

Sprechen in audiovisuellen Medien

Systematik für eine medienästhetische Analyse multimodaler Kommunikation

Clara Kindler-Mathôt, Cornelia Müller, Jana Katharina Junge, Katerina Papadopoulou

Abstract Der Beitrag illustriert anhand einer Fallstudie aus dem Korpus des DFG/NCN Forschungsprojektes »Multimodal Stancetaking«, wie sich die Multimodalität des Sprechens in audiovisuellen Medien untersuchen lässt. Ansatzpunkt der medienästhetischen (filmwissenschaftlich-linguistischen) Analyse ist die systematische Berücksichtigung der Verschränkung körperlich-sprachlicher und audiovisueller Ausdrucksformen, die das Material einfordert. Sprechen im medialen Raum entfaltet sich in audiovisuell orchestrierten Gestalten, die die Zuschauer*innen an den Bildschirmen als Wahrnehmungsgestalten sehen, hören und fühlen. Als Ausdrucksbewegungen modulieren und generieren sie Affekte. Es sind also die audiovisuell orchestrierten sprachlich-gestischen Ausdrucksformen, die eine kulturwissenschaftliche transdisziplinäre Perspektivierung dieser in vielfachem Sinne multimodalen medialen Ökologien notwendig machen. Dies erfordert ein radikales Umdenken der methodischen Zugriffe auf das Sprechen in audiovisuellen Medien.

Keywords Affekte; Ausdrucksbewegung; Methoden der Multimodalitätsforschung; Politische Kommunikation; Sprache und Gestik; Mediale Ökologien

1 Einleitung

Tagtäglich begegnet uns Sprache eingewoben in die verschiedensten medialen Ökologien. Ob als Nachrichtensendungen, Videospiele, TikTok-Videos oder Zoom Konferenzen – in der spezifischen medialen Form oder Orchestrierung

der audiovisuellen Bilder verschmelzen sprachlich-gestische Ausdrucksformen mit denen der audiovisuellen Formate.

Seit einigen Jahren widmet sich die Medienlinguistik solchen multi-modalen Phänomenen u.a. als Verflechtungen von Wort, Bild, Körper und Medium (vgl. z.B. Opiłowski 2017; Stöckl 2012). Viele Analysen verbleiben jedoch ausschließlich im methodischen Werkzeugkasten der Linguistik, z.B. Erzählen und Argumentieren in Social Media Kommentaren (Michel/Pfurtscheller 2021) oder Sprecherwechsel bei Zoom (Xia 2023). Auf diese Weise können zwar u.a. strukturelle Fragen der Gesprächsorganisation durchaus erfolgreich analysiert werden, weitestgehend unberücksichtigt bleibt jedoch, wie audiovisuelle Medien grundlegend die Wahrnehmung der Rezipient*innen strukturieren, wie sie Affekte generieren, modulieren und mobilisieren. Anders formuliert, die medienlinguistischen Analysen bedienen sich primär eines linguistischen Analyseinstrumentariums. Wie wichtig es ist, audiovisuelle Bilder auch als spezifischen Ausdrucksmodus ernst zu nehmen, möchten wir mit diesem Beitrag unterstreichen.

Filmwissenschaftliche Arbeiten zeigen, dass die Orchestrierung audiovisueller Bewegtbilder ständig neue Rahmungen kognitiver Prozesse hervorbringt, indem sie Zuschauer*innen unmittelbar affizieren (Kappelhoff 2018: 9). Die kinematografische Inszenierung erfahren Zuschauer*innen körperlich als ein geteiltes Zuschauergefühl (Bakels 2017; Kappelhoff 2004b; Kappelhoff 2008; Kappelhoff/Bakels 2011; Sobchack 1992). Bezogen auf die Multimodalität des Sprechens in audiovisuellen Medien gerät damit die ästhetische Gestaltung audiovisuellen Materials in den Blick. Wenn also Aussagen in kleine Videos verpackt, Körper in Kameraausschnitten gerahmt und abgeschnitten werden, wenn Videos nachträglich bearbeitet und geschnitten werden, wenn diese audiovisuellen Bilder dann eine bestimmte rhythmische Dynamik entfalten, wenn sie schlicht bestimmte Aspekte filmisch in das Zentrum der Aufmerksamkeit rücken, dann strukturiert diese Inszenierung sowohl die kognitive als auch die affektive Rezeption solcherart gestalteten Sprechens.

Aus diesem Blickwinkel stellen die Multimodalität des Sprechens in der Interaktion von Angesicht zu Angesicht und die audiovisuelle Multimodalität keine getrennten Kommunikationskanäle dar. Vielmehr formen sie gemeinsam eine vieldimensionale Wahrnehmungsgestalt, »an orchestrated multidimensional experiential gestalt« (Müller/Kappelhoff 2018: 45). Diese multidimensionale Erfahrungsgestalt entfaltet sich im Prozess der Wahrnehmung audiovisueller Bewegtbilder. Es ist ihre affektive Qualität, die die Zuschauer*innen unmittelbar körperlich berührt, mitnimmt, erregt oder beruhigt. In

diesem Sinne grundiert die körperliche Erfahrung der audiovisuellen Bilder sämtliche Prozesse der Bedeutungskonstitution einschließlich des dargestellten Sprechens.

Der vorliegende Beitrag stellt deshalb einen transdisziplinären (filmwissenschaftlich-linguistischen) Zugang zur Analyse der Multimodalität des Sprechens in audiovisuellen Medien vor. Der Ansatz macht die Wahrnehmungsperspektive auf solche medienspezifischen Sprachphänomene über das philosophisch-anthropologische Konstrukt der Ausdrucksbewegung theoretisch und empirisch greifbar. Damit »entgrenzt« er die linguistische Perspektive der Erforschung von Sprache in doppelter Hinsicht: vom Lautlich-Schriftlichen zum Multimodal-Mündlichen und von der Kommunikation von Angesicht zu Angesicht zur mediatisierten Mündlichkeit. Er präsentiert Ergebnisse des binationalen Forschungsprojektes »Multimodal Stancetaking: Expressive Movement and Affective Stance. Political Debates in the German Bundestag and the Polish Sejm«. Gegenstand der hier vorgestellten Studie ist eine Rede, die im Deutschen Bundestag im Rahmen der Haushaltsdebatte 2019 gehalten wurde.

Ansatzpunkt der medienästhetischen Analyse ist die systematische Berücksichtigung der Verschränkung körperlich-sprachlicher und audiovisueller Ausdrucksformen, die das Material einfordert. Das medial dokumentierte Sprechen entfaltet sich in audiovisuell orchestrierten Gestalten, die wir als Zuschauer*innen an den Bildschirmen der Laptops, Handys oder Fernsehgeräten als Wahrnehmungsgestalten sehen, hören und fühlen. Als audiovisuelle Ausdrucksbewegungen formen sie die Empfindungen der Zuschauer*innen, generieren und modulieren deren Affekte. Es ist somit die Beschaffenheit des Materials, es sind die audiovisuell orchestrierten sprachlich-gestischen Ausdrucksformen, die eine kulturwissenschaftlich transdisziplinäre Perspektivierung dieser in vielfachem Sinne »multimodalen« medialen Ökologien notwendig machen. Diese »Entgrenzung« der Perspektive auf sprachliche Phänomene erfordert ein radikales Umdenken auch der methodischen Zugriffe auf das Sprechen in audiovisuellen Medien. Darum geht es in diesem Beitrag.

Der filmwissenschaftlich-linguistische, im Kern medienästhetische Ansatz wurde im Rahmen verschiedener Forschungsprojekte kooperativ von Film- und Sprachwissenschaftler*innen (unter der Leitung von Hermann Kappelhoff und Cornelia Müller) entwickelt und vor allem am Beispiel multimodaler Metaphorik erprobt (Greifenstein et al. 2018; Horst 2018; Kappelhoff/Müller 2011; Müller/Kappelhoff 2018; Müller/Schmitt 2015). Er wurde in den vergangenen Jahren in einem deutsch-polnischen Forschungsprojekt

zum Thema »Multimodal Stancetaking: Expressive Movement and Affective Stance. Political Debates in the German Bundestag and the Polish Sejm«¹ weiterentwickelt. Im Zentrum dieser Forschungen steht die Frage nach der Analyse affektiver Positionierung der Redner*innen (*affective stancetaking*) in audiovisuell dokumentierten politischen Reden im Deutschen Bundestag und im polnischen Sejm. Auch hier dient das historische Konzept der Ausdrucksbewegung als theoretischer und methodologischer Bezugspunkt (Kappelhoff 2004b; Kappelhoff/Müller 2011; Müller/Kappelhoff 2018; Scherer et al. 2014; Schmitt et al. 2014; Horst et al. 2014).

Dass diese transdisziplinäre Perspektive sowohl als theoretische Motivation als auch als analytischer Bezugspunkt für die Analyse multimodalen Sprechens in audiovisuellen Medien fungieren kann, illustrieren wir anhand einer Fallstudie zu einer Rede im Deutschen Bundestag. Mit dieser Fallstudie dokumentieren wir zudem die wesentlichen analytischen Schritte der Datenauswahl, Aufbereitung und Auswertung, die wir im Rahmen des deutsch-polnischen Forschungsprojekts entwickelt haben. Theoretischer Rahmen und daraus abgeleitete Analysemethodik bauen auf dem oben genannten transdisziplinären filmwissenschaftlich-linguistischen Ansatz zur empirischen Analyse multimodaler Metaphoriken auf. Der Beitrag skizziert damit theoretische, methodologische und empirische Ergebnisse aus den oben genannten Forschungskontexten zu sprachlicher und audiovisueller Multimodalität. Er gliedert sich in die folgenden Abschnitte: (2) »Grundannahmen« skizziert den theoretischen Rahmen aus dem sich die im weiteren Beitrag vorgestellte medienästhetische Methode zur Analyse des Sprechens in audiovisuellen Medien ableitet; (3) »Sprachliche und audiovisuelle Multimodalität in der Analyse« stellt die medienästhetische Methode vor, wie sie im Rahmen des DFG/NCN Forschungsprojekts für die Analyse affektiver Positionierung in deutschen und polnischen Parlamentsreden weiterentwickelt wurde. Dabei illustriert Abschnitt (3) einerseits das analytische Prozedere für das gesamte Datenkorpus (3.1, Aufbereitung und Exploration der multimodalen Daten,) und illustriert andererseits am Beispiel der kurzen Parlamentsrede, wie sich das Konzept der »Ausdrucksbewegung« in konkrete, intersubjektiv überprüfbare

¹ DFG/NCN gefördertes Forschungsprojekt »Multimodal Stancetaking: Expressive Movement and Affective Stance« (MU 1587/3-1) unter der Leitung von Cornelia Müller (Europa-Universität Viadrina, Frankfurt (Oder), DE) und Maciej Karpiński (Adam-Mickiewicz-Universität, Poznań, PL) (2020–2024).

analytische Schritte überführen lässt (3.2, Die Analyse von Ausdrucksbewegungen als Ansatzpunkt für das Zusammenspiel von sprachlicher und audiovisueller Multimodalität). Der Beitrag schließt mit weiterführenden (4) Überlegungen zum Potential eines kulturwissenschaftlichen (filmwissenschaftlich-linguistischen) Zugriffs auf das Sprechen in audiovisuellen Medien in seinen je spezifischen medialen Ökologien.

2 Grundannahmen einer medienästhetischen Perspektive

Empirische Methoden leiten sich von theoretischen Grundannahmen zum Gegenstand, den sie adressieren, ab, bleiben dabei jedoch häufig implizit. Die im Folgenden skizzierten Grundannahmen formulieren medientheoretische Positionen, die für die Untersuchung sprachlicher Multimodalität in medienspezifischen Kontexten besonders relevant sind.

2.1. Audiovisuelle Bilder formen ästhetische Erfahrungen der Rezipient*innen

In einer solchen Perspektive sind audiovisuelle Bilder gerade nicht Repräsentationen einer unabhängig von ihnen existierenden Wirklichkeit, sondern formen die perzeptiven Schemata ihrer Rezipient*innen und modulieren deren Empfindungen. Es sind die Inszenierungen dieser Bewegtbilder, die unseren ›Blick‹ durch die Kamera auf die Welt strukturieren und die uns ganz unmittelbar berühren und bewegen. Diese Position der Zuschauer*innen als verkörpertes Durchleben und Erfahren audiovisueller Bilder gilt es analytisch einzuhören (Kappelhoff 2018: Kap. 3.3-3.5).

An dieser Erfahrungsqualität setzt die hier skizzierte medienästhetische Perspektive an. Medientheoretischer Bezugspunkt ist Kappelhoffs Konzept der filmischen Ausdrucksbewegung, das er in seiner kulturhistorischen Auseinandersetzung mit dem Theater der Empfindsamkeit und dem Hollywood Melodram (Kappelhoff 2004b) entwickelt hat. Ausdrucksbewegung bezieht sich auf die Expressivität filmischer Kompositionen als Bewegungserfahrung – und hierin liegt eine der Schnittstellen zwischen audiovisueller und sprachlich-gestischer Kommunikation (Kappelhoff 2004a; Müller/Kappelhoff 2018: Kap. 9).

Kappelhoff entwickelt das Konzept u.a. in Anknüpfung an Plessners philosophische Anthropologie, die von einem unmittelbaren, körperlichen Verste-

hen, einer interkorporalen Affizierung durch Bewegungserfahrung ausgeht.² Was ist damit gemeint? Zunächst einmal geht es darum, den Affekt als Qualität der Bewegung zu fassen, sei es im gestischen oder im filmischen Bewegungsbild. Plessners Idee der Ausdrucksbewegung verortet den Affekt, den Ausdruck *in der Bewegung selbst*. Die Bewegung ist hier gerade nicht Symptom innerer Vorgänge, die sich auf diese Weise einen Weg nach außen suchen, sondern sie *ist Affekt, sie ist Ausdruck*. Auch in der Phänomenologie Merleau-Pontys wird dieser Gedanke entwickelt: »[Ü]brigens fasse ich Zorn oder Drohung nicht als hinter den Gesten verborgene psychische Fakten, ich sehe vielmehr den Zorn der Gebärde an: sie lässt nicht lediglich denken an Zorn, sie ist der Zorn« (Merleau-Ponty 1968: 218). Das Verstehen dieser Ausdrucksgestalt ist dann auch ein unmittelbares, körperliches, verkörpertes Verstehen. Die Qualität der Bewegung einer ärgerlichen oder freudigen Geste wird unmittelbar erlebt, als bewegend auf welche Weise auch immer – heiter-schwingend, dramatisch-aufbrausend oder explosiv (vgl. auch Bühler 1933). Das Bewegungsverstehen, von dem hier die Rede ist, ist also ein verkörpertes Verstehen, ein leibliches in Merleau-Pontys Sinn.

2.2. Audiovisuelle Bilder entfalten sich als multidimensionale Erfahrungsgestalten

Sprachliche und filmische Multimodalität realisiert sich in Form multidimensionaler Erfahrungsgestalten. Diese Gestalten lassen sich theoretisch und analytisch als filmisch-sprachliche Ausdrucksbewegungen fassen. Sie entfalten sich als zeitlich strukturierte Bewegungen audiovisueller Bilder und sprachlich-gestischer Bewegung. Das hier zugrundeliegende Konzept der Ausdrucksbewegung bezeichnet einen historisch-spezifischen affektiven Modus, eine »Matrix der Gefühle«, die ihren Ausgang im französischen Theater des 18. Jahrhunderts nahm und prägend für die Gestaltung audiovisueller Gefühle im europäisch-amerikanischen Kino wurde (Kappelhoff 2004b; 2016)³.

-
- 2 Wie die Ausdrucksbewegung als methodischer Ansatzpunkt für die Analyse audiovisueller und sprachlicher Multimodalität fungieren kann, zeigen wir in diesem Band in einem weiteren Beitrag (Müller). Vgl. hierzu auch Müller/Kappelhoff (2018) sowie Greifenstein et al. (2018).
- 3 Vgl. hierzu auch die Arbeiten in den von Kappelhoff und Wedel herausgegebenen Reihen »Poetologien audiovisueller Bilder« und »Cinepoetics« bei De Gruyter sowie Müller/Kappelhoff (2018).

Wie eine ganze Reihe empirischer Studien aus den vergangenen Jahren belegen, strukturieren audiovisuelle Ausdrucksbewegungen jedoch auch andere mediale Formate wie etwa TV-Nachrichten, TV-Reportagen und Werbung für Baumärkte ebenso wie Wahlwerbung politischer Parteien, Musikvideos oder Social Media Kampagnen (Horst 2018; Müller/Kappelhoff 2018; Schmitt 2020; Scherer 2024). Audiovisuelle Ausdrucksbewegungen modulieren als Bewegungsbilder die kognitiven und affektiven Prozesse der Zuschauer*innen. Sie artikulieren ein ›Denken der Bilder‹, das das Denken der Zuschauer*innen rahmt, prägt und ihm eine affektive Tönung gibt. Affektivität wird dabei als Bewegungsqualität der sich entfaltenden Ausdrucksbewegungen körperlich erfahren und in Form von Affektdynamiken analytisch greifbar (Kappelhoff 2018: Kap. 1,3.2; Müller/Kappelhoff 2018: Kap. 8, 9; Scherer et al. 2014).

2.3 Mediale Ökologien strukturieren die Wahrnehmung von Sprache in audiovisuellen Medien

Multimodaler Sprachgebrauch in audiovisuellen Medien lässt sich nicht unabhängig von seiner medialen Ökologie erfassen. Überall, wo Sprache uns in medialen Kontexten begegnet – ob in Politreportagen, TikTok-Videos oder Videospielen – haben wir es mit komplex orchestrierten multimodalen Ausdrucksformen zu tun. Der vorliegende Beitrag illustriert, wie sich diese Verschränkung der Multimodalität des Sprechens mit der audiovisuellen Multimodalität mit Bezug auf das Konzept der Ausdrucksbewegung systematisch untersuchen lässt, und wie sich auf diese Weise spezifische Formen multimodaler Affizierung rekonstruieren lassen. Er nimmt damit die spezifischen medialen Ökologien multimodaler Sprachdaten in den Blick, welche erstens durch ihre spezifischen Ausdrucksformen geprägt sind und zweitens nicht ohne die Interaktion durch und mit Nutzer*innen zu denken sind. In der Interaktion zeigen sich die medialen Ökologien als »kulturelle Praxis, in der Menschen ihre subjektive Wahrnehmungswirklichkeit als Teil einer gemeinsamen geteilten Welt zu fassen suchen« (Kappelhoff 2018: 10). Durch die wechselseitige Bedingtheit (Reflexivität) von Ausdrucksformen und sozialer Praxis kann die Analyse der Modulation von spezifischen Ausdrucksformen als Ansatzpunkt für die Rekonstruktion von Prozessen der Zuschaueraffizierung nutzbar gemacht werden. Unterstützt wird dieser Aspekt der Reflexivität durch medientheoretische Arbeiten zu medialen Ökologien, die gerade auch die »Wechselwirkungen zwischen den sozialen,

politischen und psychischen Dimensionen von Medien und ihren materiellen Grundlagen« (Löffler/Sprenger 2016: 12) in den Blick nehmen.

Auf der einen Seite ist da die Multimodalität des Sprechens in der Interaktion von Angesicht zu Angesicht, die sich als dynamisches Zusammenspiel von Blick, Hand- und Körpergesten mit der gesprochenen Äußerung darstellt. Auf der anderen Seite ist da die audiovisuelle Multimodalität, d.h. die audiovisuelle Inszenierung des Sprachgebrauchs, die sich in der Orchestrierung von Kammerführung, Cadrage, Einstellungen, Montage und Ton realisiert (vgl. Müller, i.d.Bd.). Aus einer Wahrnehmungsperspektive formen beide Ebenen eine untrennbare Einheit, die sich zeitlich als Ausdrucksbewegung im Moment der Wahrnehmung, im Prozess des Zuschauens und Zuhörens entfaltet (Müller 2019; Kappelhoff/Müller 2011). Diese multidimensionale Wahrnehmungsgestalt ist der Ansatzpunkt für unsere Analyse. Sie erlaubt es uns zu rekonstruieren, wie verschiedene mediale Ökologien die Wahrnehmung von Sprache in audiovisuellen Medien strukturieren. Auf diesen Aspekt weisen auch Löffler und Sprenger hin: »Untersucht werden soll, wie sich die Strukturen der Wahrnehmung, des Denkens und des Verhaltens mit der Einführung neuer Medien auf diesen verschiedenen Ebenen verändern« (Löffler/Sprenger 2016: 12).

Wie sich diese Grundannahmen als Methodologie für die Analyse einer Multimodalität des Sprechens in audiovisuellen Medien konkret umsetzen lässt, illustrieren wir im nächsten Abschnitt.⁴

3 Sprachliche und audiovisuelle Multimodalität in der Analyse: Systematik der Analyse und Fallstudie

Der hier vorgestellte transdisziplinäre Ansatz verbindet eine filmwissenschaftliche Analyse audiovisuell generierter Affektdynamiken (»audiovisuelle Multimodalität«: empirische Medienästhetik, eMAEX⁵, AdA Filmontology;

-
- 4 Für eine detaillierte theoretische Einbettung verweisen wir auf Kappelhoff (2018), den Beitrag von Cornelia Müller (i. d. Bd.) sowie auf Müller und Kappelhoff (2018) und die oben genannten Publikationen, die im Umfeld der Arbeiten zur Kinematik audiovisueller und körperlich-sprachlicher Metaphorik (»Cinematic Metaphor«) entstanden sind.
- 5 »Electronically mediated analysis of expressive movement«: »eMAEX – Eine systematisierte Methode zur Untersuchung filmischer Ausdrucksqualitäten«, Online unter: <https://www.empirische-medienästhetik.fu-berlin.de/emaex-system/index.html> (30.01.2024).

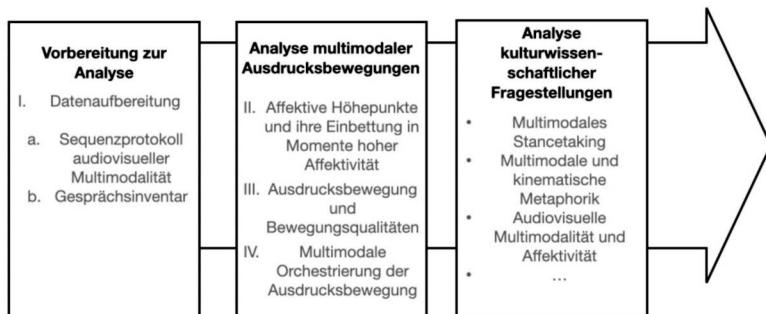
vgl. Müller/Kappelhoff 2018; Kappelhoff/Müller 2011; Kappelhoff/Bakels 2011; Bakels et al. 2020a) mit der linguistischen Analyse multimodaler Ausdrucksge-
stalten (»Multimodalität des Sprechens«: Diskursdynamik, Gestenforschung, Interaktionsforschung; vgl. Müller 2010, 2024a/b; Müller/Tag 2010; Bressem et al. 2013). Die im Folgenden dargestellte Systematik der Aufbereitung und Analyse des Sprechens in audiovisuellen Medien basiert auf diesen methodischen Grundlagen und wurde im Rahmen des DFG/NCN geförderten binationalen Forschungsprojekts »Multimodal stancetaking: Expressive movement and affective stance« weiter entwickelt. Sie umfasst drei zentrale, aufeinander aufbauende Analysephasen, die in Abbildung 1 in ihrem Verlauf dargestellt sind: (1) die Aufbereitung der Daten für die Analyse, (2) die detaillierte Analyse der audiovisuellen Ausdrucksbewegung und (3) die Analyse spezifischer Fragestellungen bezogen auf die Multimodalität des Sprachgebrauchs in audiovisuellen Medien. Im Rahmen des Forschungsprojekts zu parlamentarischen Reden ist dies die Frage danach, wie die aufgezeichneten Bilder des multimodalen Sprechens im parlamentarischen Kontext die Zuschauer*innen affizieren und wie sie die jeweiligen Positionierungen der Sprecher*innen im Sinne eines affektiven *Stance* strukturieren (vgl. Goodwin et al. 2012). In der hier präsentierten Fallstudie fokussieren wir darauf, wie im Zusammenspiel von sprachlicher und audiovisueller Multimodalität die Affekte der Zuschauer*innen generiert, moduliert und mobilisiert werden. Wir argumentieren, dass die Wahrnehmung der Multimodalität des Sprechens in audiovisuellen Medien grundsätzlich affektiv durchformt ist und dass dies in der Analyse sprachlich-gestischer Formen von Bedeutungskonstitution Berücksichtigung finden muss. Wie sich dies in der konkreten Analyse einholen lässt, zeigen wir im Folgenden.

Wir konzentrieren uns auf die ersten beiden Schritte der Analyse: (3.1) Datenaufbereitung und (3.2) Analyse multimodaler Ausdrucksbewegungen als methodischer Ansatzpunkt für die Analyse multimodaler Sprachdaten in audiovisuellen Medien. Die hierfür ausgewählte Rede ist Teil des Datensatzes des bereits erwähnten DFG/NCN Forschungsprojektes. Das Korpus besteht aus den Reden der Finanzdebatte 2019 des polnischen Sejm und des Deutschen Bundestages.⁶ Die als Videomaterial dokumentierten Reden

6 Im Rahmen des Projekts stehen Reden im Fokus, welche als offizielle Daten in den Archiven, wie etwa dem Parlamentsarchiv, gesammelt und für die Gesellschaft frei verfügbar sind. Solche Archive fungieren als »Gedächtnis« des Parlaments, aber sie laden auch die Menschen draußen ein, an den Debatten ihrer gewählten Vertreter*in-

umfassen insgesamt 681 Minuten und stellen die offiziellen Aufnahmen des Deutschen Bundestages und des polnischen Sejm dar, die in den Parlamentsarchiven für die Öffentlichkeit aufbewahrt werden. Die Aufnahmen sind frei zugänglich und bieten den Bürger*innen die Möglichkeit, Anteil an den parlamentarischen Debatten zu nehmen. Das deutsche Datenmaterial (323 Minuten Videomaterial) umfasst insgesamt 38 Redebeiträge von unterschiedlicher Länge (02:33-35:08 Min). Im Vergleich dazu beinhaltet das polnische Material (358 Minuten Videomaterial) insgesamt 105 deutlich kürzere Redebeiträge (01:00-26:00 Min). Bei der folgenden exemplarisch analysierten Rede handelt es sich um den Beitrag eines Mitgliedes der SPD-Fraktion (Dr. Jens Zimmermann). Sie hat eine Gesamtlänge von 5:32 Min.⁷

Abb. 1: Systematik der Datenaufbereitung und Analyse im Überblick



3.1 Aufbereitung und Exploration der multimodalen Daten

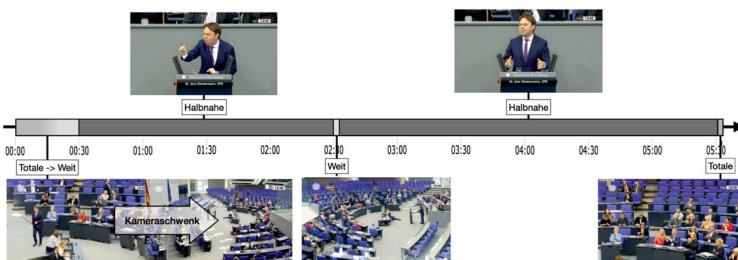
Der erste Schritt der Analyse besteht in der Aufbereitung der Daten. Dazu werden parallel ein Sequenzprotokoll der Rede, ein schriftliches Gesprächsinventar (vgl. Deppermann 2008: 31–35) in einer Excel Tabelle und eine Annotation in der Transkriptionssoftware Elan erstellt.

nen teilzunehmen. Das Projekt untersucht die audiovisuellen Materialien, die von den jeweiligen Archiven der Haushalts- und Finanzdebatten in den deutschen und polnischen Parlamenten zur Verfügung gestellt werden.

7 Die Rede ist vom 11.09.2019 und ist über die offizielle Website des Bundestages einsehbar: <https://www.bundestag.de/mediathek?videoid=7388015#url=L21lZGlh dGhla292ZXJsYXk/dmlkZW9pZD03Mzg4MDE1&mod=mediathek> (08.09.2023).

Das Sequenzprotokoll ermöglicht einen ersten Zugang zur Ebene der audiovisuellen Orchestrierung und wird im Sinne eines vereinfachten *Einstellungsprotokolls* (vgl. Hickethier 2012: 37f.) erstellt. Dazu liegt der Fokus auf den Wechseln der Kameraeinstellungen und -perspektiven über den Verlauf der gewählten Sequenz, wobei auch Aspekte der Bildraumkomposition zutage treten. Das filmische Vokabular orientiert sich ebenfalls an den standardisierten Bezeichnungen wie z.B. in der AdA Filmontologie (Bakels et al. 2020a) aufgeführt. Abbildung 2 zeigt das Einstellungsprotokoll der hier betrachteten Rede als Zeitstrahl.

Abb. 2: Darstellung der audiovisuellen Multimodalität: Sequenzprotokoll der Rede von Jens Zimmermann (Kameraeinstellungen und -perspektiven der Rede im zeitlichen Verlauf)



Die Rede weist eine, wenn auch zurückhaltende, audiovisuelle Inszenierung auf, die für die Aufnahmen des Korpus typisch ist. Größtenteils steht dabei Zimmermann im Zentrum des Bildraumes und damit auch im Zentrum der Aufmerksamkeit der Zuschauer*innen an den Bildschirmen. Die beiden Stills oberhalb des Zeitstrahls zeigen diese Perspektive aus sogenannter Normalsicht in halbnaher Einstellung. Nur einmal unterbricht ein Schnitt die Perspektive und die Kamera wechselt in eine weite Einstellung und zeigt das Parlament aus einer Aufsicht. Zimmermanns Stimme klingt währenddessen aus dem Off weiter. Die Rede ist von zwei weiteren Einstellungswechseln gerahmt, die mit dem Auf- und Abtreten Zimmermanns zum Pult einhergehen. Zu Beginn wird Zimmermann in seinem Gang nach vorne von der Kamera begleitet, um darauf in einem Schwenk und Zoom-Out das gesamte Parlament von links oben sichtbar werden zu lassen. Die letzte totale Einstellung ist erneut statisch. Vergleichbare Rahmungen zeigen sich im gesamten Datenkorpus der Bundestagsreden.

Das Gesprächsinventar ermöglicht einen ersten deskriptiven und kommentierenden Zugriff auf die gesamte Rede. Hier werden die Metadaten der Rede erfasst: Datum, Redner*in, und es erfolgt eine erste deskriptive Charakterisierung des/der Redner*in. In einer Sichtung wird die Rede weiter in interaktive Einheiten (*interactive units*) unterteilt. Die Einheiten sind entweder durch Beifall oder durch hörbare und/oder sichtbare Zwischenrufe begrenzt, die wiederum Reaktionen auf Seiten der sprechenden Person hervorrufen können. Die *Timecodes* dieser Einheiten dienen später der besseren Orientierung im Transkript und der Elan-Datei.

Im Gesprächsinventar wird die Rede zunächst als orthographisches Transkript festgehalten. Für Reden aus dem Bundestag kann dazu das online verfügbare Plenarprotokoll der Sitzung herangezogen werden. Die Transkription umfasst neben der Rede auch interaktive Aspekte wie Zwischenrufe und Applaus. In der Kommentarspalte werden parallel erste Beobachtungen von Aufälligkeiten der sprachlichen und audiovisuellen Modalitäten notiert. Auf diese Weise erschließt sich das vorliegende Material den Analysierenden in seiner Spezifik.⁸ Dies ist wichtig, denn damit kommt gewissermaßen das Material, in seiner audiovisuellen Orchestrierung multimodalen Sprechens, selbst zu Wort. Hypothesen werden so aus dem Material heraus entwickelt und nicht in Form präexistenter Fragen auf das Material projiziert.

Abb. 3: Gesprächsinventar „Zimmermann-Rede“

Time of Interactive Units	Speaker	Orthographic Transcription	General Comments
00:00:00-00:00:04	WK, JK	der SPD-Fraktion. (Johannes Kahr [SPD]: Bester Mann! – Beifall bei der SPD)	from WK's close-up to tracking shot of JZ from zooming out high right angle shot
00:00:05-00:00:09	(WK)	- (Vizepräsident Wolfgang Kubicki: Das sind sie doch! – Heiterkeit bei Abgeordneten der SPD)	high right angle shot further zooms out, podium adjustment sound
00:00:09-00:00:59	JZ	Sehr geehrter Herr Präsident. Meine Damen und Herren. Der Bundeshaushalt trägt die Überschrift „Investitionen und Chancen im Wandel“. Und wenn ich mir anschau, was in der Digitalpolitik in diesem Herbst alles passiert: Alleine in dieser Woche ist der Bericht der Wettbewerbskommission 4.0 vorgelegt worden. Das Digitalkabinett tagt. Wir werden den Digitalgipfel haben. Wir haben eine Konferenz der Vereinten Nationen, das Internet Governance Forum, hier in Berlin. Die Blockchain-Strategie der Bundesregierung wird vorgestellt und auch die Datenethikkommission präsentiert ihre Ergebnisse.	cam shot is still zooming out until furthest point, showing JZ stepping back and forth behind the podium, until cam changes to close-up of JZ; he engages in some sudden accentuated discursive beat gestures, employing also his head and upper body and occasional

Abbildung 3 zeigt die ersten drei Zeilen des Gesprächsinventars der Zimmermann-Rede. Die Kommentarspalte dokumentiert die oben beschriebene Kamerafahrt (vgl. Abb. 2) in zeitlicher Relation mit der sicht- und hörbaren Handlung. Hier rahmt sie die Zeitspanne, vom Aufruf des Redners bis zum Beginn seiner Rede am Pult. Sie zeigt den von Applaus und Zwischenrufen begleit-

8 Zur Reliabilität der Analyseschritte vgl. Abschnitt 3.2.1.

teten Weg des Redners zum Pult und den Beginn seiner Rede. Die Bewegung der Kamera und der Zoom auf den Sprecher am Ende der Kamerafahrt setzen den Auftakt der Rede in Szene. Im Gesprächsprotokoll werden zudem erste Auffälligkeiten der sprachlichen Modalitäten notiert. Die Abbildung verweist beispielsweise auf rhythmische diskursstrukturierende Gesten, die nicht nur mit den Händen ausgeführt werden, sondern »employing also his head and upper body [...]« (Abb. 3: General Comments). Diese Dokumentation erlaubt einen ersten materialbasierten, deskriptiven Zugriff auf die später ausführlich analysierte Orchestrierung potenzieller multimodaler Ausdrucksbewegungen.⁹

3.2 Die Analyse von Ausdrucksbewegungen als Ansatzpunkt für das Zusammenspiel sprachlicher und audiovisueller Multimodalität

Wir haben eingangs erwähnt, dass unsere kulturwissenschaftliche Perspektive auf die Multimodalität des Sprechens in audiovisuellen Daten die Position der Zuschauer*in einnimmt und dass audiovisuelle Bilder und die dort gezeigten Formen multimodalen Sprechens sich im Akt des Zuschauens als multidimensionale Erfahrungsgestalten entfalten. Diese Gestalten sind unsere methodischen Ansatzpunkte. Wir fassen sie als audiovisuelle Bewegungsbilder, deren Bewegungsentfaltung als audiovisuelle Ausdrucksbewegungen in ihren affektiven Qualitäten beschrieben werden (vgl. hierzu ausführlich Müller/Kappelhoff 2018: Kap. 9, sowie der Überblick bei Scherer et al. 2014).

Das heißt, in diesem Schritt der Analyse nehmen wir die Ausdrucksbewegungen als filmisch orchestrierte Bewegungsfiguren in den Blick (eMAEX; siehe Müller/Kappelhoff 2018, Appendix). Wir gehen dabei in drei Schritten vor: Zunächst fragen wir danach, wie man Ausdrucksbewegungen als audiovisuelle und sprachlich-körperliche Gestalten ›identifiziert‹ (3.2.1). Im nächsten Schritt beschreiben wir die affektive Qualität der audiovisuell orchestraerten Ausdrucksbewegung (3.2.2) und analysieren sodann im Detail, wie gestische, sprachliche und gesprächsorganisatorische Aspekte in der multimodalen Orchestrierung der Ausdrucksbewegung zusammen wirken (3.2.3).

⁹ Die detaillierte Annotation in Elan kann an dieser Stelle aus Platzgründen nicht ausgeführt werden. Sie wird jedoch im Rahmen der multimodalen Analyse immer wieder aufgegriffen.

3.2.1 Affektive Höhepunkte und ihre Einbettung in Momente hoher Affektdynamik

Wir nähern uns den audiovisuellen Daten in einer phänomenologischen Grundhaltung, d.h. die analytische Perspektive ist die einer betrachtenden – oder besser – zuschauenden Person am Fernsehgerät oder Laptop. Die Bewegungsbilder – bestehend aus den audiovisuellen Bildern, die Körperbewegungen zeigen und sich als zusammengehöriges, dynamisches Ganzes entfalten – weisen Bewegungsqualitäten auf, die von den Zuschauer*innen leiblich erfahren werden. Die Rede wird im Ganzen in Elan angesehen und währenddessen Bereiche markiert, die sich durch besonders hohe Affektivität vom Rest der Rede unterscheiden. Diese werden als affektive Höhepunkte in der Annotation festgehalten. Dabei folgt die »Identifikation« der Höhepunkte nicht dem Prinzip des Addierens einzelner Ausdrucksmodalitäten, sondern vielmehr dem der zeitlichen Entfaltung multidimensionaler Ausdrucksgestalten. Diese Höhepunkte treten als deutlich wahrnehmbare Veränderungen der Intensität von der sogenannten *baseline*, d.h. dem Grundrhythmus der Rede, als z.B. explosive, stetig steigende oder auch plötzlich abfallende Qualitäten hervor. *Baseline* bezeichnet dabei den spezifisch, individuell-charakteristischen Sprachduktus – oder die multimodale Dynamik des Sprechens, von der sich der Höhepunkt absetzt. Affektive Höhepunkte und *baseline* sind nur als aufeinander bezogene Qualitäten des multimodal orchestrierten Sprechens zu fassen. Ausgehend von diesen affektiven Höhepunkten wird die audiovisuelle Bewegungsgestalt beschrieben, in die diese eingebettet sind. Diese Gestalten werden in Elan festgehalten und als Momente hoher Affektdynamik bezeichnet. Sowohl die Identifikation der affektiven Höhepunkte als auch die Beschreibung der Affektdynamiken wurde intersubjektiv und interkulturell in Tandems abgesichert (*two independent coder procedure*)¹⁰. Die aus diesen Tandems hervorgehenden Dynamiken ähneln in ihrer Grundidee der Kendon'schen Segmentierung der zeitlichen Entfaltung gestischer Bewegungen in *preparation*, *stroke* und *recovery* (2004: 111f.): Die einzelnen Phasen der Bewegungsgestalten können, genau wie im Falle von Handgesten, spezifische

¹⁰ Selbst in bilingualen Tandems konnten im Rahmen des *Momo Stance* Projektes deutliche Überschneidungen in den Markierungen von Höhepunkten und den Momenten hoher affektiver Intensität erreicht werden. Dies spricht für den Ansatz des Zuschauer*innengefühls und gegen die Perspektive individueller Bewertung. Siehe hierzu auch die Anmerkungen bezüglich der Annotation audiovisueller Daten und »pattern-seeking« (Bakels et al. 2020b: 115; Herv. i. Orig.) im Rahmen der AdA Filmontology.

Bewegungskonfigurationen und Dynamiken entfalten. Die *preparation* etwa kann sich als langsamer und kontinuierlicher Aufbau oder aber als kurze, abrupte starke Rhythmisierung hin zum Höhepunkt darstellen. Das gleiche gilt für den affektiven Höhepunkt und das Abklingen in der Phase der *retraction*. Im Falle der politischen Reden beobachten wir oft ein Einhergehen von *retraction* und Ende einer interaktiven Einheit, in der der Moment hoher affektiver Intensität im Applaus verklingt, so auch bei Zimmermann. Die Dynamiken werden abschließend mit einem Vektorzeichenprogramm (hier Inkscape) in ihren Qualitäten, ihrem zeitlichen Verlauf und ihrer Position innerhalb der gesamten Rede visualisiert.

Abb. 4: Ein affektiver Höhepunkt (rot) eingebettet in einen Moment hoher Affektdynamik (orange) und weitere abgeschwächte affektive Höhepunkte (grau) in der Zimmermann Rede



Abbildung 4 verdeutlicht, dass in Zimmermanns Rede nur ein außerordentlich affektiver Höhepunkt identifiziert wurde, welcher in einen Moment hoher affektiver Intensität eingebettet ist. Dieser befindet sich etwa in der Mitte der Rede und hat eine Länge von 46 Sekunden. Der Höhepunkt (markiert in rot) beginnt mit einer stetig ansteigenden, affektiven Intensität in der Vorbereitungsphase und endet mit einem abrupten Abfall (markiert in orange). In grau sind vier Bereiche in der Rede abgebildet, die ebenfalls eine höhere affektive Intensität aufweisen und ihrerseits durch unterschiedliche Dynamiken gekennzeichnet sind. So wechseln sich langsame und stetige Anstiege mit plötzlichen Anstiegen und abrupten Abbrüchen ab. Die Abbildung dokumentiert, dass in der etwa fünfminütigen Rede von Zimmermann die affektive Dynamik etwa nach zwei Minuten Redezeit deutlich zunimmt und dann bis zum Ende der Redezeit in abgeschwächter, fast rhythmischer Form immer wieder präsent ist. Auf diese Weise lassen sich spezifische affektive Profile für die Reden der verschiedenen Sprecher*innen sichtbar machen. Hier fokussieren wir uns auf den Moment der Zimmermann Rede, der die höchste affektive Intensität aufweist.

3.2.2 Ausdrucksbewegungen und ihre Bewegungsqualitäten

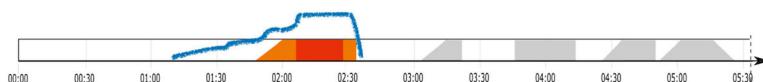
Ausgehend von den Momenten hoher Affektdynamik und den darin eingebetteten Höhepunkten lassen sich nun die audiovisuellen Ausdrucksbewegungen bestimmen (Müller/Kappelhoff 2018: Kap. 9; Scherer et al. 2014).¹¹ Auch bei der Ausdrucksbewegung haben wir es mit einer körperlich gefühlten und erlebten Einheit zu tun, die im Moment des Betrachtens der Rede entsteht. Sie entwickelt sich aus dem Zusammenspiel sprachlicher und audiovisueller Multimodalität zu einer Bewegungsgestalt. Diese Bewegungsgestalt ist ein sich zeitlich entfaltendes Bewegungsbild. Die Ausdrucksbewegung ist eine integrierte Ganzheit, deren affektive Qualität die Bewegungsqualität des sich entfaltenden Bewegungsbildes ist. Ihre Orchestrierung entsteht im Zusammenwirken der Multimodalität des Sprechens (semantische, pragmatische, syntaktische, gestische, interktionale, prosodische Aspekte) auf der einen und der audiovisuellen Multimodalität (Einstellungsperspektiven, Montage, Bildraumgestaltung, Kamerabewegungen) auf der anderen Seite. Affektivität entfaltet sich hier als Gestalt, als integriertes Ganzes. Als Ausdrucksbewegung bildet sie einen zusammengehörigen Intensitätsbogen mit einem dynamischen Rhythmus. Ihre Grenzen ergeben sich aus dieser Bewegungsgestalt. Die Analysen in Tandems haben nicht nur innerhalb der Analyseteams, sondern auch im deutsch-polnischen Vergleich eine hohe intersubjektive Übereinstimmung ergeben, was sowohl die Grenzen als auch die Qualitäten der Ausdrucksbewegungen als ganzheitliche multidimensionale Gestalten angeht.

Unsere vergleichenden Analysen im Projekt haben gezeigt, dass die Grenzen der Ausdrucksbewegung mit den im vorherigen Analyseschritt beobachteten Grenzen der Momente hoher affektiver Intensität zusammenfallen können, aber nicht müssen. Tatsächlich beginnen die spezifischen Dynamiken der Ausdrucksbewegungen häufig deutlich, *bevor* die affektive Qualität in besonderem Maße ansteigt und als Moment besonderer affektiver Intensität erkennbar wird. So verhält es sich auch mit der Ausdrucksbewegung im Beispiel von Jens Zimmermanns Rede. Abbildung 5 zeigt, wann und wie sich

¹¹ Ausdrucksbewegungen sind selbstverständlich nicht nur auf Momente hoher Affektdynamik beschränkt. Die Beschränkung ist dem analytischen Fokus des Projekts auf Momente hoher Affektdynamik geschuldet. Die Analyse der Qualität der Ausdrucksbewegung ermöglicht es dann, verschiedene Formen hoher Affektdynamik zu unterscheiden.

die Qualität des Ausdrucks um den Moment hoher affektiver Dynamik herum entfaltet (blaue Kurve).

Abb. 5: Affektiver Höhepunkt (rot) und Moment hoher affektiver Intensität (orange) eingebettet in die dynamische Entfaltung der Ausdrucksbewegung (blau)



Um die affektiven Dynamiken und Qualitäten der Ausdrucksbewegungen sprachlich zu erfassen, werden diese so beschrieben, dass der gefühlte und erlebte Eindruck der Dynamik möglichst bewahrt wird.¹² Die Forschenden nehmen damit ihre eigene körperliche Erfahrung der Rede zum methodischen Ansatzpunkt, denn auch sie sind grundsätzlich Zuschauende am Bildschirm. Die oben visualisierte Ausdrucksbewegung, in die der affektive Höhepunkt der Rede Zimmermanns eingebettet ist, entfaltet ihre affektive Qualität wie folgt:

Die Ausdrucksbewegung beginnt mit einem langsamen und stetigen Anstieg der Intensität, die sich in einem schneller und lauter werdenden Sprachgebrauch äußert und von akzentuierten gestischen Bewegungen begleitet wird. Gefolgt vom Gelächter des Publikums steigt die Intensität weiter. Die hohe Intensität wird gehalten und erstreckt sich in einer Art Plateau. Das Plateau ist von stark akzentuierten rhythmischen Bewegungen der Hände und des Oberkörpers gekennzeichnet und entfaltet so eine persistierende, repetitive, affektive Qualität. Eher abrupt – mit einem Lösen der Gerichtetheit des Körpers, einer tiefen Vorbeugung und einem Absenken der Stimme – lässt die Intensität der Ausdrucksbewegung nach.

Bei der sprachlichen Fassung der Bewegungsqualitäten stützen wir uns insbesondere auf ein Vokabular, das im Kontext der Beschreibung interaffektiver Prozesse in Mutter-Säugling Interaktion von Daniel Stern (2011) entwickelt und für die Beschreibung der affektiven Dynamiken und Qualitäten in der Filmanalyse von Sarah Greifenstein (2020) weiterentwickelt wurde.

¹² Die Beschreibung der Ausdrucksbewegung folgt dabei dem für die Filmanalyse entwickelten Verfahren der Analyse filmischer Ausdrucksbewegungen in Form von kurzen Vignetten, die die affektiven Qualitäten beschreibend erfasst (eMAEX).

Methodisch-deskriptiv fokussiere ich entsprechend Bewegungsverben (oder ihre Substantivierungen) für die gesamte Verlaufsform (z.B. das Klopfen) und Adjektive zur Präzision der sinnlichen Eigenschaften dieser Bewegung (das sanfte Klopfen, das eilig hastige Klopfen, das harte-energische Klopfen). (Greifenstein 2020: 69)

Die Qualitäten der Ausdrucksbewegungen werden als Qualitäten einer ganzheitlichen Gestalt greifbar. Sie zeichnen sich durch eine verschränkte Wahrnehmung von Bewegung, Kraft, Zeit, Raum und Gerichtetheit aus (Stern 2011: 13). Sterns Vokabular zielt dabei auf die Beschreibung der Formen, die die Bewegungsdynamiken ergeben. Es enthält Begriffe, die die Bewegung sowie die Intensität der Bewegung im zeitlichen Verlauf erfassen. Die sprachliche Beschreibung der sich in der Zeit entfaltenden Gestalt der Ausdrucksbewegung versucht – genuin phänomenologisch – die im Prozess des Zuschauens erlebten affektiven Qualitäten zu greifen. Dazu beschreibt sie die dynamischen Qualitäten und Intensitäten der situativen Bewegung. Sie konzentriert sich folglich auf die dynamische Bewegung und ihre sinnlichen Eigenschaften.

3.2.3 Multimodale Orchestrierung der Ausdrucksbewegung

An die Beschreibung der Bewegungsqualitäten der Ausdrucksbewegung schließt sich die detaillierte Analyse sprachlich-gestischer Multimodalität innerhalb der Ausdrucksbewegung an. Wie genau gestaltet sich das Zusammenspiel der sprachlichen und audiovisuellen Multimodalität in der aufgezeichneten Rede? Hier kommt die Spezifik der multimodalen Orchestrierung in den Blick, genauer das dynamische Zusammenspiel der Modalitäten, die den spezifischen Rhythmus einer Ausdrucksbewegung modulieren. Kappelhoff (vgl. 2008: 206) beschreibt diesen Rhythmus als die alles umspannende Ebene: Sie ist Struktur und Organisation des ästhetischen Wahrnehmungsvorgangs der kognitiven, affektiven und perzeptiven Prozesse des Zuschauers. Im Rhythmus, so Kappelhoff weiter, wird eine Bewegung zur Ausdrucksbewegung, zu einem kinematografischen Bild, in dem Gefühle nicht repräsentiert, sondern im Empfinden realisiert werden. Die Entfaltung dieses Rhythmus kann empirisch in der Analyse der multimodalen Orchestrierung der Ausdrucksbewegungen rekonstruiert werden. Neben dem Rhythmus von Handgesten und Intonation betrifft dies etwa auch rhythmische Veränderungen der Lautstärke, der Kamerabewegungen oder der Schnitttechniken, sowie interaktive Aspekte wie ein plötzliches Aufwallen von Applaus oder Zwischenrufen. Wie bereits angeführt, sind nicht immer alle Modalitäten

gleichermaßen an der Gestaltung dieses Rhythmus beteiligt. So zeigt der Blick auf die Ebene der audiovisuellen Multimodalität in der Zimmermann Rede eine klare Zurückhaltung in der Inszenierung. Die Kamera verbleibt durchgehend in einer stabilen Position und zeigt Zimmermann in halbnaher Einstellung von Kopf bis Hüfte am Rednerpult, welches zentral im Bildraum steht. Mit dieser Einstellung geht eine besondere Perspektivierung auf den Sprecher einher.

Die sprachlich-gestische, multimodale Orchestrierung der Ausdrucksbewegung und ihre Momente hoher Affektivität verweben sich mit den Inhalten der Rede. Wie diese Verschränkung sprachlicher und audiovisueller Multimodalität in solchen Momenten konkret funktioniert, wollen wir am Beispiel der Zimmermann Rede nun etwas genauer beschreiben. Worum geht es inhaltlich überhaupt? Und was ist der thematische Bezugspunkt des affektiv hochaufgeladenen Moments? Das zentrale Thema der gut fünf Minuten langen Rede ist die »Digitalpolitik der Bundesregierung«. Zimmermann beginnt mit einer Aufzählung der sogenannten »Tagesordnungspunkte« im Bereich der Digitalisierung des aktuellen politischen Geschehens und folgert, dass sie, die Sozialdemokrat*innen, diesen digitalen Wandel mitgestalten möchten. Nach abklingendem Applaus eröffnet er einen thematisch anschließenden Aspekt: klare Regeln auf dem Arbeitsmarkt für große digitale Konzerne. An dieser Stelle setzt die oben in ihren dynamischen Qualitäten beschriebene Ausdrucksbewegung an. Hier beginnt der kontinuierliche Anstieg der affektiven Intensitäten, den wir im Abschnitt 3.2.2 in seinen affektiven Qualitäten beschrieben und in Abbildung 5 als Bewegungsgestalt visualisiert haben. Abbildung 6 zeigt die Ausdrucksbewegung nun als gesprächsanalytisches Transkript.

Zimmermann lobt die Arbeit von Arbeitsminister Hubertus Heil, der den digitalen Strukturwandel aktiv gestalte (Zeile 001–010). Im weiteren Verlauf entfaltet sich interaktiv ein Missverständnis rund um Zimmermanns Empörung über die Behauptung, im Haushalt von Hubertus Heil würde »das Geld zum Fenster rausgeschmissen« (Zeile 011–014). Dieses beginnt mit Gelächter und Applaus aus dem Publikum (015), was zu einer Unterbrechung Zimmermanns führt.

An dieser Stelle beginnt die Entfaltung des affektiven Höhepunkts als Plateau hoher affektiver Intensität (visualisiert im gepunkteten Rahmen). Hier verdeutlicht Zimmermann erneut die Relevanz der Gestaltung des digitalen Strukturwandels (016–019) und endet mit dem Hinweis auf eine mögliche Konsequenz, sollte diese Relevanz ignoriert oder verharmlost wer-

den (019–020), wobei die affektive Qualität deutlich abnimmt. Das Ende der Ausdrucksbewegung ist dann deutlich markiert durch die Pause des Redners und Applaus im Publikum und zeichnet sich (im Gegensatz zum Beginn) durch einen abrupten Abfall der affektiven Intensität aus. Im Rest der Rede widmet Zimmermann sich weiteren Aspekten des Themenbereiches »Digitalpolitik der Bundesregierung«, es kommt aber zu keinem weiteren vergleichbar explosiven affektiven Ausbruch.

Abb. 6: Transkription der Ausdrucksbewegung (blaue Umrandung) nach GAT 2 in Verbindung mit Gistenannotation in der Sprachzeile: Nummerierte Gesture Units ({}{}) und Gesture Strokes (Fettsatz)

001	JZ	und deshalb (.) ist es GU:T, dass WIR-
002		mit hubertus HEIL einen sozialdemokratischen arbeitsminister haben,
003	JZ	¹ {der auch (.}) LIEfert an dieser stelle. wir [ham, ^h die NATIONale weiterbildungsstrategie-
004	P	[(Beifall)]
005	JZ	wir ham das qualifizierungs CHANengesetz- ^h wir ² {ham-
006		die DENKfab rik in der digitalen arbeit- und wir haben auch das arbeit von MORgen gesetzt.
007		und ³ {es ist WICHTig-} (.}) ⁴ {DIESEN digiTALen strukturWANDEL-
008		^h Aktiv zu ge) STALTEN; ^h ⁵ {denn die beSCHÄFtigt en,
009		^h die haben zu RECHT an einigen stellen auch beDENken ⁶ {wie es} weitergeht;
010		und ⁷ {deswegen ist es} WICHTig, ^h 7{dass wir dor}t} {A Kiv HAnDeln.}
011		¹⁰ {und wenn ich HÖre meine damen und herrn; ^h in HAUSHalt von hubertus hEIL-
012		dürde das geld zum FENster rausgeschmissen werden.
013		^h dann ist das nicht FALSCH- (.}) <>stockend> ^h geNUG- (.) zu- (> problematisIE:ren=
014		¹⁴ {weil das GELD- das geld--ja da können sie LACHen, da können sie LA}Chen meine] kollegen von der ef de: pe;
015	P	[(Gelächter, Applaus)]
016	JZ	¹¹ {aber an DIESer stelle den DIGItalen strukturWANDEL zu gestALTEN; ^h das ist eine MAMmutaufgabe;
017		^h und die MENsche die angst haben Ihnen job durch digitalisier) ung zu verlieren,
018		¹² {h um die müsste wir uns genAU) so kümmern, ^h wie die mensche die ang ¹³ {st haben durch den !KOH!leausstieg ihren} job zu verlieren;
019		¹⁴ {h und wenn wir uns SO um die kümmern; ^h so VERÄCHTlich wie SIE grade} gelAcht haben;
020		¹⁵ {h dann sind DAS [die nächsten]; ^h wie die extre[MIS]ten wählen werden meine damen und herrn.]
021	P	[(Applaus)]
		[das war nicht verächtlich) das war ein] netter SPREcher.]

Um die multimodale Orchestrierung der Ausdrucksbewegung gesprächsanalytisch zu fassen, wurde sie nach GAT 2 (Selting et al. 2009) transkribiert. Dies ermöglicht nicht nur einen Einblick in die sequenzielle Dynamik der Rede, sondern ist auch eine Voraussetzung für eine mikroanalytische Rekonstruktion der sprachlichen Multimodalität, hier insbesondere des Zusammenspiels von Rede und Geste innerhalb der Ausdrucksbewegung. Die Mikroanalyse fokussiert im vorliegenden Beispiel auf den Höhepunkt der affektiven Intensität in der Ausdrucksbewegung, also den Anstieg, das Plateau und den markant abfallenden Ausklang. Neben der Kennzeichnung der Momente hoher affektiver Intensität (Anstieg und Ausklang: Rahmen in orange, Plateau: Rahmen rot-gepunktet) wird das Transkript durch eine einfache Notation der Handgesten erweitert. Diese Notation orientiert sich am multimodalen Transkriptionssystem von Müller (1998: Appendix).

Wie sich die perzeptive Erfahrungsgestalt der Ausdrucksbewegung in ihrer multimodalen Orchestrierung genauer beschreiben lässt, illustrieren die folgenden Unterkapitel am Beispiel des Zusammenspiels interaktiver und sprachlich-gestischer Dynamiken (3.2.3.1) sowie einer Analyse der gestischen Komplexität (3.2.3.2). Beides sind Aspekte, die die Spezifität der Ausdrucksbewegung mit ihrem charakteristischen Bewegungs- und Intensitätsprofil in der Zimmermann Rede in besonderer Weise auszeichnen.

3.2.3.1 Ausdrucksbewegung als sprachlich-gestisches und interaktives Geschehen

Wenn wir uns die identifizierten und auf ihre expressiven Qualitäten hin beschriebene Gestalt der Ausdrucksbewegung nun mit Blick auf das sprachlich-gestische und interaktive Geschehen betrachten, dann zeigt sich, dass ganz dem Diskurstyp »Rede« entsprechend, Zimmermann der primäre Sprecher ist. Dies wird nicht nur auf gesprächsstruktureller Ebene, sondern auch in der filmischen Inszenierung deutlich, die Zimmermann durchgehend zentral im Bildraum aus einer frontalen, halbnahen Perspektive zeigt. Aus der Perspektive der Zuschauer*innen ähnelt dies einer direkten Anrede. Dennoch ist die Rede kein Monolog, sondern kann dialogisch verstanden werden, da sie interaktiv gemeinsam mit dem Publikum geformt wird. Dies zeigt sich deutlich an den drei Stellen, an denen das Publikum sich mit Redebeiträgen in die Rede »einklinkt«. Während der Beifall in Zeile 004 keinen Einfluss auf Zimmermanns Redefluss nimmt, führen Gelächter und Applaus in Zeile 015 zu einer Unterbrechung, die auch den Umschwung vom Anstieg zum Plateau des affektiven Höhepunkts innerhalb der Ausdrucksbewegung markiert. Dieser Moment ist somit Resultat eines *interactive boosts*, d.h. einer hohen interaktiven Dynamik, rund um einen »Versprecher« Zimmermanns.

Was passiert hier nun genau? Den Anstieg affektiver Intensität markiert eine doppeldeutige Formulierungen und ein stockender Redefluss. In Zeile 013 stolpert Zimmermann zunächst über eine Formulierung und sein Redefluss gerät ins Stocken. Dadurch kommt es zu einer Bedeutungsverschiebung. Die eigentlich intendierte Problematisierung der Behauptung des Geldverschwendens (Zeile 012) kehrt sich durch die intonatorische Markierung von »FALSCH-(.)« (013) als Phrasenende in eine doppelte Verneinung und damit in die gegenteilige Bedeutung: Die Behauptung ist wahr. Zimmermann repariert das semantische Problem durch eine syntaktische Weiterführung der Äußerung (013), die sich jedoch durch wiederkehrendes Stocken ungewöhnlich dehnt. Trotz dieser eigentlich erfolgreichen syntaktischen und semantischen Repara-

tur bricht laut hörbares Gelächter und Applaus aus (015). Zimmermann stockt erneut in seiner Äußerung, wiederholt diese und bricht sie dann schließlich ab: »=weil das GELD- das geld-« (014). Dies markiert den Beginn des affektiven Höhepunkts, das Plateau erstreckt sich dann insgesamt über die Zeilen 014–19. Zimmermann wendet sich körperlich und pragmatisch direkt den »Störern« zu, indem er sich von seiner bisher mittigen Ausrichtung mit seinem gesamten Körper nach rechts orientiert und Richtung Publikum spricht »da können sie LACHen, da können sie LACHen meine kollegen von der ef de: pe:« (014). Deutlich lauter und schneller spricht er nun mit wiederkehrenden Überbetonungen direkt in Richtung der FDP (die jedoch nicht im Bild sichtbar wird). Das Plateau hoher affektiver Intensität geht einher mit seiner Anklage an die FDP, dem Vorwurf, sie würde die Dringlichkeit der Lage erkennen und die Bedürfnisse der Menschen lächerlich machen. Die politische Konsequenz aus diesem Fehlverhalten bildet den Abschluss der interaktiven Sequenz und der Ausdrucksbewegung: »°h dann sind DAS die nächsten, °h die die extre!MIS!ten wählen werden meine damen und herrn« (Zeile 020).

Zimmermann markiert das Ende seines Redebeitrages mit einer Wendung des Körpers zurück zur Mitte, senkt den Blick auf seine Notizen und schaut dann erneut in das gesamte Publikum, bzw. zu uns als Zuschauer*innen an den Bildschirmen. Währenddessen ertönt erneuter Applaus, ebenso ein Zwischenruf, vermutlich aus Reihen der FDP, der versucht, das Missverständnis noch aufzulösen (»das war nicht verächtlich das war ein netter verSPRECHer-«) und die Ausdrucksbewegung findet ein eher abruptes Ende.

Die Ausdrucksbewegung mit der langsam ansteigenden und dann gleichbleibend hohen affektiven Intensität und einem plötzlichen Abfall der Affekt-dynamik am Ende wird somit interaktiv getriggert. Die Unterbrechung durch das Publikum wirkt gleichsam wie ein interaktiver *boost*, ein interaktives Anheizen der affektiven Dynamik, und leitet den abrupten Wechsel nicht nur des Sprechaktes, sondern auch der Dynamik ein.

Die multimodale Orchestrierung der Ausdrucksbewegung bezogen auf die Multimodalität des Sprechens zeichnet sich dann durch eine deutliche Intensivierung der Lautstärke, der Sprechgeschwindigkeit und auch des Sprachrhythmus im affektiven Höhepunkt, sowie eine Zunahme komplexer gestischer Bewegungen aus. Die Ausdrucksbewegung ist somit durch einen pragmatischen *shift* vom Erklären zum Anklagen und eine Veränderung in der Rhythmisität im affektiven Höhepunkt charakterisiert. Wie sich dies in

der Entfaltung zeigt, verdeutlicht das Transkript in Abbildung 7.¹³ Es zeigt den Beginn des Moments hoher affektiver Dynamik (orange Umrandung) und den Moment des plötzlich explodierenden Höhepunkts (rot gepunktete Umrandung), ausgelöst durch den interaktiven Trigger (Applaus und Gelächter).

Abb. 7: Transkription gestisch-sprachlicher Orchestrierung im Übergang vom Anstieg (orange) zum Plateau (rot-gepunktet) der Ausdrucksbewegung

Dass dieser affektive Höhepunkt sprachlich und gestisch orchestriert ist, zeigt sich anhand einer Mikro-Analyse der verwendeten Gesten. So beobachten wir am Übergang vom Anstieg zum Plateau der Ausdrucksbewegung zwei verschiedenen Handformen, welche im Transkript als S₂ und S₃ im zeitlichen Verlauf notiert und in kleinen Zeichnungen zusätzlich visualisiert werden. Es

13 Die gestische Notation unterhalb der verbalen Transkription orientiert sich an Müllers (1998) Notation und zeigt die Handgesten als Zeitformen mit Phasen der *Preparation* (P), *Stroke* (S) oder *Multi-Stroke* (MS) und *Retraction* (R), in ihrer Relation zur sprachlichen Äußerung.

handelt sich um zwei Gesten mit je verschiedenen kommunikativen Funktionen: eine rekurrente Geste (*S₂*), genauer den gestischen Präzisionsgriff (»precision grip«; Müller 2014; Kendon 2004) und eine Zeigegeste (*S₃*), hier realisiert mit ausgestrecktem Arm und Zeigefinger. Dabei wird deutlich, dass der Präzisionsgriff *S₂* in der Anfangsphase der Anstiegs der affektiven Dynamik deutlich überwiegt und *S₃*, also die Zeigegeste, exakt im Moment des pragmatischen Umschwungs von Erläuterung zu Anklage auftritt (Zeile 014). Dieser Umschwung von der dynamisch ansteigenden affektiven Qualität zum Plateau hoher Affektdynamik innerhalb der Ausdrucksbewegung realisiert sich damit nicht nur sprachlich, sondern auch auf der Ebene der kommunikativen Funktionen, der gleichzeitig verwendeten Gesten. Wie genau lässt sich dies nun aus Sicht der Gestenforschung beschreiben und begründen?

Die Ringgeste ist Teil des Repertoires rekurrenter Gesten und ist regelmäßig in metakommunikativer modaler Funktion zu beobachten, wo sie die Präzision einer sprachlichen Aussage artikuliert (Bressem/Müller 2014: 1583). Ihre kommunikative Funktion leitet sich aus der präzisen Greifbewegung ab, die erforderlich ist, um mit Zeigefinger und Daumen kleine Objekte ergreifen zu können. Sie wird daher auch als Präzisionsgriff beschrieben (Bressem/Müller 2014: Kap. 1583; Kendon 2004; Müller 2014). Wenig überraschend wird sie deshalb häufig im Kontext politischer Diskussionen verwendet (vgl. auch Neumann 2004). Mit der Ringgeste heben Sprecher*innen spezifische Aspekte des Gesagten nicht nur hervor, sondern behaupten gestisch deren Präzision und hohe Relevanz. So auch in der im Transkript dargestellten Äußerung, in der Ringgesten rhythmisch in einzelnen Schlägen auf Fokusakzente (HAUSHalt, FENster, geNUG, GELD) gesetzt oder in rhythmischen Auf- und Abbewegungen über längere Äußerungsteile getragen werden (z.B. »da würde das geld zum«). Die Ringgeste zieht sich im Wechsel zwischen einzelnen gestischen Höhepunkten und repetitiven Bewegungen über die komplette gestische Einheit, bis sie zum Ende von der Zeigegeste mit ausgestrecktem Arm abgelöst wird. Die rhythmisierte Verwendung der Ringgeste orchestriert somit den Anstieg der affektiven Dynamik. Der Wechsel von Ring- zu Zeigegeste markiert die Anfangsphase des Höhepunkts affektiver Intensität, und er geht mit einem Wechsel der kommunikativen Funktion der Handgeste einher. Nun werden die Störer auf den Seiten der FDP gestisch aufgezeigt (»ja da können sie LACHen«). Allerdings verschiebt sich die Funktion der Zeigegeste im Verlauf des affektiven Höhepunktes vom Hinweisen auf die Quelle der Störer im Publikum (d.h. einer referentiellen Zeigegeste) zum anklagenden Zeigen (d.h. eine pragmatische Zeigegeste) (Kendon 2004: 159) (»°h und wenn wir uns SO um die küm-

mern; „h so verÄCHTlich wie SIE gerade gelAcht haben; „h dann sind DAS die nächsten, „h die die extre!MIS!ten wählen«). Mit dem Wechsel von der Ringzur Zeigegeste, der hohen Bewegungsdynamik der rhythmisierten Zeigegesten, der ostentativen Ausrichtung des Oberkörpers und Kopfes auf die Gruppe der Störer parallel zur hochaufgeladenen sprachlichen Anklage entfaltet sich der Höhepunkt der Affektdynamik.

Wir halten fest: Auch mit Blick auf die Entfaltung der Gesten zeigen sich die unterschiedlichen Rhythmisierungen, affektiven Dynamiken der Ausdrucksbewegung in den drei Phasen Anstieg, Plateau und Abfallen. In ihren wechselnden kommunikativen Funktionen (Präzision der Aussage, Identifikation der Störenden und Anklage) verstärken sie die Sprechhandlungen. Sie haben zudem durch die Verkörperung der dynamischen Bewegungsqualitäten einen erheblichen Anteil an der Formierung der Ausdrucksbewegung.

3.2.3.2 Ausdrucksbewegung und gestische Komplexität

Bereits bei einem ersten Blick auf die Sequenz wird deutlich, dass sich die hohe affektive Intensität der Ausdrucksbewegung sowie ihr dynamisches Profil mit einer Ausdehnung gestischer Einheiten und einer Häufung gestischer Höhepunkte (*Strokes*) verbindet. Der folgende Abschnitt betrachtet daher etwas genauer, wie sich Formen gestischer Komplexität über den Verlauf der gesamten Ausdrucksbewegung entfalten.

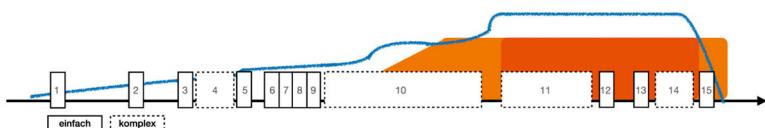
Müller (2024 a/b) betrachtet Gesten, aufbauend auf Kendons (2004) Ansatz, als lineare Zeitformen, die aus einer analytischen Perspektive in gestische Einheiten (*gesture units*) unterteilt werden können. Eine gestische Einheit ist dabei durch Ruhepositionen (rp) begrenzt (Kendon 2004: 111f.). Gestische Einheiten können weiter in einfache oder komplexe Formen unterschieden werden. Die Unterscheidung ist abhängig von der Anzahl der gestischen Phrasen (*gesture phrases*), die sich innerhalb einer gestischen Einheit entfalten. Gestische Phrasen umfassen, in Abweichung zu Kendon¹⁴, Vorbereitung, Höhepunkt und Rückbewegung der Geste. Eine einfache Einheit beinhaltet nur eine Phrase, während eine komplexe Einheit zwei oder mehr Phrasen umfasst (z.B. die in Abb. 7 visualisierte gestische Einheit 10).

Abbildung 8 zeigt eine schematische Darstellung der gestischen Einheiten und ihrer Komplexität, die die Ausdrucksbewegung in der zunehmend affektiv aufgeladenen Sequenz auszeichnet. Dargestellt auf dem Zeitstrahl als kleine

14 Kendon schließt in die gestische Phrase nur Vorbereitung und Höhepunkt der Bewegung ein (vgl. Kendon 2004: 111ff.).

Boxen, reihen sich insgesamt fünfzehn gestische Einheiten. Eine durchgezogene Linie verweist auf eine einfache, eine gestrichelte Linie auf eine komplexe gestische Einheit. Bereits auf dieser Ebene zeigt die Ausdrucksbewegung von Zimmermann eine deutliche Steigerung der Komplexität, insbesondere mit dem Beginn der ansteigenden affektiven Dynamik (orange). Während die gestischen Einheiten 1 bis 9 (mit Ausnahme von Einheit 4) kurze, simple Formen darstellen, die außerdem zu Beginn durch anhaltende Ruhephasen getrennt sind, sind die gestischen Einheiten 10 bis 15 deutlich komplexer und länger.

Abb. 8: Steigende gestische Komplexität im Verlauf der Ausdrucksbewegung



Diese mit der Entfaltung der Ausdrucksbewegung zunehmende Komplexität der gestischen Sequenzen wird auch mit Blick auf die gestische Phrasenstruktur deutlich. Zum einen steigt ihre Anzahl auf bis zu elf Phrasen in einer gestischen Einheit an und bildet damit eine rhythmisierte Bewegungssequenz (vgl. Einheit 10). Zum anderen werden die gestischen Phrasen an sich komplexer, da sie zunehmend aus Phrasen mit komplexen gestischen Höhepunkten (*multi-strokes*) bestehen. Im Gegensatz zu einfachen Single-Stroke Phrasen (in Abb. 7 als S), umfassen diese mehrere Stroke Phrasen. Beispiele sind repetitive Gesten, wie die Kurbel-Geste (Ladewig 2014) oder das Nachzeichnen komplexer Formen. Alle komplexen gestischen Phrasen der Ausdrucksbewegung in der Rede Zimmermanns stellen Formen diskursiver Taktgebungen, *discursive beatings* oder *parsings* (Kendon 2004: 159) dar, also das rhythmische Auf- und Abbewegen der Hände (vgl. gestische Einheit 10).

Betrachten wir zudem die Entfaltung der gestischen Komplexität bezogen auf den fließenden Übergang verschiedener Handformen, dann zeigt sich, dass am Beginn der Ausdrucksbewegung einzelne gestische Betonungen und einfache gestische Einheiten (1–4, 6–9) überwiegen. Die Handformen sind hier vergleichsweise unspezifisch. Mit offener, lateral ausgerichteter Handfläche (PLOH; *palm lateral open hand*; vgl. auch Annotationssystem LASG von Bressem et al. 2013) werden rhythmische Akzente als Auf- und Abbewegungen in Relation zur Äußerung gesetzt. Häufig fallen diese auf die Fokusakzente (*die*

DENKfabrik (006), ist WICHTig- (007)) oder diskursstrukturierende Elemente wie »deswegen« (010). Sie stellen klassische Formen diskursiven *beatings* dar und haben akzentuierende und redestrukturierende Funktion (Kendon 2004: Kap. 9). Wie lässt sich diese Zunahme der gestischen Komplexität über den Verlauf der Ausdrucksbewegung genauer beschreiben?

Am Beginn der Ausdrucksbewegung sehen wir viele kleine Gesten die im seitlichen Gestenraum, teilweise halb verborgen durch die Mikrofone, ausgeführt und schnell in Ruhephasen zum Rednerpult zurückgeführt werden. Mit der Zunahme der komplexen gestischen Phrasen verändert sich der Rhythmus von diesen kleinen Akzenten zu einer betonenden und stetig ratternden Dynamik. Dies geht auch mit dem Wechsel der Handform einher. So beobachten wir die pragmatische Ringgeste ausschließlich innerhalb komplexer Einheiten (4 und 10) und überwiegend in Multi-Stroke Phrasen mit einer Variabilität von zwei bis sechs Schlägen pro Phrase. Zusätzlich rückt die gestikulierende Hand weiter nach oben in das Zentrum des Gesten- und Bildraums und gerät damit ins Zentrum des Blickfeldes der Zuschauer vor Ort und vor den Bildschirmen. Während die Hände zuvor in die Ruheposition zurückgekehrt sind, verbleiben sie nun zwischen den einzelnen gestischen Taktschlägen in gehaltener Handform im zentralen Gestenraum. Der abrupte Wechsel der Handform zur Zeigegeste führt auch zu einer neuen Dynamik der gestischen Taktschläge. Deutlich eindringlicher und vor allem raumeinnehmender sticht der ausgestreckte Finger weit in den peripheren Gestenraum. Durch einen kleinen Schritt zurück ist Zimmermanns ausgestreckter Arm nun plötzlich mittig im Bildraum. Das Vor- und Zurückstrecken des Armes ist so in seiner vollen Länge zu sehen und erlangt dadurch hohe Prominenz. Es ist unmöglich, das anklagende Zeigen nicht zu sehen. Durch die neue Position sind auch die Vorbereitungsphasen aus den Ruhepositionen deutlich präsenter, sie wirken nun wie ein Anlaufnehmen für die ausladende Bewegung. Das rhythmische Rattern ist nun zu einem ausladenden Stechen geworden, das sich in ausholenden Bewegungen rhythmisch etabliert. Die letzte Zeigegeste (gestische Einheit 15) markiert das Ende des affektiven Ausbruchs wie einen Abschlusspunkt. In einer kurzen, abrupten Vorwärtsbewegung streckt Zimmermann die Hand mit ausgestrecktem Zeigefinger nach vorne, als würde er der FDP gewissermaßen »vor den Kopf« stoßen. Diese letzte Zeigegeste bildet den rhythmischen Abschluss des affektiven und argumentativen Höhepunkts, gefolgt von einem akzentuierten Nicken des Kopfes und einer nachdrücklichen Vorwärtsbewegung des gesamten Oberkörpers, all dies zeitgleich mit dem deutlich hervortretenden Fokusakzent »extre!MIS!ten« (Abb. 6: Zeile 020).

Betrachtet man die Entfaltung gestischer Komplexität im Zusammenspiel mit der prosodischen Dynamik über den Verlauf der gesamten Ausdrucksbewegung, so zeigt sich in ihrer Rhythmisierung die zunehmende Verdichtung gestischer Bewegungsmuster. Wir sehen eine Verschiebung von eher kleinen, unscheinbaren Gesten, die seitlich vom Körper und am Rande des Bildausschnittes ausgeführt werden (am Beginn der Ausdrucksbewegung), zu sehr raumgreifenden gestischen Bewegungen, die vor dem Körper deutlich sichtbar im Zentrum des Bildraums ausgeführt werden und so den affektiven Höhepunkt als ein Plateau, das sich über einen längeren Zeitraum hinweg erstreckt, orchestrieren und die Ausdrucksbewegung schließlich mit einer letzten pointierten Zeigegeste zu einem abrupten Ende führen.

3.3 Zusammenfassung

Die hier vorgestellte Systematik der Analyse sprachlicher und audiovisueller Multimodalität zeigt am Beispiel von Parlamentsreden, wie sich die Frage nach der Generierung, Mobilisierung und Modellierung von Zuschaueraffekten systematisch rekonstruieren lässt. Ausgangspunkt ist dabei das Konzept der Ausdrucksbewegung, das einen phänomenologischen Zugang zum Datenmaterial eröffnet. Einen Zugang, der die leibliche Erfahrung der Forschenden zum (intersubjektiv überprüfbaren) Ansatzpunkt für die Analyse audiovisueller multimodaler Daten nimmt. Dass dies keine esoterische Behauptung ist, haben die Analysen im deutsch-polnischen Forschungsprojekt zur affektiven Positionierung von parlamentarischen Redner*innen zeigen können. Am Beispiel der Analysen von Momenten hoher Affektdynamik ließen sich zweifelsfrei multidimensionale Erfahrungsgestalten am Material herauspräparieren, die dann einer detaillierten Mikro-Analyse unterzogen wurden. Schon der Blick auf eine Rede zeigt, dass sich auf diese Weise Affektdynamiken verschiedener Intensität und unterschiedlicher dynamischer Verläufe ausfindig machen lassen. Die Beschreibung der Bewegungsqualitäten der Ausdrucksbewegung erfolgt sodann als Beschreibung eines audiovisuell gestalteten Bildes, d.h. die Analyse des Sprechens und Gestikulierens, der Stimm- und der Körperbewegungen ist immer audiovisuell gerahmt, orchestriert. Zuschauer*innen an den Bildschirmen, auch wenn sie Forschende sind, sehen den/die Sprecher*in so, wie sie ihr von der Kamera (von der Regie) gezeigt wird. In unserem Fall scheint das unspektakulär: Wir sehen Zimmermann immer in der halbnahen Einstellung – nie schneidet die Regie auf die Totale (und verkleinert damit die Körperbewegungen auf Miniaturformat) – aber auch in dieser Einstellung ist unsere Ana-

lyse letztlich die eines Bildraumes, in dem wir den Oberkörper des Redners in frontaler Ansicht zu sehen bekommen. Manche Gesten sind von Mikrofonen verdeckt, die Interaktion mit dem Publikum können wir nur erahnen, denn auch diese »liefert« uns die audiovisuelle Dokumentation der Rede nicht. Das heißt mithin, wenn wir von Ausdrucksbewegungen und ihren Bewegungsqualitäten als affektiven Qualitäten sprechen, dann sind dies immer Qualitäten, die als ganzheitliche Erfahrung eines audiovisuellen Bildes beschrieben werden. In diesem Sinne formulieren wir also die Position der Zuschauenden. Wir beschreiben damit unmittelbar, wie diese Bewegungsbilder uns berühren, bewegen, wie wir sie empfinden.

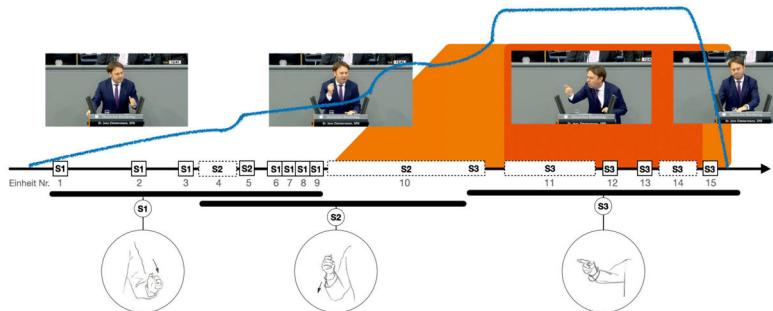
An diese phänomenologische Beschreibung schließt sich sodann eine Mikro-Analyse dessen an, was wir multimodale Orchestrierung der Ausdrucksbewegung nennen. Mit dieser Mikro-Analyse stützen wir die analytischen Einheiten, die in der phänomenologischen Analyse ausgemacht wurden, und beschreiben, wie sich die affektiven Dynamiken sprachlich-gestisch und in der audiovisuellen Inszenierung realisieren. Im vorliegenden Fallbeispiel stechen hier insbesondere die Rolle des interaktiven Geschehens sowie die sprachlich-gestischen Dynamiken als markant hervor. So zeigt sich, wie interaktive Trigger (hier die Einwürfe, Gelächter des Publikums), die sprachlich-gestische »Doppelung« der kommunikativen Funktionen (von der Beschreibung zur Anklage), die Vergrößerung der entsprechenden Gesten, die damit ins Zentrum der visuellen Aufmerksamkeit rücken (*gestisches foregrounding*; Müller/Tag 2010) sowie die bemerkenswerte Zunahme an temporaler gestischer Komplexität (längere und komplexere gestische Einheiten) den Verlauf der Ausdrucksbewegung (gleichmäßiger Anstieg, Plateau hoher Intensität und abruptes Ende) ausmachen, ihn orchestrieren.

Abbildung 9 fasst die Analysen der multimodalen Orchestrierung noch einmal zusammen. Im Verlauf der Ausdrucksbewegung können drei verschiedene Handformen identifiziert werden: S₁, die *palm lateral open hand* (PLOH); S₂, der Ring (Daumen und Zeigefinger geschlossen); und S₃, eine Zeigegeste mit ausgestrecktem Zeigefinger. Mit Blick auf die kommunikativen Funktionen zeigt sich, dass alle Gesten vordergründig pragmatische Funktionen haben.

Die multimodale Orchestrierung der Ausdrucksbewegung ist durch eine anwachsende Intensivierung charakterisiert, die sich als zunehmende Komplexität der Gesteneinheiten (von einfachen zu komplexen Einheiten) und Gesten-Phrasen (deutlich mehr Multi-Stroke Phrasen im Höhepunkt) realisiert. Gleichzeitig zeichnen sich die gestischen Bewegungen durch eine Zunahme der Rhythmisierung aus. Auch auf der Ebene der kommunikativen Funktion

lässt sich eine dynamische Veränderung beobachten, die mit der affektiven Qualität der Ausdrucksbewegung einhergeht. Während die Sequenz mit einer rein rhythmischen Betonung ohne spezifische Handform beginnt, wechselt dies im Verlauf der Sequenz über die Ringgeste, die die Präzision der zeitgleichen sprachlichen Äußerung behauptet, um plötzlich einen Sprechakt des Vorwurfs und der Anklage zu entfalten und schließlich abrupt in einem gestischen Abschlusspunkt zu enden.

Abb. 9: Gestische Formen mit pragmatischen Funktionen im zeitlichen Verlauf



Bezogen auf die audiovisuelle Multimodalität wird deutlich, dass trotz der »Zurückhaltung« der audiovisuellen Inszenierung eine klare Perspektivierung erfolgt. Die »Bühne« gehört eindeutig Zimmermann, der die gesamte Sequenz hindurch zentral im Bildraum steht. Seine Gesten sind gut zu sehen und können als rhythmische und pragmatische Komponenten der Ausdrucksbewegung zur Geltung kommen (anders etwa bei Redner*innen, die aus der Totalen gezeigt werden). Im Gegensatz dazu bleiben Zwischenrufe und Applaus aus dem Publikum rein akustisch und erscheinen damit eher zurückgesetzt. Sie sind zentraler Teil der Ausdrucksbewegung, auch wenn sie nicht dialogisch inszeniert sind, etwa durch Schuss-Gegenschuss-Montagen. Der affektive Höhepunkt der Ausdrucksbewegung ist deutlich geprägt durch die performative Sprechhandlung des Anklagens und Vorworfens.

Am Beispiel der parlamentarischen Bundestagsrede von Jens Zimmermann (SPD) haben wir illustriert, wie ein medienästhetischer Ansatz für die Analyse audiovisueller und sprachlicher Multimodalität methodisch umgesetzt werden kann. Dabei wurde zunächst die Aufbereitung und materialbasierte Exploration multimedialer Daten jeweils bezogen auf audiovisuelle

und sprachliche Multimodalität vorgestellt, um dann die Ausdrucksbewegung als eine multidimensionale Erfahrungsgestalt als methodische Größe einzuführen. Die Analyse der multimodalen Orchestrierung der Ausdrucksbewegung hat gezeigt, wie die phänomenologisch-deskriptive Identifizierung von affektiven Höhepunkten, eingebettet in Momente hoher Affektdynamik, über die Beschreibung ihrer Bewegungsfiguren empirisch rekonstruiert werden kann. Die Analyse der multimodalen Orchestrierung von Ausdrucksbewegung macht deutlich, dass sich die Ausdrucksbewegung als multidimensionale Gestalt dynamisch entfaltet, eine Gestalt, in der die Multimodalität des Sprechens immer mit der audiovisuellen Inszenierung verwoben ist. Es zeigt sich, dass Affektivität in Form von Bewegungsqualitäten und multimodalen Dynamiken nicht nur unmittelbar erfahrbar wird, sondern dass das gehörte, gesehene Sprechen immer in diese körperliche Erfahrung eingebettet ist.

Unser Ausgangspunkt ist damit die Affektivität der multidimensionalen Gestalt, nicht die sprachlich-gestische Gestaltung der Rede. Affektivität als Erfahrung grundiert den ›Inhalt‹ politischer Rede. Selbst eine zurückhaltende Inszenierung ist immer noch eine Rahmung und Positionierung des Redners als zentrale Figur. Dieser medienästhetische Ansatz, der die Wahrnehmung von Bewegungsbildern als Ausgangspunkt für die Analyse nimmt, hat die Verschränkung von sprachlicher und audiovisueller Multimodalität dokumentiert. Konsequent ist der Ausgangspunkt also nicht die Sprachanalyse, sondern die Analyse der affektiven Modulation der Rede als leibliche Erfahrung eines sich in der Zeit entfaltenden audiovisueller Bewegungsbildes.

Was die Zuschauer berührt ist weniger der Inhalt des Gesagten als das *Wie*. Diesem *Wie* sind wir hier auf verschiedenen Ebenen nachgegangen. Im Fall der Rede von Jens Zimmermann standen sprachlich-gestische und interaktive Aspekte im Vordergrund. Für andere Reden können dies andere Faktoren sein, die die Dynamik besonders triggern, auch kann die audiovisuelle Gestaltung stärker daran beteiligt sein, indem sie z.B. in Momenten hoher affektiver Expressivität eines Sprechers von der Halbnahen auf die Totale wechselt und damit die Sichtbarkeit des Körperausdrucks bis ins Mikroskopische verkleinert.

4 Abschließende Überlegungen und Fazit

Der vorliegende Beitrag hat einen medienästhetischen Ansatz zur Analyse der Multimodalität des Sprechens in audiovisuellen Medien vorgestellt, der einen

kulturwissenschaftlichen Zugang zu verschiedensten Phänomenen sprachlicher und audiovisueller Multimodalität eröffnet. Dabei haben wir argumentiert, dass die empirische Analyse multimodaler Sprachdaten, die in audiovisuellen Medien erscheinen, eine Reflexion und methodologische Berücksichtigung ihrer spezifischen medialen Ökologien notwendig macht.

Der transdisziplinäre Ansatz verbindet deshalb eine filmtheoretische Analyse der audiovisuell generierten Affektdynamiken mit einer linguistischen Analyse ihrer multimodalen Orchestrierung (Gesprächsanalyse, Gestenforschung, Interaktionsforschung). Dabei wurde deutlich, dass sich beide Perspektiven wechselseitig stützen. Ziel war es dabei, die Position der Zuschauer*innen zu rekonstruieren, die das Sprechen in audiovisuellen Medien als multidimensionale Bewegungs- und Zeitgestalten wahrnehmen – und nicht als isolierte Sätze, die aneinandergereiht, hier und da gestisch ausgeschmückt, oder von Wechseln der Kameraperspektive »abwechslungsreicher« gestaltet werden.

Indem wir von der körperlich erfahrbaren Entfaltung affektiver multidimensionaler Erfahrungsgestalten, sogenannten Ausdrucksbewegungen, ausgehen und nicht von einem spezifischen linguistischen Phänomen, nehmen wir einen wesentlichen Perspektivwechsel vor. Ein solch transdisziplinärer Ansatz deckt nicht nur die Komplexität und Vielschichtigkeit multimodalen Sprachgebrauchs auf, er ermöglicht auch eine Objektivierung der Wahrnehmungsdimension. Was wir in unserer Analyse beschrieben haben, sind keine subjektiven, individuellen Gefühle oder inneren Haltungen. Vielmehr handelt sich um ein geteiltes »Zuschauergefühl« (Kappelhoff 2018: 144–146; Kappelhoff/Bakels 2011), das sich in der Verschränkung mit den audiovisuellen Bewegungsbildern entfaltet. Die detaillierte Rekonstruktion der multimodalen Orchestrierung dieser Bilder kann gewissermaßen als »empirischer Nachweis« der »gefühlten« audiovisuellen Ausdrucksbewegung angesehen werden. In ihr fügen sich rhythmische, semantische und pragmatische Bedeutung in einem komplexen Zusammenspiel von Gesten, sprachlicher Äußerung und medialer Ökologie zu einer multidimensionalen Wahrnehmungsgestalt.

Wir schließen mit der Frage, was die Analyse vor einem kulturwissenschaftlichen Hintergrund zeigen kann. Dazu betrachten wir die Rede in ihrer Einbettung in mediale Ökologien als ein frei zugängliches Videodokument des offiziellen Archivs des Bundestages. Wir argumentieren, dass die Öffnung des Kontexts heraus aus der filmwissenschaftlich-linguistischen Betrachtung der medialen Ökologien hinein in gesellschaftliche Ökologien spezifische Formen einer Generierung, Modulation und Mobilisierung von Affekten aufdeckt,

die wie eine übergreifende Ebene der Bedeutung die Analysen rahmt und gesellschaftlich kontextualisiert. Wir beziehen uns hier auf Kappelhoff (2016; 2018). Er führt aus, dass sich im Moment des Sehens ein Wahrnehmungsraum entfaltet, in dem »die ästhetischen, semiotischen und pragmatischen Koordinaten tradierter Sinnkonstruktionen neu figuriert werden« (Kappelhoff 2018: 13f.). Dieser Wahrnehmungsraum ist immer eingebettet in ein Gefüge von sozial, kulturell und historisch situierten, gesellschaftlichen Ökologien (vgl. Kappelhoff 2018: 14, in Bezug auf Bourdieus Soziologie der legitimen Kultur). Die Inszenierung ist Teil eines etablierten »Medien-Konsums«, welcher wiederum das Sehen der audiovisuellen Daten formt. In diesem Konsum werden die Bewegtbilder zu einer »kulturellen Praxis« (Kappelhoff 2018: 10). Im Rahmen der im Archiv des Bundestages als audiovisuelles Dokument von Parlamentsdebatten der Öffentlichkeit zugänglich gemachten Rede von Zimmermann ist diese Wirklichkeit eine Form des geteilten Verständnisses von demokratischem Geschehen, in dem Streit und Auseinandersetzung akzeptierter und wichtiger Bestandteil politischen Handelns sind, welche jedoch in einem kontrollierten Rahmen verlaufen. Diese Kontrolliertheit zeigt sich in der Perspektivierung und der Affektmodulation. Wem wird wann die Bühne gegeben? Wie werden Streit und Auseinandersetzung inszeniert? Das Beispiel von Zimmermann verdeutlicht, dass selbst in Momenten direkten »Angriffs« aus dem Publikum, welche als Gefahr der primären Sprecherrolle angesehen werden können, kein Kontrollverlust dieser Rolle erfolgt. Die mediale Ökologie moduliert eine vermeintliche Neutralität, die den aktuellen Sprecher stützt. Es entfaltet sich ein Gefühl von Ordnung, Struktur und Kontrolliertheit. Die Kamera hält Zimmermann zentral, sie gibt ihm Raum. Die Auseinandersetzung wird nicht als ein gesichtsbedrohender Akt Zimmermanns erfahren. Es werden keine hämischen Gesichter in Großaufnahmen gezeigt, keine Schenkellklöpfer oder Klamauk in Szene gesetzt. Die audiovisuelle Form der Rede reiht sich so in den Kanon des Archivs des Bundestages ein, in dem eine Form von demokratischem Handeln und Geschehen erfahren wird.

Mit einer medienästhetischen Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien verschiebt sich der Gegenstand sprachwissenschaftlicher Analyse von einer kontextfreien Sprachbetrachtung zu einer kontextreichen Sprachanalyse, die Sprache als sozio-kulturelles Phänomen in ihren medienkulturellen Ökologien und damit in einer kultursissenschaftlichen Perspektive betrachtet.

Literaturangaben

- Bakels, Jan-Hendrik (2017): *Audiovisuelle Rhythmen. Filmmusik, Bewegungskomposition und die dynamische Affizierung des Zuschauers*, Berlin/Boston: De Gruyter.
- Bakels, Jan-Hendrik/Thomas Scherer/Jasper Stratil/Henning Agt-Rickauer (2020a): AdA Filmontology – a machine-readable Film Analysis Vocabulary for Video Annotation. In: Estill, Laura/Jennifer Giuliano (Hg.): *15th Annual International Conference of the Alliance of Digital Humanities Organizations*, DH 2020, Ottawa, Canada, July 20–25, 2020, Conference Abstracts.
- Bakels, Jan-Hendrik/Matthias Grotkopp/Thomas Scherer/Jasper Stratil (2020b): Computergestützte Filmanalyse im Spannungsfeld von Datenmodellen und Gestalttheorie. In: *montage AV* 29:1, 99–118.
- Bressem, Jana/Cornelia Müller (2014): A repertoire of German recurrent gestures with pragmatic functions. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva Ladewig/David McNeill/Sedinha Tessendorf (Hg.): *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/New York: De Gruyter Mouton, 1575–1591.
- Bressem, Jana/Silva H. Ladewig/Cornelia Müller (2013): Linguistic annotation system for gestures (LASG). In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Sedinha Teßendorf (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/New York: De Gruyter Mouton, 1098–1124.
- Bühler, Karl (1933): *Ausdruckstheorie. Das System an der Geschichte aufgezeigt*, Jena: Verlag von Gustav Fischer.
- Deppermann, Arnulf (2008): *Gespräche analysieren. Eine Einführung*, Wiesbaden: Verlag für Sozialwissenschaften.
- Goodwin, Marjorie/Asta Cekaite/Charles Goodwin (2012): Emotion as Stance. In: Sorjonen, Marja-Leena/Anssi Perakyla (Hg.), *Emotion in Interaction*, Oxford: Oxford University Press, 16–41.
- Greifenstein, Sarah (2020): *Tempi der Bewegung – Modi des Gefühls. Expressivität, heitere Affekte und die Screwball Comedy*, Berlin/Boston: De Gruyter.
- Greifenstein, Sarah/Dorothea Horst/Thomas Scherer/Christina Schmitt/Hermann Kappelhoff/Cornelia Müller (Hg.) (2018): *Cinematic Metaphor in Perspective. Reflections on a Transdisciplinary Framework*, Berlin/Boston: De Gruyter Mouton.
- Hickethier, Knut (2012): *Film- und Fernsehanalyse*, 5. aktualisierte und erweiterte Aufl., Stuttgart: J.B. Metzler.

- Horst, Dorothea (2018): *Meaning Making and Political Campaign Advertising. A Cognitive-Linguistic and Film-Analytical Perspective on Audiovisual Figurativity*, Berlin/Boston: De Gruyter.
- Horst, Dorothea/Franziska Boll/Christina Schmitt/Cornelia Müller (2014): Gesture as interactive expressive movement: Inter-Affectivity in face-to-face-communication. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressem (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/Boston: De Gruyter Mouton, 2112–2125.
- Kappelhoff, Hermann (2004a): Unerreichbar, unberührbar, zu spät. Das Gesicht als kinematografische Form der Erfahrung. In: *montage/av* 13:2, 29–53.
- Kappelhoff, Hermann (2004b): *Matrix der Gefühle. Das Kino, das Melodrama und das Theater der Empfindsamkeit*, Berlin: Vorwerk 8.
- Kappelhoff, Hermann (2008): Zuschauergefühl. Die Inszenierung der Empfindung im dunklen Raum des Kinos. In: Spiekermann, Geraldine/Beate Söntgen (Hg.), *Tränen*, München: Fink, 195–206.
- Kappelhoff, Hermann (2016): *Genre und Gemeinsinn. Hollywood zwischen Krieg und Demokratie*, Berlin/Boston: De Gruyter.
- Kappelhoff, Hermann (2018): *Kognition und Reflexion. Zur Theorie filmischen Denkens*, Berlin/Boston: De Gruyter.
- Kappelhoff, Hermann/Cornelia Müller (2011): Embodied meaning construction. Multimodal metaphor and expressive movement in speech, gesture, and feature film. In: *Metaphor and the Social World* 1:2, 121–153.
- Kappelhoff, Hermann/Jan-Hendrik Bakels (2011): Das Zuschauergefühl. Möglichkeiten qualitativer Medienanalyse. In: *Zeitschrift für Medienwissenschaft* 5:2, 78–96.
- Kendon, Adam (2004): *Gesture: visible action as utterance*, Cambridge/New York: Cambridge University Press.
- Ladewig, Silva H. (2014): The cyclic gesture. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressem (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Interaction*, Berlin: De Gruyter Mouton, 1605–1618.
- Löffler, Petra/Florian Sprenger (2016): Medienökologien. Einleitung in den Schwerpunkt. In: *Zeitschrift für Medienwissenschaft* 14:1, 10–18.
- Merleau-Ponty, Maurice (1968): *The Visible and the Invisible*, Evanston, IL: Northwestern University Press.

- Michel, Sascha/Daniel Pfurtscheller (2021): »Ich bin seit Montag Zuhause in Quarantäne« – Zur Verbindung von Erzählen und Argumentieren in Social-Media-Kommentaren zu politischen Reden. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 51:2, 255–279.
- Müller, Cornelia (1998): *Redebegleitende Gesten: Kulturgeschichte – Theorie – Sprachvergleich*, Berlin: Berlin Verlag Arno Spitz.
- Müller, Cornelia (2010): Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. In: *Sprache und Gestik* 1:41, 37–68.
- Müller, Cornelia (2014): Ring-gestures across cultures and times: Dimensions of variation. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressem (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Interaction* 2, Berlin: De Gruyter Mouton, 1511–1522.
- Müller, Cornelia (2019): Metaphorizing as embodied interactivity: What gesturing and film viewing can tell us about an ecological view on metaphor. In: *Metaphor and Symbol* 34, 61–79.
- Müller, Cornelia (2024a): Mimesis as as-if action. In: Zywickzynski, Przemysław/Slawomir Wacewicz/Monika Boruta-Żywiczyńska/Johan Blomberg (Hg.), *Perspectives on pantomime: evolution, development, interaction*, Amsterdam/Philadelphia: John Benjamins, 220–243.
- Müller, Cornelia (2024b): A toolbox for methods of gesture analysis. In: Cienki, Alan (Hg.), *The Cambridge Handbook of Gesture Studies*, Cambridge: Cambridge University Press, 182–216.
- Müller, Cornelia/Christina Schmitt (2015): Audio-visual metaphors of the financial crisis: meaning making and the flow of experience. In: *Revista Brasileira de Linguística Aplicada* 15:2, 311–342.
- Müller, Cornelia/Hermann Kappelhoff (2018): *Cinematic Metaphor. Experience – Affectivity – Temporality*, Berlin/Boston: De Gruyter Mouton.
- Müller, Cornelia/Susanne Tag (2010): The Dynamics of metaphor: Foregrounding and activating metaphoricity in conversational interaction. In: *Cognitive Semiotics* 6, 85–120.
- Neumann, Ranghild (2004): The semantics and pragmatics of everyday gestures. Proceedings of the Berlin conference, April 1998. In: Müller, Cornelia/Roland Posner (Hg.), *The conventionalization of the ring gesture in German discourse*, Berlin: Weidler, 217–223.
- Opiłowski, Roman (2017): Forschungsfelder der Medienlinguistik und neue Herausforderungen für Medienlinguistik 3.0. In: Iwona Bartoszewicz/

- Joanna Szczę/Artur Tworek (Hg.), *Grenzen der Sprache – Grenzen der Sprachwissenschaft I*, Wrocław/Dresden: Neisse, 179–190.
- Scherer, Thomas J. J. (2024): *Inszenierungen zeitgenössischer Propaganda. Kampagnenfilme im Dienste des Gemeinwohls*, Berlin/Boston: De Gruyter.
- Scherer, Thomas/Sarah Greifenstein/Hermann Kappelhoff (2014): Expressive Movements in Audiovisual Media. Modulating Affective Experience. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressem (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/Boston: De Gruyter Mouton, 2081–2092.
- Schmitt, Christina (2020): *Wahrnehmen, fühlen, verstehen, Metaphorisieren und audiovisuelle Bilder*, Berlin/Boston: De Gruyter.
- Schmitt, Christina/Sarah Greifenstein/Hermann Kappelhoff (2014): Expressive Movement and Metaphoric Meaning Making in Audio-Visual Media. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressem (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/Boston: De Gruyter Mouton, 2092–2112.
- Selting, Margret/Peter Auer/Dagmar Barth-Weingarten/Jörg Bergmann/Karin Birkner/Elizabeth Couper-Kuhlen/Arnulf Deppermann/Peter Gilles/Susanne Günthner/Martin Hartung/Friederike Kern/Christine Mertzlufft/Miriam Morek/Frank Oberzaucher/Jörg Peters/Uta Quasthoff/Wilfried Schütte/Anja Stukenbrock/Susanne Uhmann (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung* 10, 353–410.
- Sobchack, Vivian (1992): *The Address of the Eye. A Phenomenology of Film Experience*, Princeton: Princeton University Press.
- Stern, Daniel N. (2011): *Formen der Vitalität. Die Erforschung dynamischen Erlebens in Psychotherapie, Entwicklungspsychologie und den Künsten*, Frankfurt a.M.: Brandes & Apsel.
- Stöckl, Hartmut (2012): Medienlinguistik. Zu Status und Methodik eines (noch) emergenten Forschungsfeldes. In: Held, Gudrun/Christian Grösslinger/Hartmut Stöckl (Hg.), *Pressetextsorten jenseits der »News«. Medienlinguistische Perspektiven auf journalistische Kreativität*, Frankfurt a.M.: Suhrkamp, 13–34.
- Xia, Qiang (2023): Turn-taking in video-mediated and co-present dialogues. A corpus-based study of German. In: *Register Aspects of Language in Situation (REALIS)* 2:4, 1–29.

Clara Kindler-Mathôt, M.A.
kindler@europa-uni.de

Cornelia Müller, Prof. Dr.
cmueller@europa-uni.de

Jana Junge, M.A.
junge@europa-uni.de

Katerina Papadopoulou, M.A.
Europa-Universität Viadrina
Große Scharrnstrasse 59
15230 Frankfurt (Oder)
papadopoulou@europa-uni.de