

Assessing reception and individual responses to media
messages: relevance, computer techniques, quality and
analysis of Real-Time Response Measurements (RTR–M)

Rezeptionsbegleitende Erfassung individueller Reaktionen auf
Medieninhalte:
Bedeutung, Varianten, Qualität und Analyse von Real-Time-Response-
Messungen

Jürgen MAIER
Professor, Universität Koblenz-Landau,
GERMANY
maierj@uni-landau.de

Abstract: Media messages have an important impact on political attitudes and behavior. Hence, the question arises - how is media content received at an individual level? This question is difficult to investigate with the social science research methods currently available. Measuring an individual's real-time reaction to auditory or visual media content using computers, i.e. real-time response (RTR) measurement, can be a fruitful approach to address the question. This article describes various RTR techniques, discusses the reliability and validity of RTR measurements and outlines strategies on how to analyze RTR data.

Keywords: impact of media messages, real-time response (RTR) measurement, RTR techniques, reliability and validity of RTR measurement, analysis of RTR data

Evaluer la réception et les réponses individuelles aux messages des médias : pertinence, techniques informatiques, qualité et analyse des mesures de réponses en temps réel

Résumé : Les messages médiatiques ont un impact important sur les attitudes et les comportements politiques. Ceci soulève la question de savoir comment le contenu médiatique est réceptionné à un niveau individuel. Cette question est difficile à explorer à l'aide des méthodes de recherche en sciences sociales dont nous disposons actuellement. Mesurer la réaction en temps réel d'un individu à des contenus médiatiques audio ou visuels par ordinateur, c'est-à-dire mesurer la réponse en temps réel (RTR) peut être une approche fructueuse. Cet article décrit différentes techniques RTR, commente la solidité et la validité des mesures RTR et dégage des stratégies pour l'analyse des données RTR.

Mots-clés : impact des messages médiatiques, mesure de réponses en temps réel (RTR), techniques RTR, fiabilité et validité de mesure RTR, analyse de données RTR

1. Die wachsende Bedeutung von Massenmedien und die Schwierigkeit aussagekräftiger Wirkungsmessungen

Die Bedeutung von Kommunikation wächst rapide. Immer schneller werden neue, noch leistungsfähigere Kommunikationstechnologien entwickelt (Fang 1997; Merten 1994; Poe 2011). Gleichzeitig wächst die Zahl der verfügbaren Medien und die Beschäftigung mit massenmedial verbreiteten Informationen nimmt kontinuierlich zu: Während beispielsweise in Deutschland 1970 durchschnittlich nur etwa dreieinhalb Stunden pro Tag in die Nutzung von Massenmedien investiert wurden, beträgt der Medienkonsum vierzig Jahre später im Mittel knapp zehn Stunden. Dabei verwenden Rezipienten nach wie vor etwa 75 Prozent dieser Zeit auf die Nutzung von Fernsehen und Radio (van Eimeren/Ridder 2011: 8).

Im Zusammenhang mit politischen Medieninhalten spricht aber noch eine weitere Entwicklung dafür, dass die Bedeutung von Massenmedien zunimmt. So ist in nahezu allen westlichen Demokratien zu beobachten, dass die langfristigen Bindungen von Bürgern an die politischen Parteien erodieren: Parteiidentifikationen werden immer seltener und immer schwächer (Dalton 2004: 33). Dieser als Dealignment bezeichnete Prozess hat zur Folge, dass der Einfluss der Parteibindung auf die individuelle Wahrnehmung und Bewertung politischer Inhalte und Akteure abnimmt. Immer mehr Bürger können also bei der Selektion, Interpretation und

Bewertung politischer Medieninhalte *nicht* auf traditionelle Heuristiken zurückgreifen. Vielmehr schlägt die Wirkung von Medieninhalten bei der Gruppe parteipolitisch Ungebundener praktisch ungefiltert auf kurzfristig veränderbare Einstellungen durch (vgl. z.B. Holbrook 1996: 14).

Die These vom wachsenden Einfluss der Massenmedien ist zwar überaus plausibel, doch der Nachweis von Medienwirkungen ist keineswegs trivial. So kommen Untersuchungen, die mit Bevölkerungsumfragen arbeiten, immer wieder zu dem Befund, dass die nach Kontrolle relevanter Drittvariablen gemessenen Medienwirkungen auf politische Einstellungen und Verhaltensabsichten minimal sind. Die Gründe sind vor allem methodischer Natur, denn mit den üblicherweise erhobenen Mediennutzungsdaten lässt sich nur schwer erfassen, *welche Beiträge* die Befragten tatsächlich rezipiert haben (vgl. zusammenfassend Maier 2007a: 388-389). Das Problem der internen Validität von Messungen stellt sich bei experimentellen Designs nicht, da dort bekannt ist, mit welchen Informationen die Versuchsteilnehmer konfrontiert wurden. Tatsächlich können in Experimenten immer wieder Medieneffekte nachgewiesen werden. Solche Experimente stoßen aber an ihre Grenzen, wenn es darum geht herauszufinden, *welche Aspekte eines Beitrags* Wirkungen entfalten. Dieses Problem entsteht dadurch, dass die verwendeten Medienstimuli aus Einzelinformationen zusammengesetzte Informationspakete sind. Da der Stimulus aber nur in seiner Gesamtheit untersucht wird, kann man zwar anschließend eine Aussage darüber treffen, ob er *insgesamt* eine Wirkung hinterlassen hat. Darüber, welche Bestandteile des Informationspakets für die Wirkung verantwortlich sind, kann man aber letztlich nur spekulieren.

An einem Beispiel lässt sich dieses Problem verdeutlichen. Gerhard Schröder wurde in der Fernsehdebatte 2005 auf die Rolle seiner Frau im Bundestagswahlkampf angesprochen. Am Ende seiner Ausführungen erfolgte eine Liebeserklärung an seine Frau:

„Ich sage: Meine Frau hat das gute Recht, erstens, die Wahrheit zu sagen, denn es ist die Wahrheit, und zweitens, sich zu äußern, wann immer sie das für richtig hält. Und sie äußert sich engagiert. Ich sag's noch einmal: Sie lebt das, was sie sagt, und ich füge hinzu: Das ist nicht zuletzt der Grund, warum ich sie liebe.“

Diese für ein solches Wahlkampfformat zweifellos ungewöhnliche Aussage wurde von den Massenmedien aufgegriffen. Insbesondere wurde dabei über die Wirkung der Liebeserklärung spekuliert (vgl. Reinemann 2007). In einem gewöhnlichen Experiment mit einer Experimentalgruppe, die das TV-Duell gesehen hat, und einer Kontrollgruppe, die dem Stimulus nicht ausgesetzt wurde, kann man dieses Rätsel nicht auflösen. Man kann zwar durchaus untersuchen, ob die Rezipienten der Fernsehdebatte nach dem TV-Duell andere Kognitionen, Einstellungen und Verhaltensabsichten aufweisen, als vor dem Duell und – wenn man Experimental- und Kontrollgruppe miteinander vergleicht – ob tatsächlich das

Verfolgen der Fernsehdiskussion kausal für die gemessenen Änderungen verantwortlich ist. Angesichts der zahlreichen verbalen oder visuellen Informationen, die im Laufe einer Debatte von den Kandidaten und den Moderatoren erteilt werden, bleibt jedoch im Dunkeln, welche davon für die experimentell gemessenen Veränderungen verantwortlich sind.

Diesbezügliche Erkenntnisse über die Stärke der Wirkung spezifischer Aspekte von Medieninhalten sind jedoch von großer theoretischer und praktischer Bedeutung. So wäre es auf diese Weise möglich, die *psychologischen Mechanismen*, welche hinter der Herausbildung und Veränderung von politischen Einstellungen stehen, genauer zu untersuchen. Hinweise, wie Rezipienten auf einzelne Aussagen reagieren, würden erhebliche Fortschritte bei der sozialwissenschaftlichen Theoriebildung bedeuten. Für Praktiker bestünde der Wert solcher Erkenntnisse beispielsweise darin, die Wirkung von Kampagnen gezielt zu steigern.

2. Real-Time-Response-Messung: Sekundengenaue Erfassung von Medienwirkungen in Echtzeit

Real-Time-Response-Messungen („RTR-Messungen“) bieten eine solche Möglichkeit. Die RTR-Technik sieht eine *rezeptionsbegleitende* Erfassung der Bewertung audiovisueller Medieninhalte vor (West/Biocca 1996: 120). Rezipienten können mit Hilfe solcher Systeme ihre spontanen Reaktionen auf den präsentierten – auditiven oder visuellen – Medieninhalt mitteilen. Zu diesem Zweck erhalten sie die Möglichkeit, ihre positiven und negativen Eindrücke über eine Eingabeeinheit an einen Computer zu übermitteln. Dieser erfasst die Reaktionen sekundengenaue, sodass (durch den Vergleich von Reaktionen und Stimulus) festgestellt werden kann, wer wann wie auf den Medieninhalt reagiert hat. Auf diese Weise ist es also möglich, die „black box“ der individuellen Informationsverarbeitung ein Stück weit zu öffnen und die spezifische Bewertung der Inhalte medialer Informationspakete zu untersuchen. Wie ertragreich das sein kann, soll nochmals an dem oben angeführte Beispiel verdeutlicht werden: Mit Hilfe von RTR-Messungen wurde nachgewiesen, dass Schröders Liebeserklärung *keinen* besonderen Eindruck auf die Zuschauer der Fernsehdebatte gemacht hat (Reinemann/Maurer 2007). Die Einschätzung zahlreicher Experten und Politiker, die Aussage des damaligen Bundeskanzlers habe eine starke Wirkung auf die Fernsehzuschauer gehabt, konnte empirisch nicht belegt werden.

Ursprünglich wurde das Verfahren der RTR-Messung zur Evaluation von im Radio ausgestrahlten Unterhaltungsprogrammen verwendet (Fiske/Lazarsfeld, 1945; Peterman, 1940; Peatman/Hallonquist 1945). Die Möglichkeiten der RTR-Messung werden seitdem in zahlreichen (akademischen) Feldern genutzt (zur Geschichte der RTR-Technologie vgl. z.B. Biocca/David/West 1994; Millard 1992; Levy 1982). So spielt in der Musikwissenschaft die Messung von Echtzeitreaktionen auch heute noch eine Rolle (z.B. Brittin 1996). Den Nutzen der Technologie hat auch die

Werbewirkungsforschung für sich erkannt (z.B. Bente/Aelker/Fürtjes 2009; Hughes 1992). Genutzt werden RTR-Systeme aber auch, um das Persuasionspotenzial von Informationskampagnen zu untersuchen (z.B. Baggaley 1986; Tedesco/Ivory 2009). Der Fokus akademischer Forschung liegt vor allem auf der Analyse politischer Medieninhalte im TV: Fernsehreportagen (z.B. Meyer/Séгур 2009), Propagandafilme (Hovland/Lumsdaine/Sheffield 1949), Wahlwerbesendungen (z.B. Kaid 2009) oder TV-Duelle (z.B. Faas/Maier 2004; Maurer/Reinemann 2003; Maurer et al. 2007). RTR-Messungen eignen sich – etwa wenn sie mit experimentellen Designs kombiniert werden – auch zur Analyse spezieller Fragen, wie beispielsweise einem Vergleich der Bedeutung verbaler und visueller Informationen (z.B. Maier/Maier 2009; Roessing/Jakob/Petersen 2009), oder für kulturvergleichende Wirkungsstudien (Maier/Strömbäck 2009).

Es existieren verschiedene technische Systeme zur Erfassung von Echtzeitreaktionen, die sich insbesondere hinsichtlich des Skalenniveaus der gesammelten Daten unterscheiden. Zum einen existieren sogenannte *Druckknopfsysteme* („Push Button“). Diese Technik wurde in den 1930er Jahren von Paul F. Lazarsfeld und Kollegen in den USA entwickelt. Die Originalversion verfügte über zwei Knöpfe, über welche Rezipienten positive bzw. negative Eindrücke anzeigen konnten (vgl. Peterman 1940). Heute werden auch Systeme eingesetzt, die über mehr Knöpfe – und damit komplexere Möglichkeiten der Reaktionsmessung – verfügen (vgl. z.B. Faas/Maier 2004).¹ Ein Limit hinsichtlich der Anzahl der Druckknöpfe wird letztlich durch den Nutzer und seine Fähigkeit gesetzt gleichzeitig mit mehreren Bewertungsdimensionen umzugehen. Unabhängig von der Anzahl der verwendeten Druckknöpfe können mit dieser Technologie aber nur binäre Daten gesammelt werden: Immer dann, wenn ein Rezipient auf den Stimulus reagiert, wird dies aufgezeichnet. Reagiert der Rezipient nicht, wird auch kein Signal übermittelt; der Druckknopf wird somit wieder auf die Ausgangsposition zurückgesetzt („reset mode“). Damit werden für den Forscher individuelle Reaktionsschwellen sichtbar – also die Stellen im Stimulusmaterial, die den Rezipienten zu einer via RTR gemessenen Reaktion provozieren.

Auf der anderen Seite existieren *Drehreglersysteme* („Dial“), mit denen Rezipienten ihre Reaktionen auf einer Skala abstufen können.² Während ältere Systeme meist mit 7-stufigen Skalen arbeiten, erlauben neuere Anlagen, das Skalenformat frei zu wählen. Mit Drehreglern können quasi-metrische Daten für maximal zwei Bewertungsdimensionen erfasst werden – z.B. Eindrücke von zwei Kandidaten, die im Rahmen eines TV-Duells miteinander diskutieren (vgl. z.B. Maurer/Reinemann 2003; Maurer et al. 2007). Da vom Nutzer die Eingabe eines

¹ Die uns bekannten Systeme sind selbstprogrammiert und können mit Hilfe eines handelsüblichen Computers eingesetzt werden.

² Weiterhin gibt es auch Schiebereglersysteme. Diese haben ähnliche Eigenschaften wie Drehreglersysteme. Alternativen zum klassischen Druckknopfsystem wurden bereits in den 1940er und 1950er Jahren entwickelt (vgl. Twyford 1953).

Skalenniveau gefordert wird, bedeutet dies, dass die von ihm wahrgenommenen Eindrücke gewichtet und gegebenenfalls miteinander verrechnet werden müssen (Maier/Faas 2009). Drehreglersysteme operieren im sogenannten „latched mode“: Die gewählte Einstellung bleibt erhalten bis der Rezipient diese verändert. Damit übermittelt der Rezipient zu jedem Zeitpunkt Messwerte. Datenlücken entstehen hier nicht. Allerdings sind Reaktionsschwellen nicht mehr ohne weiteres identifizierbar. Sie offenbaren sich nur dann, wenn Veränderungen der Drehreglereinstellung vorgenommen werden. Der Schluss, dass bei unveränderter Drehreglerposition Rezipienten nicht auf den Medieninhalt reagieren, ist jedoch nicht korrekt (vgl. hierzu auch ausführlich Maier/Faas 2009).

Druckknopf- und Drehreglersysteme operieren also sehr unterschiedlich. Während Drehreglersystem den Vorteil haben, quasi-metrische, lückenlose und damit sehr anschauliche Daten zu liefern, liegt die Stärke von Druckknopfsystemen darin, dass positive und negative Bewertungen zu mehreren Dimensionen separat erfasst und damit Informationsverarbeitungsprozesse detaillierter abgebildet werden können (vgl. Tabelle 1). Zudem sind Druckknopfsysteme auch mit einem geringen Budget zu realisieren. Angesichts der Unterschiedlichkeit der gesammelten Messdaten ergänzen sich die beiden Technologien gegenseitig.

		Druckknopf („Push Button“)	Drehregler („Dial“)
Skalenniveau	der	nominal	quasi-metrisch
gemessenen Daten		(dichotom)	
Modus der Messung		„reset mode“	„latched mode“
Reaktionsschwellen		erkennbar	eingeschränkt erkennbar
Simultan	erfassbare	nicht definiert	maximal 2
Dimensionen			

Tabelle 1. Zentrale Eigenschaften von Drehregler- und Druckknopfsystemen

3. Reliabilität und Validität von Real-Time-Response-Messung *Reliabilität*

Druckknopf- und Drehreglersysteme können sich nur sinnvoll ergänzen, wenn sie nicht völlig identische oder völlig konträre Ergebnisse produzieren. Im ersten Fall könnten man auf eine Technologie verzichten, in letzterem Fall würde jedes System etwas völlig anderes messen und die Messergebnisse könnten nicht miteinander in Beziehung gesetzt werden. Dass diese Voraussetzung erfüllt ist, belegen Untersuchungen, die beide Messsysteme hinsichtlich Reliabilität und Validität direkt miteinander *vergleichen* (Maier/Faas 2009; Maier et al. 2007;

Reinemann et al. 2005).³ Sie zeigen, dass Rezipientenreaktionen an ganz ähnlichen Stellen des verwendeten Stimulusmaterials – hier: das zweite TV-Duell im Bundestagswahlkampf 2002 – zu beobachten sind. Die zu Zeitreihen aufbereiteten Messwerte beider Systeme korrelieren moderat miteinander ($r=0,39$, $p<0,01$). An Schlüsselstellen der Fernsehdebatte fallen die Zusammenhänge aber deutlich höher aus (maximal beträgt $r=0,69$, $p<0,01$). Dies deutet darauf hin, dass mit beiden Technologien reliable Messungen spontaner Reaktionen auf rezipierte Medieninhalte erzielt werden können.

3.1. Interne Validität

Weiterhin konnten die bisher vorliegenden Studien zeigen, dass beide Verfahren intern valide sind – also das messen, was sie messen sollen (Maier et al. 2007; Reinemann et al. 2005). Dies gilt einerseits im Hinblick auf die *Konstruktvalidität*, denn sowohl die via Druckknopf, als auch die via Drehregler gesammelten Messwerte hängen in der erwarteten Weise mit anderen Konstrukten zusammen. Untersuchungen zur Wahrnehmung und Wirkung von Fernsehdebatten zeigen beispielsweise, dass die spontanen Eindrücke von Parteianhängern in Bezug auf „ihren“ Kandidaten signifikant positiver ausfallen, als die Bewertungen des politischen Gegners. Andererseits kann die interne Validität beider Messtechnologien auch im Hinblick auf ihre *Kriteriumsvalidität* belegt werden (Maier et al. 2007; Reinemann et al. 2005). So korrespondierte das bei Debattenrezipienten *nach* dem TV-Duell gefällte Urteil über den Sieger der Fernsehdiskussion, mit den *während* der Debatte gemessenen spontanen Reaktionen.

3.2. Externe Validität

Im Gegensatz zur Reliabilität und zur internen Validität von RTR-Systemen liegt zur externen Validität nur eine Studie vor, die die Druckknopf- und die Drehreglertechnologie miteinander vergleicht. Die Frage, die sich hier stellt ist, inwieweit sich die Ergebnisse von Experimenten, bei denen RTR-Messungen eingesetzt wurden, generalisieren lassen. Angeregt durch die allerdings nur auf Drehregler bezogenen Befund der Untersuchung von Maurer und Reinemann (2009), die zu dem Ergebnis kommen, dass die Verwendung von RTR-Systeme die externe Validität von Experimenten einschränkt, untersucht Maier (2011) den Einfluss solcher Echtzeitmessungen auf die Fähigkeit von Probanden, sich an Ausschnitte aus einer Fernsehdebatte zu erinnern. Dabei wurde der Konfliktgehalte der Debatte variiert. Im Ergebnis zeigt sich, dass RTR-Messungen bei der

3 Neben vergleichenden Untersuchungen liegen einige Studien vor, die sich Fragen der Reliabilität und Validität von Druckknopf- oder Drehreglersystemen widmen (vgl. z.B. Baggaley 1987; Biocca/David/West 1994; Boyd/Hughes 1992; Fahr/Fahr 2009; Hollonquist/Suchman 1944; Hovland/Lumsdaine/Sheffield 1949; Hughes/Lennox 1990; Maier/Faas 2009; Peatman/Hallonquist 1945, 1950; Peterman 1940; Reinemann/Mauer 2009; Schwerin 1940; Thorson/Reeves 1986; Wolf 2010).

konfliktarmen Debatte unproblematisch sind. Streiten die Kandidaten allerdings heftig, erschweren RTR-System die Erinnerung an den Medieninhalt. RTR-Messungen lenken unter diesen Bedingungen also vom Medieninhalt ab und die Generalisierbarkeit anderer experimentell gewonnener Daten wird eingeschränkt. Dabei ist es unerheblich, ob die Probanden Drehregler- oder Druckknopfsysteme bedient haben.

Wie aber empfinden Probanden die Verwendung von RTR-Systemen? Wir wollen an dieser Stelle die Daten des erwähnten Experiments noch einmal heranziehen und die subjektiven Einschätzungen der Probanden zu Drehregler- und Druckknopfsystemen vergleichen analysieren. Das Datenmaterial stammt aus zwei Experimenten zur schleswig-holsteinischen Fernsehdebatte 2009 (zum Studiendesign vgl. ausführlicher Maier 2011). Im ersten Experiment wurden Ausschnitte der Debatte gezeigt, in denen die Kandidaten ohne Unterbrechung ihre Standpunkte vortragen konnten („konfliktarme Debatte“). Im zweiten Experiment wurden Sequenzen präsentiert, in denen die Kandidaten heftig miteinander gestritten haben („konfliktreiche Debatte“).⁴ In beiden Untersuchungen wurde eine Probandengruppe mit siebenstufigen Drehreglern ausgestattet (N=24 bzw. N=39). Eine andere Gruppe erhielt die Gelegenheit, die Performanz der beiden Kandidaten mit insgesamt vier Druckknöpfen zu bewerten (je eine eindeutig gekennzeichnete Taste auf einer PC-Tastatur für positive bzw. negative Eindrücke pro Kandidat; N=22 bzw. N=23).⁵

	Konfliktarme Debatte		Konfliktreiche Debatte	
	Druckkno	Drehregle	Druckkno	Drehregle
	pf („Push Button“)	r („Dial“)	pf („Push Button“)	r („Dial“)
Das Bedienen der Eingabeeinheit hat mich von der Debatte abgelenkt.	-1,4	-0,8	-1,5	-1,1

⁴ Die Zusammenschnitte waren wie folgt aufgebaut: Bei beiden Experimenten führte der Moderator zunächst in die Bedeutung und die Regeln der Debatte ein (Dauer: 27 Sekunden). Daran schloss sich ein Trailer zum Thema „Energie und Umwelt“ an (60 Sekunden). Darauf folgte die Stellungnahme der Kandidaten zu diesem Thema. Diese dauerte 267 (Experiment 1) bzw. 291 Sekunden (Experiment 2), wobei 84 Sekunden der hier ausgewählten Passagen identisch waren.

⁵ In jedem Experiment hat eine Kontrollgruppe das Duell ohne RTR-Technik verfolgt (N=21 bzw. N=29). Die Versuchspersonen – Studierende der Universität Koblenz-Landau –, wurden hinsichtlich des Geschlechts parallelisiert. In Experiment 1 waren 64 Prozent, in Experiment 2 54 Prozent der Versuchsteilnehmer weiblich. Bei beiden Versuchen gab es zwischen den drei Versuchsgruppen keine signifikanten Unterschiede hinsichtlich der Geschlechterverteilung ($p > 0,05$). Gleiches gilt für das politische Interesse, sowie für die Parteiidentifikation mit einer Regierungspartei (CDU/CSU, FDP) bzw. einer Oppositionspartei (SPD, Bündnis 90/Die Grünen, Die Linke).

Die Möglichkeiten, die mir die Skala/Tastatur zur Be- wertung der Kandidaten geboten haben, waren ausreichend.	0,7	0,9	0,7	1,1
Wenn beide Kandidaten gesprochen haben, war es schwer, eine Bewertung abzugeben.	-1,2	-0,7	0,3	0,2
Wenn beide Kandidaten zu sehen waren, war es schwer, eine Bewertung abzugeben.	-0,7	-0,8	-0,7	-0,9
Ich musste häufig auf die Skala/Tastatur schauen, um sicherzugehen, dass ich auch die richtige Bewer- tung abgebe.	-0,7 ^a	0,2	-0,7	-0,3

a: $p < 0,05$, b: $p < 0,01$, c: $p < 0,001$

d: 5-Punkte-Skalen von -2 („stimme überhaupt nicht zu“) bis +2 („stimme voll und ganz zu“)

Tabelle 2. Subjektive Bewertung^d von RTR-Systemen nach Konfliktgehalt der Debatte

Die in Tabelle 2 dokumentierten Ergebnisse zeigen (bis auf eine Ausnahme) keine signifikanten Unterschiede in der Handhabbarkeit von Drehregler- und Druckknopfsystemen – und zwar unabhängig vom Konfliktgehalt der Debatte. Generell lenken beide Technologien kaum vom präsentierten Medieninhalt ab. Bewertungen sind auch dann möglich, wenn beiden Kandidaten im Bild sind oder reden.⁶ Die Probanden fühlten sich zudem nicht durch die Möglichkeiten, die die Skala bzw. die Tastatur bot, in ihrem Urteil eingeschränkt. Die einzige statistisch bedeutsame Differenz besteht hinsichtlich der Frage, ob man häufig auf die Skala bzw. Tastatur geschaut hat, um sicherzustellen, das richtige Urteil abgegeben zu haben. Dies kam bei Probanden, die mit dem Druckknopfsystem ausgestattet waren, deutlich seltener vor als bei Versuchspersonen, die Drehregler bedienten. Dieser

⁶ Allerdings wird dies bei beiden Technologien (statistisch signifikant) schwieriger, wenn der Konfliktgehalt zunimmt.

Unterschied findet sich allerdings bei Rezeption der konfliktreichen Debatte allerdings nicht mehr.⁷

4. Auswertung von Real-Time-Response-Messung

RTR-Daten eröffnen zahlreiche Auswertungsoptionen (vgl. hierzu auch ausführlich Biocca/David/West 1994). Grundsätzlich können RTR-Analysen hinsichtlich des Aggregationsniveaus der verwendeten Daten unterschieden werden, also ob auf Individualdaten (also die für jeden einzelnen Probanden für jede Sekunde gesammelten Messwerte) oder auf Aggregatdaten (also auf über alle oder bestimmte Teilgruppen von Probanden hinweg zusammengefasste Daten) zurückgegriffen wird. Außerdem unterscheiden sich Auswertungen von RTR-Daten im Hinblick auf den Kompressionsgrad der erhobenen Datenzeitpunkte – also ob diese zu Blöcken zusammengefasst werden (z.B. nach bestimmten Aussagen oder Themenblöcken), oder ob gar soweit komprimiert wird, dass am Ende z.B. nur ein über alle Beobachtungszeitpunkte errechneter Mittelwert vorliegt. Entlang beider Dimensionen der Datenmatrix können also Verdichtungen vorgenommen werden, um die Masse der erhobenen Datenpunkte beherrschbar zu machen (vgl. Tabelle 3).

Typisch für die Auswertung von RTR-Messungen ist vor allem die Arbeit mit aggregierten Daten. Eine mittlere Verdichtung wird dann gewählt, wenn die Entwicklung von Bewertungsverläufen über alle Probanden oder über bestimmte Probandengruppen beschrieben werden soll (z.B. Faas/Maier 2004). Um der Deskription Substanz zu verleihen wird dabei gerne auf Datenpunkte fokussiert, die eine besonders starke positive oder negative Bewertung des Stimulusmaterials erkennen lassen oder bei denen die interindividuelle Homo- oder Heterogenität von Echtzeitreaktionen besonders ausgeprägt ist (vgl. z.B. Maurer/Reinemann 2003; Reinemann/Maurer 2005).

Eine maximale Verdichtung von Aggregatdaten wird oft gewählt, wenn es um eine Gesamtbewertung des präsentierten Medienstimulus oder bestimmter Teile des Medienstimulus geht (z.B. Bewertung eines Werbespots oder Abschneiden von Kandidaten in einer TV-Debatte). Das Problem bei dieser Vorgehensweise liegt vor allem darin, dass bei der Zusammenfassung der Messzeitpunkte implizit die Annahme getroffen wird, dass alle Datenpunkte mit gleichem Gewicht in ein Gesamturteil eingehen. Dass hier Zweifel angebracht werden müssen zeigen Untersuchungen, in denen RTR-Daten (etwa zur Bewertung von Kandidaten in einer TV-Debatte) mit einem externen Kriterium (etwa der Frage nach dem in Fragebögen erhobenen Debattensieger) verglichen wird. Beide Maße hängen stark, aber

⁷ Unklar ist, warum dies so ist. Möglicherweise mussten sich die Probanden bei der Rezeption der konfliktreichen Debatte stark auf den Inhalt der Diskussion konzentrieren, so dass sie keine Zeit mehr hatten, die Richtigkeit ihrer Eingaben durch einen Blick auf den Drehregler bzw. die Tastatur zu überprüfen.

keinesfalls perfekt miteinander zusammen (Maier/Faas 2004). Offenbar gewichten Rezipienten einzelne Sequenzen des ihnen präsentierten Medieninhalts unterschiedlich stark, wenn sie aufgefordert werden, ein Gesamturteil zu fällen.

	Individualdaten	Aggregatdaten
Verwendung aller Messzeitpunkte	Zeitreihe über alle Messzeitpunkte für jeden Probanden (keine Verdichtung)	Zeitreihe über alle Messzeitpunkte für die Gesamtheit oder Teilgruppen von Probanden (mittlere Verdichtung bei Verzicht auf die Abbildung interindividueller Unterschiede)
Zusammenfassung von Messzeitpunkten	Lage- und Streuungsmaße für jeden Probanden über bestimmte Abschnitte der Zeitreihe bzw. über die gesamte Zeitreihe (mittlere Verdichtung bei Verzicht auf die Abbildung der RTR-Daten inhärenten Dynamik)	Lage- und Streuungsmaße für die Gesamtheit oder Teilgruppen von Probanden über bestimmte Abschnitte der Zeitreihe bzw. über die gesamte Zeitreihe (maximale Verdichtung bei Verzicht auf die Abbildung interindividueller Unterschiede und der RTR-Daten inhärenten Dynamik)

Tabelle 3. *Möglichkeiten zur Verdichtung von RTR-Daten*

Weitaus seltener werden Individualdaten verdichtet. Dies geschieht meist nur dann, wenn individuelle Reaktionen mit anderen Datenquellen (z.B. einem Fragebogen) verknüpft werden sollen. RTR-Daten werden auf diese Weise mit anderen Persönlichkeitsmerkmalen der Probanden in Beziehung gesetzt – etwa wenn es darum geht, den Einfluss der während einer Fernsehdebatte gemessenen spontanen Reaktionen auf die Wahrnehmung des Debattensieger oder auf Veränderung von Kandidatenbewertungen zu untersuchen (z.B. Maier 2007b). Allerdings erweist es sich auch hier als problematisch, dass von einer Gleichgewichtung der Messzeitpunkte ausgegangen wird.

Analysen *ohne* jegliche Verdichtung der gesammelten RTR-Daten liegen bislang nicht vor. Allerdings dürfte hier ein erhebliches Analysepotenzial von RTR-Daten liegen. Hilfreich könnten hier statistische Verfahren sein, die vor allem in der Psychologie (Verfahren zur Analyse von Messwiederholungen) und der Ökonomie (Zeitreihenanalysen) verwendet werden. In beiden Fällen darf aber ein zusätzliches Problem nicht aus dem Auge verloren werden, das bei Messungen mit der Dial-Technik verursacht wird: Im Unterschied zur Druckknopftechnik sind die

gemessenen Daten hier nicht voneinander unabhängig. Eine gemessene Drehreglereinstellung ist nämlich nicht nur ein Produkt der aktuellen Reaktion auf das Stimulusmaterial, sondern – aufgrund des „latched mode“ – auch eine Funktion des zuvor eingestellten Drehreglerwerts.

5. Fazit

Angesichts der Tatsache, dass individuelle politische Einstellungen – und folglich auch politisches Verhalten – zunehmend eine Funktion der Wahrnehmung und Verarbeitung von Medieninhalten darstellen, ist eine Antwort auf die Frage, wie diese Medieninformationen von Rezipienten verarbeitet werden, von zentraler Bedeutung für die Weiterentwicklung sozial- und verhaltenswissenschaftlicher Theorien. Der Zugriff auf diese Frage ist mit den gängigen Methoden der empirischen Sozialforschung jedoch nicht oder nur eingeschränkt möglich. RTR-Messungen bieten hier eine interessante Alternative: die rezeptionsbegleitende, computergestützte Erfassung von spontanen Reaktionen auf auditive oder visuelle Medieninhalte macht es möglich, individuelle Informationsverarbeitungsprozesse wie unter einem Brennglas zu betrachten.

Hierfür stehen im Wesentlichen zwei Techniken zur Verfügung, die nicht nur miteinander vergleichbare, sondern vor allem sich ergänzende Messergebnisse liefern. Darauf deuten die vorliegenden vergleichend angelegten Untersuchungen zur Reliabilität und internen Validität von Druckknopf- und Drehreglertechnik hin. Komparative Studien zur externen Validität der beiden Technologien sind hingegen Mangelware. Die hier präsentierten Ergebnisse zur subjektiven Einschätzung von RTR-Systemen durch Teilnehmer eines Laborexperiments zeigen, dass der Einsatz von Echtzeitmessungen aus Sicht von Probanden weitgehend unproblematisch ist. Um aber ein fundiertes Urteil zu fällen, ob die Generalisierbarkeit von Forschungsergebnissen durch die Verwendung von RTR-Technologie beeinträchtigt wird, ist weitere Forschung notwendig. Dies gilt auch im Hinblick auf die effiziente Analyse von RTR-Daten. Gegenwärtig besteht die Tendenz, die Messergebnisse stark zu verdichten – eine Praxis, die das in den Daten enthaltene Informationspotenzial nur zu einem Bruchteil ausschöpft. Die zukünftige Forschung auf diesem Gebiet wird sich deshalb verstärkt der Frage zuwenden müssen, wo die Möglichkeiten und Grenzen der RTR-Technik liegen, und wie man die bei RTR-Messungen entstehenden großen Datenmengen angemessen analysieren kann. Gelingt dies, haben wir ein weiteres Instrument zur Verfügung, um die Verarbeitung von Medieninhalten besser zu verstehen.

References

- Baggaley, J. (1986). Formative Evaluation of Educational Television. *Canadian Journal of Educational Communication*, 15, 29-43.

- Baggaley, J. (1987). Continual Response Measurement. Design and Validation, *Canadian Journal of Educational Communication*, 16, 217-238.
- Bente, G., Aelker, L., & Fürtjes, M. (2009). Tracing Ad Experience. Real-Time Response Measurement in Advertising Research. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds., *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 117-135.
- Biocca, F., Prabu, D., & West, M. (1994). Continuous Response Measurement (CRM). A Computerized Tool for Research on the Cognitive Processing of Communication Messages. In Lang, A., Ed., *Measuring psychological responses to media*. Hillsdale, NJ: Erlbaum, 15-64.
- Boyd, T.C., & Hughes, G.D. (1992). Validating Realtime Response Measures. *Advances in Consumer Research*, 19, 649-656.
- Brittin, R.V. (1996). Listener's Preference for Music of Other Cultures. Comparing Response Modes. *Journal of Research in Music Education*, 44, 328-340.
- Dalton, R.J. (2004). *Democratic Challenges, Democratic Choices. The Erosion of Public Support in Advanced Industrial Democracies*. Oxford: Oxford University Press.
- Faas, T., & Maier, J. (2004). Mobilisierung, Verstärkung, Konversion? Ergebnisse eines Experiments zur Wahrnehmung der Fernsehduelle im Vorfeld der Bundestagswahl 2002. *Politische Vierteljahresschrift*, 45, 55-72.
- Fahr, A. (2009). Reactivity of Real-Time Response Measurement. The Influence of Employing RTR Techniques on Processing Media Content. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds. *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 45-62.
- Fang, I. (1997). *A History of Mass Communication. Six Information Revolutions*. Boston: Focal Press.
- Fiske, M., & Lazarsfeld, P.F. (1945). The Columbia Office of Radio Research. *Hollywood Quarterly*, 1, 51-59.
- Holbrook, T.M. (1996). *Do Campaigns Matter?* Thousand Oaks: Sage.
- Hollonquist, T., & Suchman, E.A. (1944). Listening to the Listeners. Experiences with the Lazarsfeld-Stanton Program Analyzer. In Lazarsfeld, P.F., Stanton, F.N., Eds., *Radio Research 1942-1943*. New York, Duell: Sloan and Pierce, 265-334.
- Hovland, C.I., Lumsdaine, A.A., & Sheffield, F.D. (1949). *Experiments on Mass Communication*. New York: Wiley.

- Hughes, D. (1992). Realtime Response Measures Redefine Advertising Wearout. *Journal of Advertising Research*, 13 (3), 61-77.
- Hughes, G.D., & Lennox, R.D. (1990). The Reliability and Validity of Realtime Response Measures. In Bearden, W., Parasuraman, A., Eds., *Enhancing Knowledge Development in Marketing*. Chicago, IL: American Marketing Association, 284-288.
- Kaid, L.L. (2009). Immediate Responses to Political Television Spots in the U.S. Elections. Registering Responses to Advertising Content. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds., *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 137-153.
- Levy, M.R. (1982). The Lazarsfeld-Stanton Program Analyzer. A Historical Note. *Journal of Communication*, 32, 30-38.
- Maier, J. (2007a). Wahlkampfkommunikation und Wahlverhalten. In Rattinger, H., Falter, J.W., Gabriel, O.W., Eds., *Der gesamtdeutsche Wähler. Stabilität und Wandel des Wählerverhaltens im wiedervereinigten Deutschland*. Baden-Baden: Nomos, 385-411.
- Maier, J. (2007b). Erfolgreiche Überzeugungsarbeit. Urteile über den Debattensieger und die Veränderung der Kanzlerpräferenz. In Maurer, M., Reinemann, C., Maier, J., Maier, M., Schröder gegen Merkel. *Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*. Wiesbaden: Verlag für Sozialwissenschaften, 91-109.
- Maier, J. (2011). Führt der Einsatz von RTR-Technik zu einer anderen Wahrnehmung von Fernsehdebatten? Ergebnisse zweier Experimente zur externen Validität von Echtzeitmessungen. *Politische Psychologie* (im Druck).
- Maier, J., & Faas, T. (2004). Debattenwahrnehmung und Kandidatenorientierung. Eine Analyse von Real-Time-Response- und Paneldaten zu den Fernsehduellen im Bundestagswahlkampf 2002. *Zeitschrift für Medienpsychologie*, 16, 26-35.
- Maier, J., & Faas, T. (2009). Measuring Spontaneous Reactions to Media Messages the Traditional Way. Uncovering Political Information Processing with Push Button Devices. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds., *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 15-26.
- Maier, J., Maurer, M., Reinemann, C., & Faas, T. (2007). Reliability and Validity of Real-Time Response Measurement. A Comparison of Two Studies of a Televised Debate in Germany, *International Journal of Public Opinion Research*, 19, 53-73.
- Maier, M., & Maier, J. (2009). Measuring the Perception and the Impact of Verbal and Visual Content of Televised Political Ads. Results from a Study with Young German Voters in the Run-Up to the 2004 European Parliamentary Elections. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds., *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 63-84.

- Maier, M., & Strömbäck, J. (2009). Advantages and Limitations of Comparing Audience Responses to Televised Debates. A Comparative Study of Germany and Sweden. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds., *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 97-116.
- Maurer, M., & Reinemann, C. (2003). *Schröder gegen Stoiber. Nutzung, Wahrnehmung und Wirkung der TV-Duelle*. Wiesbaden: Westdeutscher Verlag.
- Maurer, M., Reinemann, C., Maier, J., & Maier, M. (2007). *Schröder gegen Merkel. Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*. Wiesbaden: Verlag für Sozialwissenschaften.
- Merten, K. (1994). Evolution der Kommunikation. In Merten, K., Schmidt, S.J., Weischenberg, S., Eds., *Die Wirklichkeit der Medien. Eine Einführung in die Kommunikationswissenschaft*. Opladen: Westdeutscher Verlag, 141-162.
- Meyer, V., & Ségur, C. (2009). Immigration as Translated by the Media. (Re)Production, Representation and Response to Images. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds., *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 193-206.
- Millard, W.J. (1992). A History of Handsets for Direct Measurement of Audience Response. *International Journal of Public Opinion Research*, 4, 1-17.
- Peatman, J.G., & Hallonquist, T. (1945). *The Patterning of Listener Attitudes Toward Radio Broadcasts. Methods and Results*. Stanford: Stanford University Press.
- Peatman, J.G., & Hallonquist, T. (1950). Geographic Sampling in Testing the Appeal of Radio Broadcasts. *Journal of Applied Psychology*, 34, 270-279.
- Peterman, J.N. (1940). The "Program Analyzer". A New Technique in Studying Liked and Disliked Items in Radio Programs. *Journal of Applied Psychology*, 24, 728-741.
- Poe, M.T. (2011). *A History of Communications. Media and Society from the Evolution of Speech to the Internet*. Cambridge: Cambridge University Press.
- Reinemann, C. (2007). Völlig anderer Ansicht. Die Medienberichterstattung über das TV-Duell. In Maurer, M., Reinemann, C., Maier, J., Maier, M., 2007, *Schröder gegen Merkel. Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*. Wiesbaden: Verlag für Sozialwissenschaften, 167-194.
- Reinemann, C., Maier, J., Faas, T., & Maurer, M. (2005). Reliabilität und Validität von RTR-Messungen. Ein Vergleich zweier Studien zur zweiten Fernsehdebatte im Bundestagswahlkampf 2002. *Publizistik*, 50, 56-73.

- Reinemann, C., & Maurer, M. (2005). Unifying or Polarizing? Short-Term Effects and Postdebate Consequences of Different Rhetorical Strategies in Televised Debates. *Journal of Communication*, 775-792.
- Reinemann, C., & Maurer, M. (2007). Populistisch und unkonkret. Die unmittelbare Wahrnehmung des TV-Duells. In Maurer, M., Reinemann, C., Maier, J., Maier, M., 2007, *Schröder gegen Merkel. Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*. Wiesbaden: Verlag für Sozialwissenschaften, 53-89.
- Reinemann, C., & Mauer, M. (2009). Is RTR Biased Towards Verbal Message Components? An Experimental Test of the External Validity of RTR Measurements. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds., *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 27-44.
- Roessing, T., Jakob, N., & Petersen, T. (2009). The Explanatory Power of RTR Graphs. Measuring the Effects of Verbal and Nonverbal Presentation in Persuasive Communication. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds., *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 85-95.
- Schwerin, H. (1940). An Exploratory Study of the Reliability of the "Program Analyzer". *Journal of Applied Psychology*, 24, 742-745.
- Tedesco, J.C., & Ivory, A. (2009). Health Message Primes and Sexual Health Campaign Messages. Evaluating Viewers' Real-Time Responses. In Maier, J., Maier, M., Mauer, M., Reinemann, C., Meyer, V., Eds., *Real-Time Response Measurement in the Social Sciences. Methodological Perspectives and Applications*. Frankfurt/M.: Peter Lang, 175-192.
- Thorson, E., & Reeves, B. (1986). Effects of Over-Time Measures of Liking and Activity During Programs and Commercials on Memory for Commercials. *Advances in Consumer Research*, 13, 549-553.
- Twyford, L.C. (1953). Profile Techniques for Program Analysis. *Audio-Visual Communication Review*, 1, 243-262.
- Van Eimeren, B., & Ridder, C.-M. (2011). Trends in der Nutzung und Bewertung der Medien 1970 bis 2010. Ergebnisse der ARD/ZDF-Langzeitstudie Massenkommunikation. *Media Perspektiven*. Heft 1/2011, 2-15.
- West, M.D., & Biocca, F.A. (1996.). Dynamic Systems in Audience Response Measures. In Watt, J.H., VanLear, C.A., Eds., *Dynamic Patterns in Communication Processes*. Thousand Oaks: Sage, 119-144.
- Wolf, B. (2010). *Beurteilung politischer Kandidaten in TV-Duellen. Effekte rezeptionsbegleitender Fremdmeinungen auf Zuschauerurteile*. Baden-Baden: Nomos.