich nicht vergebener Telefonnummern, Geschäftsanschlüssen und Faxnummern verblieben 1.341 zu befragende Personen, von denen 250 tatsächlich befragt werden konnten (Ausschöpfungsquote von 19 %). Die Befragung im Jahr 2008 war auf die Rhein-Neckar-Region (Vorwahl 0621) beschränkt und widmete sich dem Thema „Familie und Beruf“. Insgesamt waren 40 studentische Interviewer beteiligt (29 weiblich, 11 männlich), keiner von diesen war bereits bei der Studie im Jahr 2007 als Interviewer tätig gewesen. Die Interviewerstäbe der beiden Studien sind also vollständig disjunkte Gruppen. Auch im zweiten Erhebungsjahr wurde eine Gabler-Häder-Stichprobe verwendet (N=4.500), die zu 1.046 zu befragenden Personen führte, von denen 359 befragt werden konnten (Ausschöpfungsquote von 34 %). Diese vergleichsweise hohe Ausschöpfungsquote könnte sich zumindest teilweise aus Regionalitätseffekten und einer verbesserten Öffentlichkeitsarbeit (Artikel in der Lokalzeitung „Mannheimer Morgen“) erklären.

Die Erfolgsquote der Interviewer wurde als Quotient der individuell realisierten Interviews und der Anzahl der Telefonate, bei denen eine Kontaktperson erreicht wurde, berechnet.⁵ In der Befragung 2007 lag die mittlere Erfolgsquote

5 Eine Schwierigkeit bei der Berechnung von Erfolgsquoten stellen Termine dar. So kann man argumentieren, es sei bereits ein Erfolg, einen Termin für ein Interview abzusprechen, und andererseits könne ein Interview, das aufgrund eines Termins realisiert wird, der von einer anderen Person vereinbart wurde, nicht als Erfolg gewertet werden. Im Laufe der Befragung zeigte sich jedoch, dass die Absprache von Terminen auf eine weiche Variante der Teilnahmeverweigerung hindeuten kann. Dies gilt umso mehr, da die Interviewer geschult waren, im Falle einer „keine Zeit“-Reaktion des Befragten unbedingt zu versuchen, einen Termin abzusprechen. Da außerdem die Anrufe nach Terminabsprache den Interviewern von der CATI-Software nicht systematisch zugewiesen wurden, ist keine Verzerrung der Ergebnisse zu befürchten.

der Interviewerinnen bei 14,3 % ($s=5,4$), im Jahr 2008 bei 16,0 % ($s=7,6$). Diese Erfolgsquote ist wohlgerne nicht identisch mit der Ausschöpfungsquote, da letztere auf der Zahl der zu befragenden Personen beruht. Die individuelle Erfolgsquote basiert hingegen auf der Zahl der getätigten Anrufe mit Personenkontakt, und pro Respondent können mehrere Anrufe notwendig sein. Eine Erfolgsquote von 16 % (wie im Jahr 2008) bedeutet, dass eine durchschnittlich produktive Interviewerin 6,25 Anrufe mit Personenkontakt benötigte, um ein Interview zu realisieren.

3.2 Messung von stimmlichen Eigenschaften

Insgesamt erklärten sich 56 der 61 weiblichen Interviewer bereit, eine Stimmprobe abzugeben. Die Aufnahme der tatsächlichen Interviews war uns technisch nicht möglich⁶ und ist datenschutzrechtlich bedenklich. Daher baten wir die Interviewer im Anschluss an die Feldphase um eine Aufnahme ihrer Stimme. Die Interviewer verlasen unter kontrollierten Bedingungen den Einleitungstext der Befragung. Wie in der realen Interviewsituation sprachen sie hierfür in ein Computer-Headset folgenden Text:

Guten Tag, hier ist <NAME INTERVIEWER/IN> von der Universität Mannheim. Wir führen eine <deutschlandweite / regionale> wissenschaftliche Umfrage zum Thema <„Soziale Sicherheit“ / „Familie und Beruf“> durch. Dazu würde ich gern mit derjenigen Person in Ihrem Haushalt sprechen, die bereits 18 Jahre alt ist und als letzte Geburtstag hatte. Können Sie mir bitte sagen, auf wen das in Ihrem Haushalt zutrifft?

Die Aufnahmen wurden jeweils kurz nach der Feldphase der jeweiligen Umfrage unter gleichen technischen Bedingungen durchgeführt. Sie entstammen also nicht der tatsächlichen Interviewsituation, sondern wurden nachträglich angefertigt. So erhielten wir valide Messungen der Stimmhöhe⁷ und ähnlicher Eigenschaften, nicht aber der Lautstärke. Diese wird in der realen Interviewsituation durch eine Reihe von teilweise nicht kontrollierbaren Faktoren beeinflusst. Hierzu zählen einerseits Einstellungen des Mikrofons bzw. Headsets und des CATI-Computers (diese sind kontrollierbar), andererseits Einstellungen am Telefon des Befragten. Letztere können vom Interviewer weder wahrgenommen noch beeinflusst werden. In unseren Analysen verzichteten wir folglich auf eine Betrachtung der Lautstärke.

6 Eine unmittelbare Aufnahme der Stimme des Interviewers ist in der eingesetzten Telefonsoftware (WinCATI) nicht vorgesehen. Auch gelang es uns mit den zur Verfügung stehenden Mitteln nicht, einen Mitschnitt der Einleitungssequenz über die Siemens-Telefonsoftware „Opti Client“ zu realisieren.

7 Die individuelle Stimmhöhe eines Menschen wird vor allem durch die Länge und Dicke der Stimmbänder, durch die Größe des Kehlkopfs bestimmt (Fährmann 1982). Im Gegensatz zu Akzentuierung, Betonung und Sprechgeschwindigkeit wird die Fundamentalfrequenz daher als ein nicht (dauerhaft) verstellbares Stimmmerkmal eingestuft (Bartsch 2008). Es ist jedoch zu beachten, dass auch habituelle Spannungszustände die Stimmhöhe mit beeinflussen können (Fährmann 1982).

Die Stimmproben wurden um die Stille am Anfang und Ende der Aufnahme gekürzt und mit dem phonetischen Analyseprogramm *praat* analysiert (Boersma/Weenink 2009). Es wurden die durchschnittliche Fundamentalfrequenz der Stimme F0, ihre Variation sowie – als Indikator für die Sprechgeschwindigkeit – die Dauer der Aufnahme bestimmt (siehe Tabelle 1). Die Frequenzberechnung erfolgte unter Verwendung der Autokorrelationsmethode in einem Frequenzrahmen von 75–600 Hz, mit 100 frames/sec (vgl. Hewlett/Beck 2006; Reetz/Jongman 2009 für eine Einführung in die Phonetik). Wenn im weiteren Verlauf von „Stimmhöhe“ gesprochen wird, ist immer die hier beschriebene durchschnittliche Fundamentalfrequenz gemeint.

Tabelle 1 Gemessene stimmliche Eigenschaften
(Mittelwerte, sd in Klammern)

	2007	2008
Stimmhöhe [Hz]	230,0 (18,5)	232,1 (16,1)
Variation [sd(F0)]	74,4 (12,8)	67,7 (15,5)
Dauer [sec]	17,8 (1,5)	17,1 (1,4)
N	29	27

Die durchschnittliche Frequenz der Stimmproben unterscheidet sich nicht substantiell zwischen den Befragungen, und auch die Sprechgeschwindigkeit variiert nicht systematisch. Deutliche Unterschiede sind jedoch hinsichtlich der Variation der Stimmhöhe zu erkennen. Die Gründe für Differenzen in der Variation der Stimmhöhe sind unklar, es ist jedoch davon auszugehen, dass sie artifiziell sind. Wir gehen also davon aus, dass es zwischen den Erhebungsjahren zu einer unbeobachteten Veränderung (u. U. am Aufnahmegerät) gekommen ist. Dies beeinträchtigt die Reliabilität der Messungen je Erhebungsjahr aber nicht. Um Verzerrungen durch diese Unterschiede zu vermeiden, werden in den statistischen Analysen stets gruppenzentrierte Maße der objektiv gemessenen Stimmeigenschaften verwendet (d. h. von den Messwerten der 2007er-Interviewer wurde der 2007er-Durchschnittswert abgezogen; bei den Interviewern aus 2008 der Mittelwert dieser Befragung).

Vor der *subjektiven Bewertung* der Stimmen wurden die Aufnahmen anonymisiert, indem der Interviewername herausgeschnitten und durch einen 45 Hz Sinuston von 2,3 Sekunden Dauer ersetzt wurde. Um Einflüsse der Aufnahme-Lautstärke auf die subjektiven Stimmratings auszuschließen, wurden die Stimmproben zudem in ihrer Lautstärke angeglichen.

Für die Ratings wurde auf eine anfallende Stichprobe von 159 Nutzern der Mannheimer Universitätsbibliothek zurückgegriffen (Dezember 2008). Das Durch-

schnittsalter der Rater lag bei 25,9 Jahren ($sd=8,9$), 98 waren weiblich, 61 männlich. Die Bibliotheksnutzer wurden beim Betreten des Ausleihbereichs der Zentralbibliothek gebeten, an einem Experiment zur Einschätzung von Stimmen teilzunehmen. Als Incentive wurde eine saisonal übliche Schokoladenfigur angeboten. Die Messung der Stimmeigenschaften erfolgte an Laptops mit der Software MediaLab, wobei jeder Rater zehn zufällig ausgewählte Stimmen in zufälliger Reihenfolge bewerten sollte.⁸ Bewertet wurde anhand von 7-stufigen Differenzialskalen, bei denen nur die Endpunkte benannt waren. In Anlehnung an die Literatur (siehe den Literaturüberblick weiter oben) wurden subjektive Angaben erhoben, inwieweit die Stimme bzw. der Sprecher kompetent, vertrauenswürdig, angenehm, freundlich oder selbstsicher erscheint. Zudem wurde abgefragt, wie deutlich ein regionaler oder ausländischer Akzent erkennbar ist, wie stark die Stimme männlich oder weiblich klingt, ob die Stimme dem Rater insgesamt gefällt und für wie gutaussehend er den Sprecher einschätzt. Auch die Einschätzung des Aussehens erfolgte ausschließlich auf Basis der Stimmprobe.⁹ Zum Abschluss wurde der Rater gefragt, wie gerne er von der Person interviewt werden würde. Wir bezeichnen letzteres im Folgenden als subjektive Bereitschaft der Teilnahme.

Um Ausstrahlungseffekte der vorhergehenden Stimmsamples zu verringern, wurden zwischen den Aufnahmen einige Takte eines amerikanischen Weihnachtslieds eingeblendet („Who Comes this Night?“). Im Durchschnitt liegen 19,8 Ratings pro Stimme vor (12*19, 44*20 Ratings), die zur Bildung des Skalenwertes gemittelt wurden. Die Werte variieren zwischen 1 und 7, wobei hohe Werte eine starke Ausprägung des jeweiligen Merkmals bedeuten (siehe Tabelle 2 für Kennwerte der Messungen). Die subjektiven Eigenschaften (kompetent, vertrauenswürdig, angenehm, freundlich, selbstsicher, gut aussehend) sind empirisch hoch miteinander korreliert und laden in einer Hauptkomponentenanalyse auf einem Faktor.

8 Die Ratings liegen für männliche und weibliche Stimmen vor (1.590 Ratings). In diesem Artikel werden aufgrund der Fallzahl nur weibliche Stimmen analysiert. Folglich werden nur die 1.108 Ratings der weiblichen Stimmen genutzt.

9 Ein Foto des Interviewers lag den Ratern nicht vor, es geht hier also um die Vorstellung des Raters. Mehrere Studien haben gezeigt, dass es einen Zusammenhang zwischen stimmlichen und körperlichen Merkmalen gibt. So berichten Zuckerman und Driver (1989) Ergebnisse in Bezug auf Merkmale des Gesichts. Huges, Dispenza und Gallup (2004) demonstrieren, dass Körpermaße mit stimmlichen Eigenschaften korrelieren (Verhältnisse von Schulter- zu Hüftumfang und Verhältnis von Taillen- zu Hüftumfang). Als mögliche Ursache für diese Zusammenhänge diskutieren sie hormonelle Faktoren. Ferner werden über stimmliche Merkmale auch Alters-Schätzungen des Senders ermöglicht (Mulac/Giles 1996). Alter ist zwar sicherlich eine wichtige, durch die Stimme vermittelte Eigenschaft, den Ratern war in unserer Studie allerdings klar, dass es sich bei den Stimmproben um in etwa gleichaltrige Kommilitonen handelt.

Tabelle 2 Messergebnisse und Reliabilität

	Gesamt	Durchschnittswerte		Inter-Rater-Reliabilität	
		2007	2008	ICC	Reliabilität ^a
kompetent	5,01	5,00	5,03	0,19	0,82
vertrauenswürdig	5,12	5,11	5,14	0,15	0,78
angenehm	4,98	5,00	4,96	0,19	0,82
freundlich	5,39	5,36	5,42	0,16	0,78
selbstsicher	4,80	4,68	4,92	0,20	0,83
maskulin	1,92	1,90	1,94	0,09	0,65
gefällt insgesamt	4,69	4,68	4,71	0,20	0,83
gut aussehend	4,71	4,78	4,65	0,17	0,80
gerne interviewt	4,52	4,53	4,52	0,17	0,80
ausl. Akzent	1,82	1,73	1,91	0,72	0,98
regionaler Akzent	2,10	2,17	2,04	0,14	0,77

a Reliabilität berechnet als $\rho_k = (k \cdot ICC) / (1 + (k-1) \cdot ICC)$, siehe z. B. Winer (1971).

Die tabellierten Reliabilitätskoeffizienten der Messergebnisse liegen größtenteils im zufriedenstellenden Bereich. Betrachtet man statt des Reliabilitätskoeffizienten die Intraklassenkorrelationen, wird jedoch deutlich, dass die Durchschnittswerte zwar insgesamt als annähernd reliable Messung angesehen werden können, es aber zwischen den einzelnen Ratern durchaus unterschiedliche Einschätzungen der gleichen Stimme gibt. Entsprechend liegen die ICCs für „weiche“ Eigenschaften wie „angenehm“ oder „kompetent“ nur um 0,19, bei einem klar definierten und einfach einzuschätzenden Merkmal wie „ausländischer Akzent“ hingegen bei 0,72. Die Reliabilitätskoeffizienten und ICCs unserer Messungen decken sich mit den Ergebnissen anderer Studien (z. B. Oksenberg et al. 1986; Groves et al. 2008).¹⁰

4 Ergebnisse

Im Folgenden diskutieren wir zunächst bivariate Befunde, ehe wir multivariate Modelle zum Einfluss stimmlicher Eigenschaften auf die Erfolgsquote weiblicher Interviewerinnen vorstellen.

10 Das bedeutet letztlich, dass eine ausreichend große Anzahl von Ratings (10-20) benötigt wird, wenn subjektive Eigenschaften von Stimmen mittels Rating-Skalen reliabel gemessen werden sollen. Die Ergebnisse von Van der Vaart et al. 2005 deuten jedoch darauf hin, dass eine Magnitude-Messung zu höheren Übereinstimmungen zwischen den Ratern führt. Für eine Diskussion messtheoretischer und praktischer Probleme beim Raten von Stimmen vgl. Shrivastav und Sapienza (2005), Kreiman et al. (2006) sowie Kreiman und Gerratt (1993).

4.1 Ergebnisse bivariater Analysen

Bivariate Ergebnisse lassen zwar nur bedingt Aussagen über tatsächliche Zusammenhänge zu, allerdings sind diese Befunde am ehesten mit denen früherer Forschungsarbeiten vergleichbar. Tabelle 3 fasst erstens Korrelationen zwischen den durch (subjektive) Ratings erhobenen Stimmerkmalen und der Erfolgsquote der Interviewerinnen zusammen. Zweitens werden auch Korrelationen zwischen diesen Merkmalen und der subjektiven Bereitschaft des Raters berichtet, an einem Interview teilzunehmen.

Tabelle 3 Bivariate lineare Korrelationen der subjektiven Stimmerkmale mit den Erfolgsmaßen

	Erfolgsquote	subjektive Bereitschaft
subjektive Bereitschaft	0,14	1,00
kompetent	0,19	0,86***
vertrauenswürdig	0,15	0,91***
angenehm	0,13	0,95***
maskulin	-0,19	-0,49***
freundlich	0,13	0,78***
selbstsicher	0,18	0,80***
gefällt insgesamt	0,16	0,97***
gut aussehend	0,12	0,81***
ausländischer Akzent	-0,28**	-0,44***
regionaler Akzent	0,20	-0,23*

*N=56. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$*

Es zeigt sich zunächst, dass die Korrelation zwischen der Erfolgsquote und der subjektiven Bereitschaft, sich von einer spezifischen Interviewerin befragen zu lassen, recht gering ist ($r=0,14$). Dies weist darauf hin, dass es sich bei dem objektiven Interviewerfolg und der subjektiven Bereitschaft offensichtlich um voneinander unabhängige Merkmale handelt, eine subjektive „Erfolgseinschätzung“ der Interviewerin also nicht zur Vorhersage ihrer tatsächlichen Performanz geeignet ist.

Sieht man von den Merkmalen eines regionalen und ausländischen Akzentes ab, so sind die (linearen) bivariaten Befunde überraschend einhellig. Alle subjektiven Indikatoren korrelieren stark mit der subjektiven Bereitschaft, wohingegen sich keinerlei signifikante Zusammenhänge zwischen den subjektiv eingeschätzten stimmlichen Merkmalen und der Erfolgsquote berichten lassen.

Als einflussreiches Merkmal, auch hinsichtlich der tatsächlichen Erfolgsquote, erweisen sich hingegen Akzente des Interviewers, und zwar insbesondere

eine ausländische Färbung. Ein ausländischer Akzent hat einen negativen Einfluss auf die Erfolgsquote ($r=-0,28$) und reduziert auch die Bereitschaft des Raters, sich von einer Frau mit einem solchen Stimmerkmal befragen zu lassen ($r=-0,44$). Ein regionaler Akzent erweist sich hingegen als der Erfolgsquote tendenziell förderlich ($r=0,20$), wenngleich regionale Sprachcharakteristika von den studentischen Ratern offenbar eher negativ bewertet werden ($r=-0,23$).

Betrachten wir nun noch die bivariaten linearen Korrelationen zwischen objektiven Stimmeigenschaften und der Erfolgsquote sowie der potentiellen Teilnahmebereitschaft des Raters. Die in Tabelle 4 dargestellten Zusammenhänge weisen insgesamt auf schwache und zumeist insignifikante lineare Effekte stimmlicher Eigenschaften hin. Lediglich die positive Korrelation zwischen der Varianz der Stimmhöhe, also der Modulation der Tonhöhe, und der Erfolgsquote ist auf dem 5-Prozent-Niveau signifikant.

Tabelle 4 Bivariate lineare Korrelationen der objektiven Stimmerkmale mit den Erfolgsmaßen

	Erfolgsquote	subjektive Bereitschaft
Stimmhöhe (gruppenzentriert)	-0,11	-0,08
Standardabweichung der Stimmhöhe (gruppenzentriert)	0,26**	0,13
Dauer (gruppenzentriert)	-0,07	-0,23

*N=56. * $p<0,1$, ** $p<0,05$, *** $p<0,01$*

Zusammenfassend lässt sich sagen, dass die subjektiven Angaben der Rater vergleichsweise stark miteinander korrelieren. Die Bereitschaft aber, von einer bestimmten Person gerne befragt werden zu wollen, ist weitgehend unabhängig von der tatsächlichen Erfolgsquote der Interviewerin. Subjektive Ratings spezifischer Stimmerkmale stehen mit der Erfolgsquote in keinem nennenswerten linearen Zusammenhang, und auch die bivariaten Ergebnisse zu objektiven Stimmeigenschaften sind unbefriedigend.

4.2 Ergebnisse multivariater Analysen

Bisherige Forschungsarbeiten zu Effekten der Stimme des Interviewers auf die Erfolgsquote von Interviewern nutzten nur selten multivariate Analysemethoden und unterstellten jeweils lineare Zusammenhänge. Diese Modellierung lässt sich jedoch nicht unmittelbar aus den oben skizzierten theoretischen Überlegungen ableiten.

Vielmehr ist es durchaus vorstellbar, dass gerade durchschnittliche Ausprägungen von Tonhöhe, deren Variation oder der Sprechgeschwindigkeit in beiden Modi der Informationsverarbeitung die Kooperationswahrscheinlichkeit des Respondenten fördern, wohingegen extreme Werte auffällig und u. U. sogar unangenehm erscheinen könnten. In den multivariaten Analysen prüfen wir daher systematisch auch nichtlineare Zusammenhänge; hierfür nehmen wir stets quadratische Terme der Prädiktoren mit in die Regressionsmodelle auf. Zudem wird in sämtlichen Modellen das Erhebungsjahr kontrolliert. Dies erscheint geboten, da die Umfragen der Jahre 2007 und 2008 in Erhebungsgebiet und Ausschöpfungsquote differieren. Unterschieden in den Ausprägungen der Prädiktoren wurde zuvor bereits durch eine gruppenspezifische Zentrierung auf die Jahres-Mittelwerte entgegengewirkt.

In Bezug auf *subjektiv* bewertete Merkmale der Stimme finden wir auch in den multivariaten Analysen keinen Zusammenhang mit der Erfolgsquote. Auch mit linearem und quadratischen Term der bewerteten Eigenschaften können wir keine Zusammenhänge zwischen der Erfolgsquote und den gerateten Charakteristika erkennen. Die Zusammenhänge sind nur von geringer Stärke und nicht einmal auf dem 10-Prozent Niveau statistisch signifikant. Die Ergebnisse der multivariaten, nichtlinearen Analysen bestätigen diesbezüglich unsere bivariaten Resultate, so dass wir die Ergebnisse zu subjektiven Merkmalen nicht weiter behandeln müssen. Der folgende Abschnitt konzentriert sich folglich auf eine multivariate Untersuchung *objektiv* gemessener stimmlicher Aspekte.¹¹

Tabelle 5 zeigt sukzessiv aufgebaute Regressionsmodelle zum Einfluss der Stimmhöhe, ihrer Variation sowie der Sprechgeschwindigkeit (Dauer in Sekunden, die für das Lesen des Einleitungstextes benötigt wurde) auf die Erfolgsquote – stets unter Kontrolle des Erhebungsjahres. Signifikante Unterschiede zwischen den Erfolgsquoten der beiden Erhebungsjahre bestehen in den multivariaten Modellen nicht. Die Koeffizienten des Erhebungsjahres verweisen aber auf die tendenziell leicht höhere Erfolgsquote im Jahr der regional durchgeführten Erhebung (2008).

Die Tonhöhe, die in den bivariaten Analysen praktisch keinen Effekt auf die Erfolgsquote hatte, erweist sich nach der Aufnahme eines quadratischen Terms als statistisch hoch signifikant und substanziell bedeutsam (Modell 1). Mit steigender Stimmhöhe steigt zunächst die Erfolgsquote an. Bei einer Tonhöhe um 225 Hertz, und damit nahe am Stichprobenmittelwert von 231 Hertz, erzielen weibliche Interviewerinnen demnach die höchste Erfolgsquote, die bei einer weiteren Zunahme der Grundfrequenz aber wieder sinkt. Die von Modell 1 vorhergesagten Erfolgs-

11 Bei der vergleichsweise geringen Fallzahl ist eine simultane Modellierung subjektiv gerateter und objektiv gemessener Einflussgrößen leider kaum möglich.

quoten sind in Abbildung 1 veranschaulicht.¹² Der nichtlineare Zusammenhang ist hier klar zu erkennen und fittet die Datenpunkte vergleichsweise gut.¹³ Der Anteil der erklärten Varianz ($R^2=0,15$) ist vollständig auf die unterschiedlichen Stimmhöhen der Interviewerinnen zurückführbar.

Tabelle 5 Determinanten der Erfolgsquote in telefonischen Interviews (Lineare Regression, unstandardisierte Koeffizienten, z-Werte in Klammern)

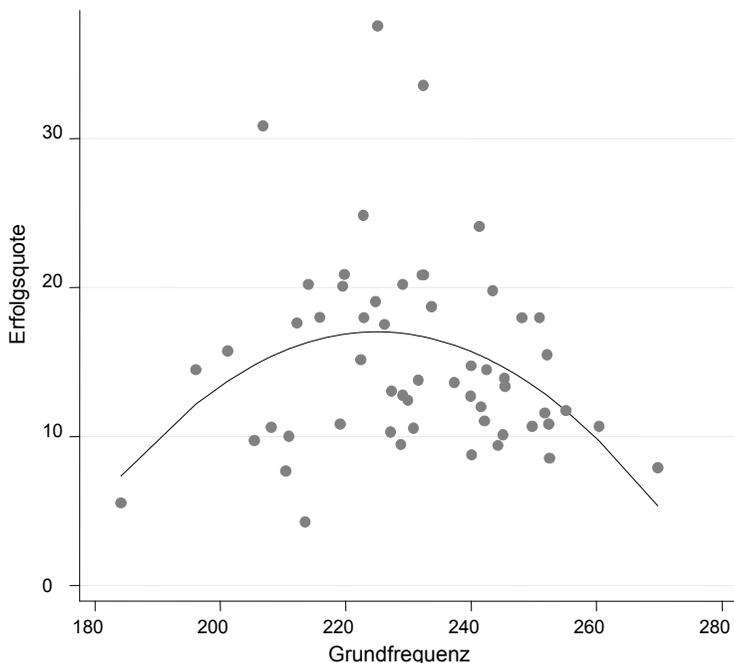
	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
<i>Stimmhöhe</i>					
linear	-0,71 (-1,43)			-0,55 (-1,11)	-0,49 (-0,99)
quadriert	-0,59*** (-2,76)			-0,56** (-2,64)	-0,45** (-2,02)
<i>Variation der Stimmhöhe</i>					
linear		1,44** (2,07)		1,26* (1,86)	1,04 (1,51)
quadriert		-0,22 (-0,64)		-0,21 (-0,64)	-0,11 (-0,32)
<i>Dauer</i>					
linear			0,14 (0,22)		-0,02 (-0,04)
quadriert			-0,73** (-2,48)		-0,43 (-1,36)
Jahr (Ref: 2007)	1,32 (0,79)	1,94 (1,12)	1,62 (0,97)	1,50 (0,91)	1,41 (0,85)
Konstante	16,19*** (12,04)	14,61*** (11,19)	15,77*** (11,99)	16,42*** (11,50)	16,81*** (11,58)
R^2	0,15	0,10	0,13	0,21	0,24
adjustiertes R^2	0,11	0,04	0,08	0,13	0,13
N	56	56	56	56	56

* $p<0,1$, ** $p<0,05$, *** $p<0,01$

Erfolgsquote in Prozent. Sowohl Stimmhöhe als auch deren Variation wurden aus Gründen der besseren Lesbarkeit vor der Berechnung der Schätzer durch 10 dividiert.

- 12 Die drei im Scatterplot auffälligen Beobachtungen mit den höchsten Erfolgsquoten verfügen zwar alle über eine in etwa mittlere Stimmlage. Tests auf Anfälligkeit des Modells gegenüber Extremwerten haben aber ergeben, dass diese und einzelne andere Datenpunkte die Regressionskoeffizienten und damit den Verlauf der Kurve nicht verzerren.
- 13 Eine nichtparametrische Lowess-Modellierung führt im Wesentlichen zu dem gleichen Kurvenverlauf.

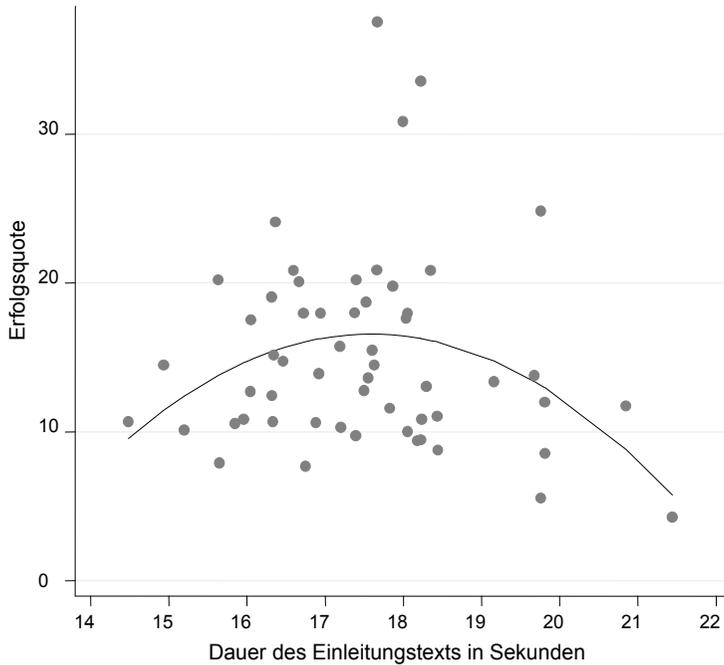
Abbildung 1 Effekte der Stimmhöhe auf die Erfolgsquote



Die Variation der Stimmhöhe (Modell 2) hatte sich bereits in den bivariaten Analysen als einflussreich erwiesen, was sich in den multivariaten Analysen zunächst auch bestätigt. Grundsätzlich wird hier ein positiver Effekt der Variation der Stimmhöhe (Modulation) errechnet, der allerdings ab einem Wert um 90 Hertz leicht abflacht und darüber hinaus keine weitere Erhöhung der Erfolgsquote mehr bedingt. Eine zu eintönige, schlecht betonte Sprechweise geht daher mit einer niedrigen Erfolgsquote einher. Die Erfolgsquote verbessert sich zunächst mit der Betonung, steigt aber bei Überbetonung nicht weiter an.

Der Effekt der Sprechgeschwindigkeit (Modell 3 und Abbildung 2) ähnelt dagegen wieder dem der Stimmhöhe: Sowohl zu langsames als auch zu schnelles Ableesen des Einleitungstextes führt zu niedrigen Erfolgsquoten. Das Maximum mit über 16,5 % wird bei einer für das Lesen benötigten Dauer von 17,5 Sekunden erreicht, also bei durchschnittlicher Sprechgeschwindigkeit. Bei einer um zwei Standardabweichungen erhöhten beziehungsweise reduzierten Dauer von 20,3 und 14,7 Sekunden liegt die vorhergesagte Erfolgsquote dahingegen nur noch bei etwa 10 %.

Abbildung 2 Effekte der Sprechgeschwindigkeit auf die Erfolgsquote



In Modell 4 wird nun simultan der Effekt der durchschnittlichen Tonhöhe während des Sprechens des Einleitungstextes und die dabei gezeigte Variation der Stimmhöhe untersucht. Es ist in diesem wie auch dem folgenden Modell jedoch zu beachten, dass – in Relation zur Zahl der verwendeten Parameter – die Zahl der Datenpunkte recht gering ist, so dass die Power der Modelle zurück geht. Entsprechend schwächen sich lineare und quadratische Effekte im Vergleich zu den Effekten in den Modellen 1 und 2 in Stärke und Signifikanz leicht ab, bleiben aber substantiell unverändert. Wird zusätzlich die Sprechgeschwindigkeit berücksichtigt (Modell 5), so bleiben signifikante Effekte lediglich der Stimmhöhe bestehen; auch die Effekte der Variation derselben sowie der Sprechgeschwindigkeit werden zumindest in ihrer Richtung nicht substantiell beeinflusst. Insgesamt erklärt dieses Modell nun ca. 24 % der Varianz der Erfolgsquote der Interviewerinnen. Dieser Anteil ist, wie Analysen ohne Berücksichtigung des Erhebungsjahres gezeigt haben, zu einem großen Teil auf objektiv gemessene stimmliche Charakteristika der Interviewerinnen zurückzuführen.

5 Diskussion und Fazit

Basierend auf Daten von 56 weiblichen Interviewern wurde in diesem Beitrag der Einfluss von stimmlichen Eigenschaften auf die Erfolgsquote in telefonischen Interviews untersucht. Hierdurch lassen sich Erkenntnisse gewinnen, die dazu dienen können, Unit-Nonresponse zu verringern und in der Folge die Datenqualität in telefonischen Befragungen zu verbessern.

Theoretisch gingen wir von einem erweiterten entscheidungstheoretischen Modell aus. Wir haben argumentiert, dass sich die Wahrscheinlichkeit eines Respondenten, am Telefoninterview teilzunehmen, erhöhen sollte, wenn die Interviewer eine Stimme haben, die als angenehm empfunden wird. In einer kalkulierten Entscheidung, wie sie in der Rational-Choice-Theorie angenommen wird, ergibt sich dies durch eine Veränderung der Kosten-Nutzen-Relation (Esser 1986; Schnell 1997). Eine angenehme Stimme, die sich durch bestimmte Merkmale auszeichnet (Apple et al. 1979; Feldstein et al. 2001; Scherer 1978), kann entweder den subjektiv empfundenen Nutzen an der Befragungsteilnahme erhöhen oder die erwarteten Kosten reduzieren, was letztlich gleichermaßen zu einer Erhöhung des Nettonutzens führt. In einem peripheren Informationsverarbeitungsmodus hingegen, wie er in dual-process-Theorien konzeptionalisiert wird (Chaiken/Trope 1999; Esser 1990; Kroneberg 2007; Petty/Cacioppo 1984), bleibt diese kognitiv anspruchsvolle Evaluation zwar aus, aber stimmliche Eigenschaften der Interviewerin können als Hinweisreize (cues) Wirksamkeit erlangen. Der entscheidende Unterschied zu einem reinen RC-Modus ergibt sich daraus, dass es sich um unbewusste Prozesse handelt, die keine aktive Evaluation und Bewertung stimmlicher Merkmale erfordern (Bartsch 2008; Zuckerman et al. 1990).

Empirisch zeigt sich, dass stimmliche Eigenschaften der Interviewer insgesamt einen nicht unerheblichen Einfluss auf ihre Erfolgsquote haben. Die Ergebnisse müssen jedoch differenziert betrachtet werden. So finden wir, dass *subjektive Ratings* der Stimmen (z. B. die Bewertung der Stimme als freundlich, angenehm o. ä.) nicht relevant für den Interviewerfolg sind. Vielmehr sind es die *objektiv gemessenen* Merkmale der Stimme, die den Erfolg als Interviewerin wirkungsvoll prognostizieren. Allerdings, und das ist ein wichtiges Ergebnis dieser Studie, ist der Zusammenhang zwischen z. B. der Tonhöhe oder der Sprechgeschwindigkeit mit den Erfolgsquoten nicht linear, sondern umgekehrt u-förmig. Solch umgekehrt u-förmige Kurvenverläufe haben sich, auch unter Kontrolle anderer Variablen, sowohl für die Stimmhöhe als auch für die Sprechgeschwindigkeit beobachten lassen. Die Maxima der vorhergesagten Erfolgsquote liegen in beiden Fällen sehr nah am Stichprobenmittelwert der beiden erklärenden Faktoren. Inhaltlich bedeutet dies, dass gerade

die *Durchschnittlichkeit* einer Stimme zu hohen Erfolgsquoten führt. Je normaler eine (weibliche) Stimme ist, desto positiver wirkt sie sich im telefonischen Interview aus, und weder schnelles noch langsames Sprechen führt zu irgendwelchen Vorteilen. Im Gegenteil ist es auch hier der mittlere Bereich, der die höchsten Erfolgsquoten aufweist. Offen und damit weiterhin erklärungsbedürftig bleibt, warum gerade die Durchschnittlichkeit kooperatives Verhalten stimuliert. Denkbar wäre, dass eine solche Stimme mit einem Attraktivitäts-Stereotyp besetzt ist. Alternativ wird in der phonetischen Forschung diskutiert, ob durch soziale Vergleichsprozesse gerade solche Stimmen überzeugen, die der des Empfängers ähnlich sind (Feldstein et al. 2001; Festinger 1954). Gegenüber der Gesamtheit der Respondenten ergäbe sich daraus ein Vorteil für Interviewer mit durchschnittlichen Stimmeigenschaften. Die Tatsache schließlich, dass wir keinen Zusammenhang zwischen den subjektiven Bewertungen der Stimme und der beobachteten Erfolgsquote, wohl aber zwischen der Erfolgsquote und objektiven Stimmeigenschaften finden, deutet darauf hin, dass der Einfluss der Stimme auf die Interviewteilnahme kein bewusster Prozess ist. Für den Fall einer Entscheidung, die auf rational-kalkulierten Nutzenerwartungen basiert, hätten wir einen stärkeren Effekt der Bewertungen erwartet. Vielmehr spricht unser Befund dafür, dass stimmliche Merkmale ihre Wirkung mitunter unbewusst über periphere Informationsverarbeitungsprozesse entfalten.¹⁴

Unsere Ergebnisse erklären auch, zumindest teilweise, den recht inkonsistenten Forschungsstand (z. B. Oksenberg/Cannell 1988; Sharf/Lehman 1984; van der Vaart et al. 2005). Offenbar ist eine rein lineare Spezifikation des Modells, wie sie in den bivariaten Korrelationsstudien unterstellt wird, den tatsächlichen Zusammenhängen nicht angemessen. Vielmehr ist die Möglichkeit nicht-linearer Beziehungen unbedingt zu beachten, was durch Befunde zu Zusammenhängen zwischen Stimme und Persönlichkeitsmerkmalen bereits nahegelegt wurde (Apple et al. 1979; Feldstein et al. 2001). Ausgehend von unseren Ergebnissen erscheint *weitere Forschung* bezüglich des Zusammenhangs von stimmlichen Eigenschaften und Interviewerfolg vielversprechend. Eine wichtige Erweiterung wäre es, die Basis der Untersuchung auszuweiten, also größere Fallzahlen zu verwenden und es dadurch zu ermöglichen, auch männliche Interviewer zu untersuchen. Unsere Arbeit zeigt zudem, dass es wichtig ist, die tatsächliche Erfolgsquote der Interviewer –

14 In Ergänzung zu dieser inhaltlichen Schlussfolgerung sollte jedoch auch die Möglichkeit eines Artefakts durch unzureichende Messung der *subjektiven* stimmlichen Eigenschaften durch Ratingverfahren in Betracht gezogen werden. Es lässt sich nicht mit Sicherheit ausschließen, dass mittels einer Messung mit Magnitude-Skalen (anstatt der in dieser Untersuchung genutzten 7-stufigen Differenzialskalen), andere Ergebnisse erzielt werden könnten. Diese Einschränkung gilt nicht für *objektive* stimmliche Eigenschaften.

und nicht die subjektive Teilnahmebereitschaft – zu untersuchen, da letztere keinen Zusammenhang mit der tatsächlichen Erfolgsquote zeigt, also nicht valide ist. Unerlässlich ist schließlich die getrennte Analyse männlicher und weiblicher Interviewer, die Verwendung multivariater Verfahren und vor allen die Berücksichtigung nichtlinearer Zusammenhänge.

Zwar sind die hier präsentierten Analysen sicherlich nicht abschließend – allein schon, da sie sich ausschließlich auf weibliche Interviewer beziehen –, dennoch lassen sich auch Ansatzpunkte für die Praxis der Umfrageforschung finden. Unsere Ergebnisse legen zunächst einmal nahe, dass in Befragungssituationen Eigenschaften des Interviewers, die über seine fachliche Kompetenz hinausgehen, eine wichtige Rolle spielen. Allein die stimmlichen Merkmale erklärten in unseren Regressionsmodellen mehr als 20 % der Varianz der Erfolgsquote. Allerdings sollte es bei der Rekrutierung von Interviewerinnen nicht darum gehen, Personen mit ganz besonderen stimmlichen Merkmalen zu selektieren. Im Gegenteil, gerade solche mit normalen, durchschnittlichen Stimmen scheinen die geringsten Verweigerungsraten zu erzielen und sind für telefonische Befragungen daher besonders geeignet.

Literatur

- Addington, D. W., 1968: The Relationship of Selected Vocal Characteristics to Personality Perception. *Speech Monographs* 35: 492-508.
- Apple, W., L. A. Streeter und R. M. Krauss, 1979: Effects of Pitch and Speech Rate on Personal Attributions. *Journal of Personality and Social Psychology* 37: 715-727.
- Aronovitch, C. D., 2001: The Voice of Personality: Stereotyped Judgments and their Relation to Voice Quality and Sex of Speaker. *The Journal of Social Psychology* 99: 207-220.
- Bartsch, S., 2008: „What Sounds Beautiful Is Good?“ How Employee Vocal Attractiveness Affects Customer's Evaluation of the Voice-to-Voice Service Encounter. S. 45-68 in: B. Strauss (Hg.): *Aktuelle Forschungsfragen im Dienstleistungsmarketing*. Wiesbaden: Gabler.
- Boersma, P. und D. Weenink, 2009: Praat, Version 5.0.46, Software zur Sprachanalyse. www.praat.org.
- Chaiken, S. und Y. Trope (Hg.), 1999: *Dual-Process Theories in Social Psychology*. New York: Guilford Press.
- de Leeuw, E. D. und W. de Heer, 2001: Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. S. 41-54 in: R. M. Groves, D. A. Dillman, J. L. Eltinge und R. J. A. Little (Hg.): *Survey Nonresponse*. New York: Wiley.
- Esser, H., 1986: Über die Teilnahme an Befragungen. *ZUMA-Nachrichten* 18: 38-47. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_18.pdf (3.3.2010).
- Esser, H., 1990: „Habits“, „Frames“ und „Rational Choice“: Die Reichweite von Theorien der rationalen Wahl (am Beispiel der Erklärung des Befragtenverhaltens). *Zeitschrift für Soziologie* 19: 231-247.
- Fährmann, R., 1982: Elemente der Stimme und Sprechweise. S. 138-164 in: K. R. Scherer (Hg.): *Vokale Kommunikation*. Weinheim, Basel: Belz.

- Feldstein, S., F.-A. Dohm und C. L. Crown, 2001: Gender and Speech Rate in the Perception of Competence and Social Attractiveness. *The Journal of Social Psychology* 141: 785-806.
- Festinger, L., 1950: A Theory of Social Comparison Processes. *Human Relations* 7: 117-140.
- Fricker, S., M. Galesic, R. Tourangeau und T. Yan, 2005: An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly* 69: 370-392.
- Gabler, S. und S. Häder, 1999: Erfahrungen beim Aufbau eines Auswahlrahmens für Telefonstichproben in Deutschland. *ZUMA-Nachrichten* 23: 45-61. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_44.pdf (3.3.2010).
- Gelinas-Chebat, C. und J.-C. Chebat, 2001: Effects of Two Voice Characteristics on the Attitudes Towards Advertising Messages. *The Journal of Social Psychology* 132: 447-459.
- Goffman, E., 1969 [1959]: *Wir alle spielen Theater*. München: Piper.
- Groves, R. M., 1989: *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, R. M., R. B. Cialdini und M. C. Couper, 1992: Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly* 56: 475-495.
- Groves, R. M. und M. P. Couper, 1998: *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R. M., D. A. Dillman, J. L. Eltinge und R. J. A. Little, 2001: *Survey Nonresponse*. New York: Wiley.
- Groves, R. M., B. O'Hare, D. Gould-Smith, J. Benki und P. Maher, 2008: Telephone Interviewer Voice Characteristics and the Survey Participation Decision. S. 385-400 in: J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japac, P. J. Lavrakas, M. W. Link und R. L. Sangster (Hg.): *Advances in Telephone Survey Methodology*. New York: John Wiley & Sons, Inc.
- Groves, R. M., E. Singer und A. Corning, 2000: Leverage-Salience Theory of Survey Participation. Description and an Illustration. *Public Opinion Quarterly* 64: 299-308.
- Hewlett, N. und J. M. Beck, 2006: *An Introduction to the Science of Phonetics*. Philadelphia: Larence Erlbaum Associates.
- Hüfken, V. und A. Schäfer, 2003: Zum Einfluss stimmlicher Merkmale und Überzeugungsstrategien der Interviewer auf die Teilnahme in Telefonumfragen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 55: 321-339.
- Huges, S. M., F. Dispenza und G. G. Gallup Jr., 2004: Ratings of Voice Attractiveness Predict Sexual Behavior and Body Configuration. *Evolution and Human Behavior* 25: 295-304.
- Kreiman, J. und B. R. Gerratt, 1993: Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research* 36: 21-40.
- Kreiman, J., B. R. Gerratt und M. Ito, 2006: When and Why Listeners Disagree in Voice Quality Assessment Tasks. *Journal of the Acoustical Society of America* 122: 2354-2364.
- Kroneberg, C., 2007: Wertrationalität und das Modell der Frame-Selektion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59: 215-239.
- Mulac, A. und H. Giles, 1996: „You're Only As Old As You Sound“: Perceived Vocal Age and Social Meanings. *Health Communications* 8: 199-215.
- Oksenberg, L. und C. F. Cannell, 1988: Effects of Interviewer Vocal Characteristics on Nonresponse. S. 257-269 in: R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls und J. Waksberg (Hg.): *Telephone Survey Methodology*. New York: Wiley.
- Oksenberg, L., Coleman, L. und C. F. Cannell, 1986: Interviewers' Voices and Refusal Rates in Telephone Surveys. *Public Opinion Quarterly* 50: 97-111.
- Petty, R. E. und J. T. Cacioppo, 1984: The Effects of Involvement on Responses to Argument Quantity and Quality – Central and Peripheral Routes to Persuasion. *Journal of Personality and Social Psychology* 46: 69-81.
- Reetz, H. und A. Jongman, 2009: *Phonetics: Transcription, Production, Acoustics and Perception*. Chichester: Wiley-Blackwell.
- Roster, C. A., R. D. Rogers, G. Albaum und D. Klein, 2004: A Comparison of Response Characteristics from Web and Telephone Surveys. *International Journal of Market Research* 46: 359-373.

- Scherer, K. R., 1978: Personality Inference From Voice Quality: The Loud Voice of Extroversion. *European Journal of Social Psychology* 8: 467-487.
- Schnauber, A. und G. Daschmann, 2008: States oder Traits? Was beeinflusst die Teilnahmebereitschaft an telefonischen Befragungen? *Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung* 2: 97-123. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.2_Heft_2/03_Schnauber.pdf (3.3.2010).
- Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Opladen: Leske und Budrich.
- Sharf, D. J. und M. E. Lehman, 1984: Relationship Between the Speech Characteristics and Effectiveness of Telephone Interviewers. *Journal of Phonetics* 12: 219-228.
- Shrivastav, R., C. M. Sapienza und V. Nandur, 2005: Application of Psychometric Theory to the Measurement of Voice Quality Using Rating Scales. *Journal of Speech, Language, and Hearing Research* 48: 323-335.
- van der Vaart, W., Y. Ongena, A. Hoogendoorn und W. Dijkstra, 2005: Do Interviewers' Voice Characteristics Influence Cooperation Rates in Telephone Surveys? *International Journal of Public Opinion Research* 18: 488-499.
- Weidmann, C., P. Schmich und S. Schiller-Born, 2008: Der Einfluss von Kontrollüberzeugungen der Interviewer auf die Teilnahme an telefonischen Befragungen, *Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung* 2: 125-147. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.2_Heft_2/04_Schmich.pdf (3.3.2010).
- Winer, B. J., 1971: *Statistical Principles in Experimental Design*. New York: McGraw-Hill.
- M. Zuckerman und R. E. Driver, 1989: What Sounds Beautiful is Good: The Vocal Attractiveness Stereotype. *Journal of Nonverbal Behavior* 13: 67-82.
- Zuckerman, M., H. Hodgins und K. Miyake, 1990: The Vocal Attractiveness Stereotype: Replication and Elaboration. *Journal of Nonverbal Behavior* 14: 97-12.

Anschrift der Autoren

Leander Steinkopf, B. A.
Ostender Straße 49a
13353 Berlin

Gerrit Bauer, Dipl. Soz.
Universität Mannheim
Mannheimer Zentrum für Europäische
Sozialforschung (MZES)
68131 Mannheim
gerrit.bauer@mzes-uni-mannheim.de

PD Dr. Henning Best
Universität Mannheim
Fakultät für Sozialwissenschaften
68131 Mannheim
best@uni-mannheim.de