

Erfahrungsraum Interaktion

Die audiovisuelle Modulation eines Gesprächs als Zuschauer*innenerfahrung

Clara Kindler-Mathôt

Abstract *Der medienlinguistische Beitrag nimmt aus einer Zuschauer*innenperspektive die komplexe Bedeutungsentfaltung audiovisueller Bilder in den Blick. Am Beispiel eines Videobeitrags des digitalen Parteitags der Partei BÜNDNIS 90/DIE GRÜNEN wird gezeigt, wie audiovisuelle und sprachlich-interaktive Formen gemeinsam ein spezifisches Bewegungsbild entfalten, welches die Zuschauer*innenerfahrung eines ›Gesprächs‹ moduliert, das nicht das Abbild vorfilmischer Realität, sondern eine neue mediale Wirklichkeit darstellt. Der theoretische und methodische Ankerpunkt des medienästhetischen Ansatzes liegt im Konzept der filmischen Ausdrucksbewegung, welcher mit Ansätzen der Interaktions- und Gesprächsanalyse ergänzt wird. Die Diskussion der Analyse betont die Unmittelbarkeit der Zuschauer*innenerfahrung und plädiert vor dem Hintergrund zeitgenössischen Medienkonsums für eine Stärkung der Medienkompetenz zur Reflexion audiovisueller Bedeutungskonstitution.*

Keywords *Zuschauer*innenerfahrung; mediale Wirklichkeit; Medienästhetik; sprachliche und audiovisuelle Multimodalität; mediale Durchformung*

1. Modulation medialer Wirklichkeiten als Zuschauer*innenerfahrung

Die Gegenwart bringt zunehmend neue Formen der audiovisuellen Kommunikation und Interaktion hervor, seien es DIY-Tutorials auf Instagram, Live-Übertragungen digitaler Parteitage oder politisch aufgeladene Kommentare auf YouTube. Insbesondere vor der Frage um die »Wirkung« audiovisueller Bilder wird die Rolle des Medienkonsums und dessen gesellschaftliche Folgen relevant. Ein Aspekt davon bezieht sich auf die Wahrhaftigkeit oder Echtheit audiovisueller Bilder, bzw. ob das, was wir sehen, real ist oder nicht – oder zumindest als solches beurteilt wird. Der Beitrag nimmt diese Frage aus einer medienlinguistischen Perspektive auf, die Ansätze der sprachwissenschaftlichen Multimodalitätsforschung mit filmtheoretischen und -analytischen Perspektiven vereint. Am Beispiel eines Videobeitrags der

digitalen Bundesdelegiertenkonferenz (DBDK) 2020 der Partei BÜNDNIS 90/DIE GRÜNEN wird gezeigt, dass auch die bewegten Bilder einer vordergründig ›beobachtenden‹ und vergleichsweise ›objektiven‹ Kamera nicht einfach Abbildung der vorfilmischen Realität sind, sondern jede Aufnahme neue *mediale Wirklichkeiten* (vgl. Luginbühl 2019) erzeugt.

Diese bedeutungskonstituierende Funktion von Medialität wird in Arbeiten der sprachlichen Multimodalitätsforschung (v.a. multimodaler Interaktionsanalyse) in der Regel nicht weiter reflektiert, denn sie sind häufig der Versuch, eine emische Perspektive auf spezifische Interaktionsformen einzunehmen. Der Artikel nimmt deshalb einen Perspektivwechsel vor, indem nicht der Bedeutungsaushandlungsprozess zwischen Personen im Bild, sondern zwischen Bild und Zuschauer*innen adressiert wird. In dieser Zuschauer*innenerfahrung (vgl. Kappelhoff 2008; Kappelhoff/Bakels 2011) verschmelzen die Multimodalität des Sprechens und die audiovisuelle Multimodalität zu einer komplexen *multidimensionalen Bedeutungsgestalt* (Müller/Kappelhoff 2018: Kap. 2.3), die im Moment des Sehens unmittelbar erfahren wird. Die spezifischen Bewegungsbilder dieser Bedeutungsgestalten formen den Erfahrungsraum der Zuschauer*innen.

Der theoretische und methodische Ankerpunkt liegt dabei im medienästhetischen Konzept der filmischen Ausdrucksbewegung (vgl. Kappelhoff 2004: Teil 1; Kappelhoff/Müller 2011; Müller/Kappelhoff 2018: Kap. 9; Scherer et al. 2014). Mithilfe dieses Ansatzes zeigt die Fallstudie am Beispiel eines ›Gesprächs‹, dass dieses nicht einfach repräsentiert wird, sondern erst durch die audiovisuelle Inszenierung als ein Bewegungsbild in der Zuschauer*innenerfahrung moduliert wird. Kamera und Schnitt setzen die Personen in ein raum-zeitliches Verhältnis, welches als Interaktion von Angesicht zu Angesicht (i.F. von AzA) – als Gespräch – erfahren wird. Diese Zuschauer*innenerfahrung ist so unmittelbar, dass sie erst in Momenten der Irritation hinterfragbar wird, da sie bspw. als Bruch in den kommunikativen Praktiken erfahren wird. Vor dem Hintergrund zeitgenössischer Formen des Medienkonsums, plädiert der Beitrag deshalb für eine Stärkung der Medienkompetenz der Zuschauer*innen, um die spezifische Medialität des Audiovisuellen reflektieren zu können.

2. Eine Frage der Perspektive: Audiovisuelle Medialität und Sprachgebrauch von Innen und von Außen

»Kommunikation [ist] immer auf Medialität angewiesen« (Luginbühl 2019: 126), sei sie schriftlich, mündlich oder gebärdet, von AzA, am Telefon oder im Fernsehen. Jede Medialität unterliegt dabei spezifischen Bedingungen, die die Kommunikation nicht nur übertragen, sondern *ausformen*, genauer, *mitkonstituieren* (vgl. ebd.). Eine Änderung der *medialen Bedingungen* (vgl. Marx/Schmidt 2019: 14ff.), hat immer

Auswirkungen auf die semiotischen Ressourcen und somit auf die Form der Bedeutungsaushandlung. Bei einem Telefonat etwa fallen alle visuellen Ressourcen weg, in einer Videokonferenz ist nur sichtbar, was in den Kameraausschnitt fällt. Dementsprechend ist auch eine Anpassung der methodischen Herangehensweisen erforderlich, was beispielsweise Ausdruck in der wachsenden Multimodalitätsforschung und Medienlinguistik findet.

Im Folgenden werden zwei Perspektiven auf Medialität und ihre »vermittelnde« Funktion eröffnet. Die erste bezieht sich auf eine »Innenperspektive«: Die interaktionsanalytischen Zugänge zu *medienvermittelter Kommunikation* thematisieren den Aushandlungsprozess zwischen den Interaktionsteilnehmer*innen. Medialität übernimmt dabei die Funktion eines Übertragers und Vermittlers. Mit dem daran anschließenden Konzept der *medialen Durchformung* wird eine Brücke zur zweiten Perspektive geschlagen, die das Medialitätsverständnis erweitert und die Rolle der Betrachter*innen audiovisueller Bilder mit einbezieht. Diese Außenperspektive wird im dritten Abschnitt durch den filmwissenschaftlichen Ansatz des *Zuschauergefühls* (Kappelhoff 2008, Bakels/Kappelhoff 2011) ausgeführt, welcher den Ausgangspunkt für die sich anschließende Analyse bildet.

2.1 Innenperspektive: Ansätze der medienvermittelten Kommunikation

In der interaktionslinguistischen Forschung zu *medienvermittelter Kommunikation* werden nicht nur Fragen nach dem Übertragungsprozess von Kommunikation mittels (meist als technische Hilfsmittel verstandener) Medien (vgl. Marx/Schmidt 2019: 9) gestellt, sondern auch nach den »spezifischen Folgen medialer Vermittlung für Interaktion und Interaktionsforschung« (Marx/Schmidt 2019: 2). Damit wird eine emische Innenperspektive eingenommen: Durch die ethnomethodologisch-konversationsanalytische Forschungslogik (vgl. ebd.: 14), die in der »möglichst guten Annäherung an die Teilnehmerperspektive der Interagierenden« (Lanwer 2019b: 5) besteht, werden für die Interagierenden sicht- und hörbare Praktiken in Bezug auf reflexive Aushandlungsprozesse analysiert. Dabei wird vor allem in Betracht genommen, wie (und ob) die vermittelnden Medien die Kernaspekte von Interaktion beeinflussen – im weitesten Sinne stellen dies Formen der Kopräsenz und Wechselseitigkeit dar (vgl. Marx/Schmidt 2019: 2) – und wie die Interagierenden damit umgehen, insbesondere in Bezug auf die sich medial verändernden semiotischen Ressourcen:

Da Medienvermittlung das gemeinsame Hier und Jetzt (je nach Medium) partiell auflöst bzw. in umgekehrter Perspektive distante und in größerem Umfang verbreitete Kommunikation überhaupt erst ermöglicht, stehen viele dieser Ressourcen und auf ihnen basierende Lösungen nicht mehr oder zumindest so nicht mehr zur Verfügung. Stattdessen kommen neue Ressourcen (etwa im Falle des Telefons

das Telefonklingeln) und entsprechende Lösungen hinzu, etwa spezifische Eröffnungssequenzen. (Marx/Schmidt 2019: 2)

Die Innenperspektive beschreibt folglich, wie die Interaktionsteilnehmer*innen in ihren gegenwärtigen Realitäten, in ihrem Hier und Jetzt agieren, wie sie wechselseitig aufeinander bezogen handeln, wie sie sich in den spezifischen Interaktionsräumen, seien sie virtuell oder real, mündlich oder schriftlich, orientieren und mit den spezifischen medialen Umwelten umgehen. Was bedeutet eine solche Perspektive für Interaktion in audiovisuellen Kontexten? Aus Perspektive der medienvermittelten Kommunikation steht auch hier in erster Linie die Bedeutungsaushandlung zwischen den Teilnehmer*innen im Blick. Im Falle von Videokonferenzen¹ etwa zeigen Studien (vgl. Haddington/Oittinen 2022; Lanwer 2019a; 2019b; Xia 2023), dass die medialen Bedingungen Einfluss auf die gegenseitige Koordinierung der Interagierenden nehmen. Die korpusbasierte Vergleichsstudie zwischen Gesprächen von AzA und Zoomkonferenzen von Xia (ebd.) zeigt so bspw. ein deutlich verändertes Timing bei den Turn-Übernahmen auf: Gespräche in der Videokonferenz neigen so zwar zu weniger Überlappungen, dafür sind diese, ebenso wie die Pausen zwischen den Turns, länger. Ebenfalls in Studien thematisiert werden die gemeinsame Blickkoordination sowie Aufmerksamkeitsfokussierung (*joint attention*; Lanwer 2019a; 2019b), die durch die technischen Affordanzen wie Kamera, Softwareinterface oder Internetverbindung beeinflusst werden.

Während die Forschung zu Videokonferenzen trotz Berücksichtigung dieser Affordanzen eine emische Perspektive der vorfilmischen Realität der Teilnehmer*innen einnimmt, öffnet die Forschung zu Fernsehgesprächen zusätzlich die Frage nach den Zuschauer*innen und wie für sie Bedeutung konstituiert wird. Da damit Fragen der Rezeption angesprochen werden, deutet sich an dieser Stelle eine Form der Außenperspektive an. Spätestens hier wird es relevant, zwischen zwei Formen der Realität zu unterscheiden: Zum einen die vorfilmische Realität, in der Interaktionsteilnehmer*innen bspw. in einem Studio sitzen und sich in einer Interaktionssituation von AzA miteinander befinden. Zum anderen die Realität die sich für die Zuschauer*innen als audiovisuelle Bewegungsbilder auf ihren Bildschirmen entfaltet. Diese ist geformt und moduliert durch die audiovisuelle Medialität: Kameraeinstellungen und Bewegungen bestimmen die für die Zuschauer*innen sichtbaren Bildräume, die Position der Mikrofone entscheidet, wer oder was zu hören ist, Schnitte und Montagesequenzen ermöglichen es, zusätzliche Bilder einzublenden und eine Form der Zeitlichkeit zu kreieren, die nicht der Zeitlichkeit der vorfilmischen Realität entsprechen muss. Luginbühl (2019; Luginbühl/Schneider

1 Mit Videokonferenzen sind sogenannte *Face-to-Screen-Interaktionen* gemeint (vgl. Lanwer 2019a; 2019b) und weniger Konferenzen, die in Besprechungsräumen mit »veränderter Ökologie« (Meier 2000: 195) stattfinden.

2020) bezeichnet deshalb diese Zuschauer*innenrealität als neue, *medial durchformte Wirklichkeit*. Dieser Ansatz bildet damit eine Brücke zwischen der Innen- und der Außenperspektive, da er beide Ebenen der Interaktion berücksichtigt.

2.2 Brücke zwischen Innen und Außen: Mediale Durchformung

Der Ansatz der medialen Durchformung beschreibt den Einfluss medialer Affordanzen auf die Zuschauer*innenerfahrung. Für Fernsehgespräche (vgl. Luginbühl 2019) bedeutet dies, dass sie nicht abgefilmte »außermediale Gespräche« (ebd.: 143) sind, die einfach übertragen werden, sondern die Affordanzen der spezifischen medialen Ökologien die Bedeutungskonstitution der Zuschauer*innen entscheidend mitbestimmen. Luginbühl und Schneider (2020) illustrieren dies am Beispiel der zweiten Präsidentschaftsdebatte von 2016 in den USA zwischen Hillary Clinton und Donald Trump. Zum einen werden die räumlichen Affordanzen des Studios und die Kameraführung von den Kontrahenten selbst als bedeutsam erfahren und dementsprechend genutzt, etwa in der aktiven Zuwendung zur Kamera. Zum anderen wird ihr Verhalten selbst aus einer Zuschauer*innenperspektive von der Medialität des Audiovisuellen, bspw. der Kameraperspektive, moduliert (vgl. ebd.: 58). Das Gespräch von AzA wird so durch die medialen Umstände des Fernsehens quasi in doppelter Weise durchformt (von Innen und von Außen), woraus eine neue Form der Medialität hervorgeht, in welcher Inhalte nicht nur vermittelt werden, sondern die Medialität selbst als Akteurin der Bedeutungskonstitution hervorgeht:

In this threefold way, the medium of television is thus a political actor: it presents oral interactions that are shaped from the outset by the institution of television and the media format of television discussion. Without the medium of television, these interactions would never take place in this way, as a result of which the medium not merely transmits the content, but becomes an actor in its own right. (ebd.: 61)

Holly (2015) spricht in diesem Zusammenhang auch von *Intermedialität* und *intermedialen Bezugnahmepraktiken*, aus welchen das Zusammenspiel von gesprochener Sprache und Audiovisualität als Prozess gemeinsamer Semiose hervorgeht. Am Beispiel der Polit-Talkshow »Maybritt-Illner« beschreibt er das intermediale Zusammenspiel, in welchem durch Kameraperspektiven und Bewegungen, Einblendung von Bildmaterial o.Ä. Bedeutungsaspekte der Interaktion verändert oder hinzugefügt werden:

Die Kameraführung in Polit-Talkshows generiert durch die Selektion von Einstellungen und Umschnitten Bedeutungskomponenten, die sprachliche Äußerungen

»transkribieren«, d.h. überformen, implizit kommentieren und dadurch »anders lesbar« machen. (ebd.: 89)

Die Ansätze dieser sogenannten *Transkription* und die Generierung zusätzlicher Bedeutungskomponenten sind den Zuschauer*innen im Moment des Sehens nicht offenkundig. In der filmischen Erfahrung sind sprachliche Äußerungen und die Medialität des Audiovisuellen nicht mehr trennbar, sondern sind in ihrer ganzheitlichen Gestalt auch bedeutungskonstituierend. In dieser Ganzheitlichkeit verschwindet die Medialität und hinterlässt nur unmerkliche Spuren. Luginbühl und Schneider (vgl. ebd.: 59) beschreiben dies, mit Verweis auf Sybille Krämer (1998), als Tendenz der Medien sich unsichtbar zu machen. Diese Unsichtbarkeit wird jedoch in Momenten der Irritation aufgeweicht: »Für wirklich hält der Filmbetrachter vor allem das, was den Eindruck von Wirklichkeit macht, für wenig eindrucksvoll wird dagegen gehalten, was diesem Illusionscharakter des Mediums nicht entspricht, sondern ihn irritiert« (Koch 2003: 223). Die mediale Wirklichkeit erscheint dann als nicht mehr kohärent, und wird hinterfragbar. Beispiele stellen Formen der Asynchronität oder Inkongruenz zwischen Bild, Ton, Bildraum und/oder Kameraführung dar, bedingt durch technische Störungen oder der Inszenierung selbst. In der unten vorgestellten Analyse (s. 3.3) zeigt sich dies exemplarisch in der ungewöhnlichen Turnkonstruktion, moduliert durch die Montage.

Während der Ansatz der medialen Durchformung grundsätzlich den Blick auf das Fernsehpublikum öffnet, verbleibt er jedoch auf Ebene, zwar durchformter aber dennoch repräsentierter Inhalte. Obwohl Luginbühl (2019) betont, dass die Bedeutungskonstitution nur aus dem Gebrauch, also aus dem Rezeptionsprozess hervorgeht, bleibt eine Reflexion dieses Prozesses aus. Ich möchte daher mit einem Blick in die medienästhetische Filmtheorie noch einen Schritt zurückgehen, indem gefragt wird, wie es überhaupt zur Konstitution von Bedeutung bei Zuschauer*innen kommt, wie sie also grundsätzlich audiovisuelle Bilder verstehen. Damit wird die wesentliche Zuschauer*innenerfahrung filmischer Bilder adressiert.

2.3 Außenperspektive: Das interaktive Moment der Zuschauer*innenerfahrung

Nach wie vor sind in der Filmwissenschaft kognitive Ansätze geläufig, in welchen filmische Bedeutung als Produkt der »Zuschaueraktivität im Rahmen eines Verstehens, Beurteilens oder Bewertens von narrativen, repräsentierten Handlungen, Szenarien und Figurenkonstellationen« (Greifenstein 2020: 24) verstanden wird.² Dabei wird das Narrativ und die Repräsentationsebene als gegeben, als offenkundig angesehen und nicht als Produkt der Rezeption von sich zeitlich entfaltenden und

2 Für eine Diskussion des Diskurses vgl. Müller/Kappelhoff (2018: Kap. 1).

angeordneten Bewegungsbildern. Müller und Kappelhoff (2018) kritisieren dies wie folgt:

The claim that basic meanings are understood, as a rule lacks a convincing model of how this understanding is accomplished. Instead, what is to be understood from film images is taken for granted, as already given: actors, actions, spaces, objects, treated as a concatenation of isolated pictorial representations. (ebd.: 30)

Der »specific media-character of audiovisual images« (ebd.: 8) wird außer Acht gelassen und damit auch der Prozess des Sehens und Verstehens audiovisueller Bilder: Sie entfalten sich im Moment des Sehens als eine zeitliche Erfahrung, welche aus dem Zusammenspiel der Inszenierungsformen wie Montagetechnik, Kameraeinstellungen usw., als orchestrierte Bewegungsfigurationen hervorgehen – als sogenannte *Bewegungsbilder* (vgl. ebd.: 23). Kappelhoff (2018) bezeichnet dies als die Poiesis des Filme-Sehens, als kreativen Prozess des Verstehens, als Entstehen von Film im Moment der Rezeption (vgl. Müller/Kappelhoff 2018: 223; Scherer 2024: 110ff.). Die Rezeption ist hier kein Dekodieren, sondern ein körperlich-sinnliches Wahrnehmen der Bewegungsbilder, die uns in ihren Rhythmen und Dynamiken direkt und unmittelbar *bewegen* (vgl. Scherer et al. 2014: 2083). Ihre Bewegungsmuster (*movement patterns*; vgl. ebd.) formen die alles umspannende Ebene der Zuschauer*innenerfahrung, indem sie den Wahrnehmungsvorgang, die kognitiven und perzeptiven Prozesse organisieren und strukturieren und so bedeutsam werden (vgl. Kappelhoff 2008: 206). Das sich so entfaltende *Zuschauergefühl* (Kappelhoff 2004; 2008; Kappelhoff/Bakels 2011; Bakels 2017) ist deshalb eine sinnlich-körperliche Erfahrung, eine Modellierung von sich entfaltenden Affekten, ein Mitfühlen (vgl. Bakels 2017: 67), eine grundlegende Form der Affizierung der Zuschauer*innen. Vor diesem Hintergrund sind die Zuschauer*innen nicht reine Empfänger*innen, sondern stehen in einer Art interaktivem Verhältnis mit den Bewegungsbilder selbst (vgl. Müller/Kappelhoff 2018: 29).

In Kappelhoffs Filmtheorie (2004; 2018) nehmen die Begriffe des *Bewegungsbildes* und der *Ausdrucksbewegung* eine zentrale Stellung ein. Er entwickelt den Begriff der filmischen Ausdrucksbewegung in einer kulturhistorischen Auseinandersetzung mit sprachwissenschaftlichen, soziologischen und psychologischen Ausdruckstheorien der ersten Dekaden des 20. Jahrhunderts (bspw. Bühler 1933; Plessner 1982; Wundt 1975) sowie den filmtheoretischen Schriften von Deleuze (*Zeit- und Bewegungsbild*; vgl. 1989; 1990). Mit der Unmittelbarkeit der Zuschau-Erfahrung und der Körperlichkeit des Filme-Sehens greift Kappelhoff Aspekte der neo-phänome-

nologischen Filmtheorie Vivian Sobchacks (1992) auf.³ Aktuelle Forschung in der Film- und Sprachwissenschaft hat die Konzepte kooperativ weiterentwickelt (Müller/Kappelhoff 2018; Greifenstein 2020; Horst 2018; Scherer 2024; Kindler-Mathôt et al. im Erscheinen) und über unterschiedliche Kontexte hinweg (insbesondere) am Gegenstand multimodaler Metaphorik in Sprache, Geste und audiovisuellen Medien methodisch erprobt und zum Ansatz einer empirischen Medienästhetik weitergedacht.

Dieser medienästhetische Ansatz grenzt sich ebenfalls von der (neuro-)psychologischen experimentellen⁴ sowie kommunikationswissenschaftlichen und medienpsychologischen Rezeptionsforschung⁵ ab, welche sich für Medienwirkung in empirisch-experimentellen Settings interessiert. Ähnlich der kognitiven Ansätze, wird hier in erster Linie Bezug zu Reiz-Reaktions-Relationen genommen, bspw. zur »systematische[n] Erfassung und Protokollierung von sinnlichen und apparativ wahrnehmbaren Aspekten menschlicher Handlungen und Reaktionen« (Gehrau 2017: 17). Damit werden vordergründig Fragen rund um intendierte Wirkungsstrategien thematisiert. Kappelhoffs medienästhetischer Ansatz versucht vielmehr, Grundmuster und Poetologien verschiedener Affizierungsformen aufzudecken und damit das unmittelbare Wahrnehmungserleben und das Filme-Sehen als Praxis und Poiesis zu verdeutlichen.⁶

Die folgende Analyse betrachtet deshalb die Inszenierung aus einer Zuschauer*innenperspektive als *Modulation einer medialen Wirklichkeitserfahrung von Interaktion*, die die Rolle von Produzent*innen und möglichen Intentionen außen vor lässt und vielmehr die audiovisuellen Bewegungsbilder in den Vordergrund stellt. Vor diesem Hintergrund wird nicht von der Repräsentation eines Gesprächs ausgegangen, sondern von einer audiovisuellen Inszenierung und Orchestrierung: Kamera und Montage setzen die Personen, deren Körperbewegungen und räumliche Situierung in ein Verhältnis, welches von den Zuschauer*innen als Interaktion, als Gespräch erfahren wird. Im Bewegungsbild verschmilzt die Multimodalität des Spre-

3 Zur ausführlichen Theoriegeschichte und Begriffskonzeption siehe Müller/Kappelhoff (2018: Kap. 9) sowie Greifenstein (2020: Kap. 2.3).

4 Siehe für eine Diskussion der Forschung bei Kappelhoff und Bakels (2011) und Greifenstein (2020).

5 Zur Unterscheidung und Ausrichtung der verschiedenen Ansätze, vgl. Gehrau (2008). Ebenfalls verweise ich auf die Zeitschriftenreihe »Rezeptionsforschung« (2004- heute) des Nomos Verlags.

6 Die theoretischen Ansätze und methodischen Vorgehensweisen wurden maßgeblich in der Kolleg-Forschungsgruppe »Cinepoetics. Poetologien audiovisueller Bilder« entwickelt. Sie sind in zahlreichen Publikationen in den von De Gruyter publizierten Schriftenreihen (Cinepoetics (2016-heute), Cinepoetics – English Edition (2018-heute), Cinepoetics Essay (2020-heute) sowie das Online-Journal mediaesthetics (2016-heute); alle zu finden unter <https://www.cinepoetics.fu-berlin.de/publications/index.html> (08.07.2024) dokumentiert.

chens in Interaktion mit der audiovisuellen Multimodalität⁷ zu einer sich zeitlich entfaltenden Bedeutungsgestalt – einem Erfahrungsraum. Diese Gestalt, dieser Erfahrungsraum *ist* die mediale Wirklichkeit, die sich im Moment des Sehens im Körper der Zuschauer*innen entfaltet.

3. Audiovisuelle Modulation einer Gesprächserfahrung

3.1 Daten und methodische Vorgehensweise

Im Folgenden analysiere ich einen Videobeitrag (5:28 Min), welcher im Rahmen der DBDK der Partei BÜNDNIS 90/DIE GRÜNEN 2020 ausgestrahlt wurde.⁸ Er gehört zu einer Reihe von Einspielern, die unter dem Titel »Neue Zeiten, neue Antworten« immer zwei Parteimitglieder, hier Claudia Roth und Georg Kurz, im Gespräch zeigen. Das Gespräch fand in der vorfilmischen Realität als Videokonferenz statt, wird im Einspieler jedoch in einer Split-Screen-Montage in editierter Form zusammengeführt, wodurch eine Art ›Gespräch‹ von AzA inszeniert wird. Die audiovisuelle Inszenierung tritt somit deutlich als modulierendes Element hervor.

Die Analyse geht deshalb von der Audiovisualität des Materials aus und versucht über die Bewegungsdynamik des sich entfaltenden Bewegungsbildes herauszuarbeiten, wie es zur Erfahrung eines ›Gesprächs‹ kommt, wie diese moduliert ist und an welchen Stellen der Erfahrungsraum durch Irritationsmomente hinterfragbar wird. Methodisch werden sowohl medienästhetische als auch gesprächs- bzw. interaktionsanalytische Ansätze zusammengeführt. Die entsprechenden Bezüge werden in den einzelnen Abschnitten erläutert. Die Analyse folgt den folgenden Schritten: Zunächst (3.2) wird das Video in Bezug auf seine audiovisuelle Inszenierung als ein vereinfachtes Einstellungsprotokoll (Hickethier 2012: 37f.) dargestellt. Weiter (3.3) wird die Dynamik des sich entfaltenden Bewegungsbildes beschrieben (vgl. Greifenstein 2020; Müller/Kappelhoff 2018; Kindler-Mathôt et al. im Erscheinen). Der daraus hervorgehende Rhythmus wird dann in 3.4-3.5 mit Erkenntnissen der Gesprächs- und Interaktionsanalyse in Verbindung gebracht (vgl. Auer 2020; Kendon

7 Die Begrifflichkeiten »Multimodalität des Sprechens« und »audiovisuelle Multimodalität« werden in Müller (im Erscheinen) und Kindler-Mathôt et al. (im Erscheinen) entwickelt: »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 Kameraführung, Cadrage, Einstellungen, Montage und Ton realisiert.« (Kindler-Mathôt et al. im Erscheinen: Kap.2.3).

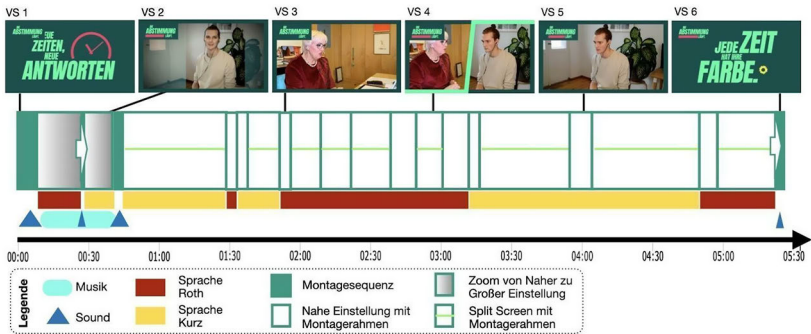
8 Abrufbar auf dem YouTube Kanal von BÜNDNIS 90/DIE GRÜNEN unter: <https://www.youtube.com/watch?v=Wp0FTzwlTJ8> (27.02.2024).

1990; Müller/Bohle 2007; Sacks et al. 1974; Selting et al. 2009). Abschnitt 3.6 bringt die Ergebnisse zusammen und beschreibt damit den Erfahrungsraum ›Gespräch‹ als komplexen Aushandlungsprozess von wechselseitigen Positionierungen.

3.2 Einstellungsprotokoll eines ›Gesprächs‹

Abbildung 1 zeigt schematisch die audiovisuelle Inszenierung des Videobeitrags als ein vereinfachtes Einstellungsprotokoll im zeitlichen Verlauf. Dies erlaubt einen ersten Überblick über die zeitliche Entfaltung der Inszenierung. Ich orientiere mich hier an Hickethier (2012: 37f.), eine ähnliche Vorgehensweise ist ebenfalls bei Kinder-Mathôt et al. (im Erscheinen) geschildert.

Abb. 1: Einstellungsprotokoll des Videobeitrags (05:28 Min).



Oberhalb des Zeitstrahls ist die auditive Ebene annotiert. Neben Musik (türkis) und anderen Soundeffekten (blau) umfasst dies auch die verbalen Äußerungen (rot für Roth, gelb für Kurz). Darüber liegt die visuelle Inszenierung. Sie stellt eine Abfolge von Montagesequenzen (dunkelgrün) und nahen Kameraeinstellungen (weiß) dar, die von Schnitten getrennt werden (senkrechte Linien). Einige Einstellungen sind exemplarisch als Videostills (in der Abbildung als VS gekennzeichnet) abgebildet, die auf die Bildraumkomposition verweisen. Die gesamte Sequenz wird von einem dunkelgrünen Montagerahmen umfasst (vgl. VS 3 und 4), ebenfalls wird im linken oberen Bildrand durchgehend ein kleines Fenster mit der Inschrift »Die Abstimmung läuft« und einem Prozessbalken eingeblendet.

Der Beitrag folgt in seiner Inszenierung einem bestimmten Muster. Er beginnt mit einer ins Bild fliegenden Einblendung eines dunkelgrünen Hintergrunds, die den Titel der Reihe zeigt (VS 1). Daraufhin folgen zwei frontale nahe Einstellungen, die die einzelnen Sprecher*innen im Bild zeigen (VS 2). Der Bildraum ist teilweise transparent-dunkelgrün kaschiert (vgl. Hickethier 2012: 48), sodass die Personen

im Bild in einem hellen Quadrat hervorgehoben sind. Mit in die Kamera gerichteten Blick stellen Roth und Kurz sich vor, während sich die Einstellung in einem langsamen Zoom bis fast in eine Großaufnahme verkleinert. In Sekunde 40 endet die Vorstellung mit einer Abblende, die das Thema des Einspielers (»Selbstbestimmung«) zeigt. Nun beginnt das »Gespräch«, welches als eine Abfolge von nahen Einstellungen auf Roth (VS 3) und Kurz (VS 5) inszeniert wird. Dominant ist dabei die Split Screen Montage (hellgrüne Linie, Bsp. VS 4), in der das Bild mittig in zwei Felder (rechts-links) aufgeteilt ist und somit beide Personen im Bildraum zeigt. Das »Gespräch« ist durch insgesamt 17 Schnitte bei stabiler Kameraeinstellung und -perspektive unterteilt. Der Beitrag endet mit einer letzten Abblende mit dem Motto der DBDK (»Jede Zeit hat ihre Farbe«).

3.3 Bewegungsbild »Gespräch«

Das Einstellungsprotokoll zeigt, dass sich die Inszenierung des »Gesprächs« deutlich von der vorangehenden »Vorstellung« unterscheidet: Die Montage der dunkelgrünen Hintergründe, die damit einhergehenden Soundeffekte sowie der frontale Zoom auf die Personen, stehen im Gegensatz zur Split Screen Montage und den vermeintlich zueinander orientierten Körpern rechts und links im Bildraum. Das »Gespräch« entfaltet dabei eine andere Dynamik, ein eigenes Bewegungsbild, welches im weiteren Verlauf der Analyse im Vordergrund steht. Wie kann das Bewegungsbild beschrieben werden, ohne von der Ebene der narrativen Repräsentation, also dem Gespräch als solchem auszugehen? Ich beziehe mich hier auf die Herangehensweise von Müller und Kappelhoff (2018) sowie Greifenstein (2020) und beschreibe die distinkte Bewegungsdynamik und Qualität des Bewegungsbildes. Greifenstein hat hierfür, in Anlehnung an Daniel Sterns (2011) Formen der Vitalität, ein spezifisches Beschreibungsvokabular entwickelt, mit dem die spezifisch-rhythmischen Bewegungsqualitäten verdeutlicht werden. Greifenstein formuliert exemplarisch:

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)

Mit Blick auf das Bewegungsvokabular kann die Dynamik des Bewegungsbildes des oben betrachteten Gesprächs als ein zügiges, geschmeidiges »Hin und Her« beschrieben werden, welches in seinem Rhythmus zunächst gleichmäßig fließend die Seiten auf einer horizontalen Ebene wechselt, dann aber ruckhafter wird und mehr und mehr einseitig dominiert. Das Bild eines Ping-Pong Spiels hilft, die spezifische Bewegung zu verstehen: Ein zügiger Ballwechsel lenkt die Aufmerksamkeit

alternierend von rechts nach links. Dabei wird der Ball jedoch nicht jedes Mal zurückgespielt, sondern es kommt zum neuen Aufschlag. Genauer: Während Kurz auf der rechten Seite den Ball annimmt und zurückspielt, führt Roth konsequent immer neue Aufschläge aus. Die Bewegung ist folglich zwar wechselseitig, aber dennoch einseitig dominierend.

Um diese deskriptive Ebene empirisch zu begründen, werfe ich in den folgenden beiden Abschnitten einen genauen Blick auf die Orchestrierung von audiovisueller Inszenierung und sprachlicher Multimodalität. Dabei werden das Zusammenspiel von Schnittmontage und Gesprächsstruktur sowie die Bildraumkomposition mit Blick auf die körperliche Koordination eines geteilten Interaktionsraums untersucht. Der Rückgriff auf gesprächsanalytische Forschung bietet dabei auch Erklärungsansätze für das sich entfaltende Dominanzgefühl und warum dieses aus Zuschauer*innenperspektive als Irritation in der Gesprächsdynamik sowie der Gesprächspraktik erfahren wird.

3.4 Rhythmus und Gesprächsstruktur

Das Zusammenspiel von filmischer Inszenierung und Gesprächsstruktur zeigt sich als starke Synchronisation: Jeder Sprecher*innenwechsel geht mit einem Schnitt einher (s. Abb.1), welcher immer den*die Sprecher*in in das Zentrum der Aufmerksamkeit rückt, da der Beginn eines neuen Turns nie im Split Screen, sondern immer in der Einzelansicht inszeniert ist. Das wahrgenommene Turn-Taking ist damit ein Produkt der Inszenierung. Das heißt, es ist kein Abbild der vorfilmischen Realität, sondern ein Teil der medialen Wirklichkeit und damit des Erfahrungsraums der Zuschauer*innen. Dennoch können die abwechselnden Turns vor dem Hintergrund der Turn-Taking Systematik von Sacks et al. (1974) betrachtet werden. Denn, wie im Falle einer dritten Person, die einem Gespräch von Aza zuhört, haben auch die Zuschauer*innen am Bildschirm ein Gefühl dafür, ob ein Gespräch ›rund‹ läuft oder Inkongruenzen auftreten.

Es wird zwischen zwei Möglichkeiten des Turn-Takings unterschieden (vgl. ebd.: 703): Erstens der *Fremdwahl* (ein*e Sprecher*in wählt den*die nächste*n Sprecher*in z.B. durch direkte Anrede oder eine Geste⁹) und zweitens der *Selbstwahl* (ein*e Sprecher*in wählt sich selbst als nächste*n). Kommt es zu keiner der Möglichkeiten, kann der*die vorherige Sprecher*in den Turn weiterführen (*current speaker continues*; Auer 2020: 165) oder das Gespräch unterbricht, bzw. endet. Die Übergänge zwischen Turns sind in der Regel fließend, ohne große Pausen oder simultanes Sprechen (vgl. Sacks et al. 1974: 700f.). Dies setzt voraus, dass sich die

9 Henne und Rehbock verweisen bereits 1979 auf die Relevanz der Multimodalität, insbesondere in Bezug auf Strategien des Turn-Takings. Zum multimodalen Turn-Taking, vgl. auch Schmitt (2005).

Sprecher*innen miteinander koordinieren und den nächstmöglichen Moment für einen Übergang, einen sogenannten *transition-relevance place* (TRP; ebd.: 703) zum einen antizipieren und zum anderen ihren nächsten Turn anschlussfähig vorbereiten: Die Interaktionsteilnehmer*innen »verarbeiten [...] die gerade entstehende TKE [Turn-Konstruktionseinheit] **während** ihrer Produktion und machen dabei bereits Vorhersagen darüber, wann sie zu einem Abschlusspunkt kommen könnte« (Auer 2020: 115; Herv. i. Orig.). Übergangsstellen gehen insbesondere mit syntaktischen und diskursiven Strukturen einher (vgl. ebd.: 117ff.) und sind meist prosodisch gekennzeichnet. Auch bei der Selbstwahl kündigen Sprecher*innen i.d.R. ihre Turn-Übernahme bereits an und machen sie somit antizipierbar, etwa durch die Veränderung der Körperhaltung in den Interaktionsraum hinein oder verstärkten Rezeptionssignalen (*pre-starters*) wie Nicken oder *hm_hms*.

Betrachten wir diese Aspekte mit der Turn-Taking Erfahrung des Videobeitrags, können zunächst vier Punkte festgehalten werden, auf welche im Folgenden eingegangen wird:

- 1) Insgesamt finden fünf Wechsel statt, die Turns sind dabei gleichmäßig in Anzahl und etwa proportional in Länge verteilt;
- 2) Vier der fünf Wechsel erfolgen durch Selbstwahl (ausgenommen zwischen Turn 2 und 3);
- 3) Trotz unmittelbarer Anschlüsse (*Latching*; Selting et al. 2009: 376) kommt es zu keinen Überlappungen;
- 4) Die meisten der Übergangsstellen sind prosodisch auffällig.

Zur genaueren Betrachtung werden Ausschnitte aus dem Gespräch als gesprächsanalytische Transkripte (nach GAT 2; Selting et al. 2009) aufbereitet.¹⁰

¹⁰ Die Orthografie wird zu Gunsten prosodischer Elemente außer Acht gelassen. So werden Pausen, Intonationsakzente und Phrasenenden mittels Kapitalen und Satzzeichen annotiert.

Abb. 2: Turn-Taking im Gespräch zwischen Roth und Kurz.

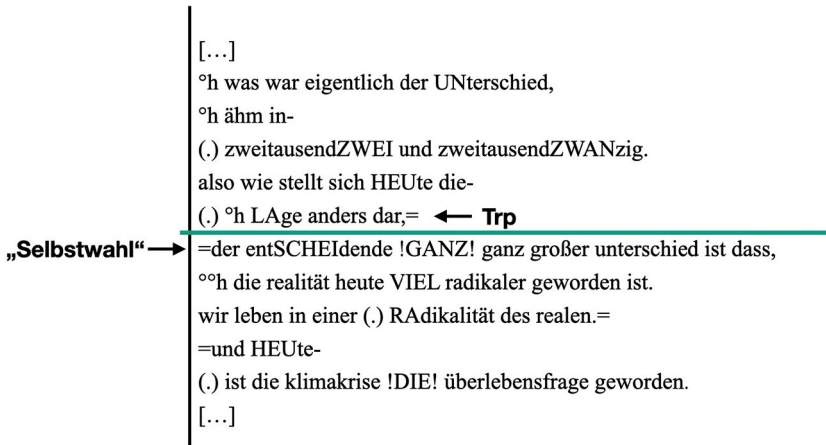
Turn Nr.	Sprecher*in	Zeile	Anfang und Ende des Turns als gesprächsanalytisches Transkript (GAT 2) (Auslassungen durch [...] gekennzeichnet)	Sprecherwechsel
1	GK:	001 002 003 [...] [...] 017 018 019 020 021 022	also ich bin ja so_n hoffnungslöser ÖKO gewesen; ich bin ja wegen so: so KLIma und umwelt sachen- (.) zu den GRÜnen gekommen- [...] [...] da macht man sich nich (.) FREI von- wenn man die part=wenn man die tür zum PARTEIbüro zu macht, das ist ja natürlich KLAR- aber die halt vor allem polItisch dafür kämpfen dass- (.) ALle menschen da- (.) ähm sich frei entfalten KÖNnen=	(Selbstwahl)
2	CR	023 024 025 026	=was BEDEUtet denn- der WERT selbstbestimmung; HEUte- für DICH,=	Selbstwahl
3	GK	027 028 [...] [...] 033 034 035 036	=ich KOMM ja aus nem- (.) nem kleinen (.) KAFF bei berchtesgaden, [...] [...] ähm (.) und für mich war das ne toTALE:- (.) äh beFREIung; irgendwie festzustellen dass das ja EIgentlich- auch ganz ANDers geht;	Fremdwahl (Frage mit Anrede)
4	CR	037 038 039 [...] [...] 072 073 074 075	für MICH war natürlich selbstbestimmung- äh ganz ganz ganz STARK- verKNÜPFT mit dem selbstbestimmungs!RECHT!. [...] [...] =und HEUte- (.) ist die klimakrise !DIE! überlebensfrage geworden. aber was bedeutet !DAS! für selbstbestimmung. was bedeutet das für selbstbestimmung für (.) MEInen LEbensstil.=	Selbstwahl
5	GK	076 077 078 079 [...] [...] 124 125 126 127 128 129 130	=was du GRAD- ähm geSAGT hast zum- (.) ähm selbstbestimmungsrecht der FRAU und so weiter eingehen.= =weil- [...] [...] = GLAUbe ich, !EI!gentlich ein schritt hin zu MEHR freiheit (.) wenn man NICHT mehr geZWUNGen ist;= =zum beispiel warten bis man VOLLjährig ist;= Bis man FÜHRerschein hat;= =TAUsende euro ausgibt für ein auto,= =und dann erst irgendwie moBIL zu sein?= [...] [...]	Selbstwahl
6	CR	131 132 133 [...] [...] 141 142 143 144 145 146 147 148	=selbstbestimmung verwirklicht FREIHEIT. (.) !SELBST! (.) bestimmung verwirklicht auch- die !WÜRDE! jedes menschen; [...] [...] wo es gruppen GIBT, die (.) die würde nicht JEDES menschen,= =sondern die WÜRde des DEUTschen menschen- oder des MÄNNlichen menschen- oder des CHRIStlichen menschen- oder des Heterosexuellen- oder des WEIßen, in den VORdergrund stellen.	Selbstwahl

In der Zuschauer*innenerfahrung wechseln sich Kurz und Roth in ihren Turns ab. Alle Wechsel (ausgenommen zwischen Turn 2 und 3) finden durch Selbstwahl statt. Die Fremdwahl entsteht aus einer Frage-Antwort Sequenz: Roth fragt, was Selbstbestimmung für Kurz heute bedeute. Syntaktisch und prosodisch mit leicht steigender Intonation zum Phrasenende (gekennzeichnet durch ein Komma), stellt der Turn eine abgeschlossene Einheit mit Handlungsaufforderung an Kurz dar. Entsprechend folgt ein Turn von Kurz mit einer inhaltlich passenden Antwort und womit eine Form von Intersubjektivität hergestellt wird: Aus Zuschauer*innenperspektive erscheint es, als ob Kurz den Turn als Fremdwahl verstanden und den TRP entsprechend antizipiert hat. Es kommt zu einem Latching, also einem unmittelbaren Anschluss der Antwort (gekennzeichnet durch Gleichheitszeichen).

Latching stellt innerhalb des ›Gesprächs‹ die Regel der Turn-Anschlüsse dar (Ausnahme zwischen Turn 3–4). Damit gehen für TRPs ungewöhnliche Tonhöhenbewegungen zum Ende der Intonationsphrasen einher, da sie gleichbleibende (gekennzeichnet als Bindestrich) oder steigende (Komma für leicht steigende, Fragezeichen für stark steigende Intonation) Tendenzen aufweisen, z. B. bei *KÖNnen* (022) oder *moBIL zu sein?* (130). Auer (2020: 153f.) bemerkt, dass solche Formen in Gesprächen der finalen Tonhöhenbewegung als »weiterweisend« gelten, also auf Unabgeschlossenheit des Turns verweisen. Die Inszenierung, genauer die Schnitte, die die Erfahrung von TRPs modulieren, führen also zu ersten feinen Irritationsmomenten, welche prägend für die Dynamik des Bewegungsbildes sind: Die schnellen kontinuierlichen Hin- und Herbewegungen zwischen den Turns, welche durch die damit einhergehenden Schnitte und die Wechsel von Split Screen zu Einzelansicht, zum einen akzentuieren, zum anderen auch eine fast rastlose Zügigkeit entfalten.

Mit Blick auf die TRPs tritt noch eine weitere Stelle hervor: In Turn 7 kommt es zu einem TRP, welcher durch erneute Selbstwahl von Roth, also einer Weiterführung, nicht zum Sprecher*innenwechsel führt. Das folgende Transkript verdeutlicht die Stelle.

Abb. 3: Ausschnitt aus Turn Nr. 4 (CR) mit ungenutztem TRP.



Wieso ist diese Stelle auffällig? Der TRP resultiert aus einer syntaktischen Fragekonstruktion mit entsprechend steigender Tonhöhenbewegung zum Phrasenende: [...] *also wie stellt sich HEUTE die- (.) °h LAge anders dar*, Er stellt damit, im Gegensatz zu den zuvor beschriebenen Äußerungen, eine abgeschlossene Turn-Konstruktion dar. Darauf beantwortet Roth die Frage selbst, indem sie direkten lexikalischen Bezug zum ersten Teil des Turns nimmt (*°h was war eigentlich der UNterschied, [...] =der entSCHEidende !GANZ! ganz großer unterschied ist dass, [...]*). Auch wenn das Selbstbeantworten von Fragen durchaus in Gesprächen vorkommt, ist an dieser Stelle die sprachliche Markierung einer Übergangsstelle durch die Phrasenintonation (insbesondere mit Blick auf die sonst auftretenden Unregelmäßigkeiten) recht markant. Ebenfalls gibt es keinen Hinweis darauf, dass es sich bei der Frage um eine rhetorische Frage handelt, da sie weder prosodisch noch metakommunikativ als solche markiert wird.

Die untypischen Formen der TRPs durch finale Tonhöhenbewegung oder Latching ohne Überlappung sind Produkte der Schnittmontage, die bspw. den Turn von Roth als Beantwortung der eigenen Frage gestaltet (in Transkript 2 als horizontale Linie). Sie stellen Irritationsmomente dar, die ausschließlich auf der Ebene der Zuschauer*innenerfahrung bestehen und kein Phänomen der vorfilmischen Realität sind. Sie sind Formen der von Holly (2015) angesprochenen Transkription, die zu einer Bedeutungsanreicherung führt. Im Falle der selbstbeantworteten Frage besteht dies in einer Irritation, die nicht die grundsätzliche Dynamik des Gesprächs unterbricht, aber dennoch als vielleicht sogar etwas unhöflich auffällt. Holly fügt in diesem Zusammenhang an, dass die Medialität damit dem*der Sprecher*in die alleinige Auktorialität entzieht, da die »performative Letztfassung« seiner Rede nicht mehr kontrolliert« (Holly 2015: 89) werden kann.

In Bezug auf den Rhythmus des Bewegungsbildes im Zusammenspiel mit der Gesprächsstruktur kann also festgehalten werden, dass die Erfahrung der Hin- und Her Dynamik eng mit der audiovisuellen Inszenierung des Turn-Takings verknüpft ist. In der Zuschauer*innenerfahrung wird, trotz der Schnitte und der anhaltenden Irritationsmomente durch die ungewöhnlich intonierten TRPs, ein Aushandlungsprozess wie in einem ›natürlichen‹ Gespräch von AzA deutlich. Die Videokonferenz wird – zumindest mit Blick auf das Turn-Taking – in der Erfahrung nicht relevant gesetzt, da keine hierfür typischen Phänomene wie lange Pausen oder Simultansprechen auftreten. In dieser Hinsicht ist das Turn-Taking kein Produkt der Repräsentation, sondern wird vielmehr im Rezipieren der Zuschauer*innen körperlich-affektiv erfahren und unmittelbar verstanden: nicht durch die ›Darstellung‹, sondern als eine spezifische Bewegungsdynamik, die die Wechselseitigkeit einer Gesprächserfahrung von AzA aufgreift, wenn auch in ungewöhnlicher Weise.

Ein weiterer Aspekt, der diese Erfahrung hervorhebt, zeigt sich in der Gestaltung des Bildraumes, insbesondere durch die bereits benannte Split Screen Montage. Die folgende Analyse zeigt, wie in der Montage die beiden Körper in einer für das Gespräch von AzA geläufigen Ausrichtung in der Zuschauer*innenerfahrung einen gemeinsamen Interaktionsraum zu etablieren scheinen.

3.5 Bildraumkomposition und Interaktionsraum

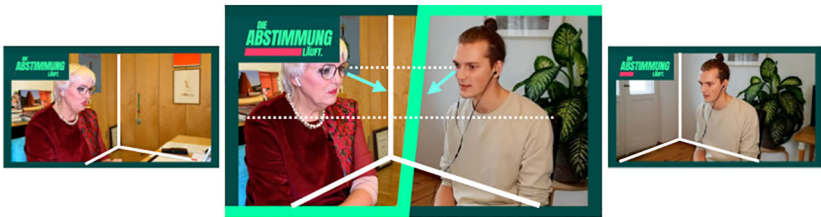
Die Aufteilung des Bildschirms in zwei unterschiedlich gestaltete Flächen sowie deren jeweilige Bildraumkomposition ist ein deutlich zu bestimmender Fall von medialer Wirklichkeit, denn die Perspektive auf zwei unterschiedliche Orte gleichzeitig ist in der vorfilmischen Realität schlicht unmöglich (s. Abb. 4). Der Split Screen bringt dabei zwei kontrastierende Aspekte zum Vorschein: Zum einen hebt die prägnante hellgrüne Linie quer durch die Bildmitte eine klare Trennung hervor, die ebenfalls mit Blick auf die einzelnen Flächen und deren Räumlichkeiten offensichtlich macht, dass sich die Personen an unterschiedlichen Orten befinden. Zum anderen schafft er dennoch eine Verbindung zwischen Personen und Räumen, die in der Zuschauer*innenerfahrung einen geteilten Interaktionsraum formen. Im Folgenden wird, diesmal mit Bezug auf interaktionsanalytische Forschung, die Konstitution dieses Raumes erläutert und damit ein weiteres Argument für die Erfahrung eines audiovisuell orchestrierten Gesprächs von AzA angeführt, da damit die zweite konstituierende mediale Bedingung inkludiert wird: Kopräsenz durch Anwesenheit.

In einem Gespräch von AzA orientieren sich die Teilnehmer*innen nicht nur sprachlich aneinander, sondern bilden in der Ausrichtung ihrer Körper einen geteilten Interaktionsraum. Müller und Bohle (2007) bezeichnen diese körperliche Koordination als das *Fundament fokussierter Interaktion*. Kendon (1990) beschreibt in seiner Arbeit zu *Spatial organization in social encounters* äußerst differenziert verschie-

dene wiederkehrende Konfigurationen: Für das Gespräch zwischen zwei Personen sind dies bspw. *vis-a-vis arrangement* (Personen stehen sich gegenüber), *L-arrangement* (Personen stehen in einem Winkel nebeneinander) oder *side-by-side arrangement* (Personen stehen nebeneinander) (vgl. ebd.: 213). Er prägt in diesem Zusammenhang den Begriff der *F-formation systems*. Diese Systeme treten immer dann auf, wenn zwei oder mehr Personen einen gemeinsamen Handlungsraum (*o-space*; ebd.: 211) zwischen sich etablieren und diesen kooperativ durch die Organisation ihres körperlichen Verhaltens aufrechterhalten. Eine F-Formation ist keine Bedingung für ein Gespräch, sie ist jedoch präferiert (vgl. Auer 2017: 10).

Wie ist der Interaktionsraum im Bildraum nun gestaltet? Mit Blick auf die zwei Bildflächen zeigen sich im Hintergrund zwei verschiedene büroartige Räume, die sich in ihrer Farbgebung abheben. Roth und Kurz sitzen jeweils aus den Außenseiten der Felder mit dem Rücken zu einer Wand an einem Tisch (bei Roth je nach Einstellung) sichtbar. Beide tragen kleine Kopfhörer, die ein Hinweis auf die Videotelefonie sind. Die Inszenierung hebt dies aber nicht weiter hervor, nur in wenigen Einstellungen ist das Endgerät von Roth im rechten Bildrand zu sehen. Die starke örtliche Trennung wird mithilfe einer Split Screen Montage überbrückt, sodass visuell ein Miteinander inszeniert wird (s. Abb. 4).

Abb. 4: Audiovisuelle Überwindung der räumlichen Trennung zu Gunsten eines geteilten Interaktionsraums.



Zunächst sind beide Personen so im Bildraum ausgerichtet, dass es im Zusammenschnitt im Split Screen erscheint, als säßen sie sich in einem offenen Winkel gegenüber: Beide nehmen jeweils etwa gleich viel Raum ein, Augen und Schultern sind auf gleicher Höhe im Bildraum gesetzt, beide sind in sitzender Position halbfrontal von der Kamera aufgenommen. In Verlängerung formen ihre Schultern deshalb eine L-Konfiguration, die in Richtung der Zuschauer*innen (also der Kameras) geöffnet ist. Verlängert werden die Achsen des Ls zusätzlich durch die Kameraperspektive in den jeweiligen Raum, die den Blick in eine Ecke freigeben. Im Split Screen sind beide Bildfelder so nebeneinander gesetzt, dass nicht nur die Körper auf gleicher Höhe ausgerichtet sind, sondern auch die Sockelleisten an den Wänden ineinander übergehen und so die beiden Räume visuell zu einem verbunden werden (in

der Abbildung als weiße durchgezogene Linie gekennzeichnet). Es entsteht so nicht nur ein L-förmiger Interaktionsraum zwischen den Körpern mit Innen und Außen, sondern auch ein gemeinsamer Raum mit zwei Wänden. Durch die Interaktion zwischen den beiden Teilbildern des Split Screens wird eine Erfahrung von Parallelität und Gleichzeitigkeit evoziert (vgl. Wulff 1991: 282), welche im *Mise-en-Scène* zum Eindruck von gemeinsamer Anwesenheit und Kopräsenz verbunden wird.

Der scheinbar geteilte Raum und die körperliche Ausrichtung zueinander sind folglich ein Produkt der Inszenierung. Die Modulation stellt einen wortwörtlich un/realen Interaktionsraum dar: unreal in der vorfilmischen Realität, real in der Erfahrung der Zuschauer*innen. Doch auch in der Konstruktion des Interaktionsraumes ergibt sich ein weiteres Irritationsmoment, welches im Blickverhalten besteht. Die aus der Abbildung 4 hervorgehende Blickrichtung auf Schulter- und Brustbereich (türkise Pfeile) würde in einem natürlichen Gespräch von AzA als unangebracht empfunden werden. Im Gegensatz zur oben beschriebenen Irritation beim Turn-Taking tritt der Blick jedoch nicht so stark hervor, da die L-Konfiguration und Ausrichtung der Körper, vor allem der gleichen Höhe der Augen und Schultern, dominanter ist. Es lässt sich zwar folglich eine intramediale Umdeutung nach Holly (2015) erahnen – der Blick auf Brusthöhe wird nur durch die spezifische mediale Konfiguration möglich – sie tritt jedoch nicht prominent hervor, sondern schmiegte sich eher in die modulierte Kopräsenz ein.

3.6. Multidimensionaler Erfahrungsraum ›Gespräch‹

Sowohl die Analyse des Rhythmus als auch der Bildraumkomposition aktivieren die Erfahrung eines Gesprächs von AzA, indem dessen mediale Bedingungen, Wechselseitigkeit und Kopräsenz durch Anwesenheit, audiovisuell konstruiert und/oder hervorgehoben werden und dadurch zu einem affektiven Zuschauer*innenerleben werden. Mit Blick auf das in 3.3 beschriebene Bewegungsbild bieten beide Ebenen der Analyse bereits empirische Anhaltspunkte, die sich vordergründig auf die medialen Bedingungen als eine Art ›Gerüst‹ des Gesprächs beziehen. In diesem abschließenden Analyseteil möchte ich zusätzlich noch auf den Inhalt, bzw. die pragmatische Ebene eingehen, in dem ich mich auf den Bedeutungsaushandlungsprozess innerhalb dieses Gerüsts beziehe. Damit wird neben den Rhythmen und Bewegungsrichtungen auch die Qualität und zunehmende Intensität des Bewegungsbildes adressiert und somit auch die spezifisch affektive Dimension des Erfahrungsraums. Abbildung 5 greift dazu das Bild des Ping-Pong Spiels auf und führt die Ebenen der Analyse zusammen.

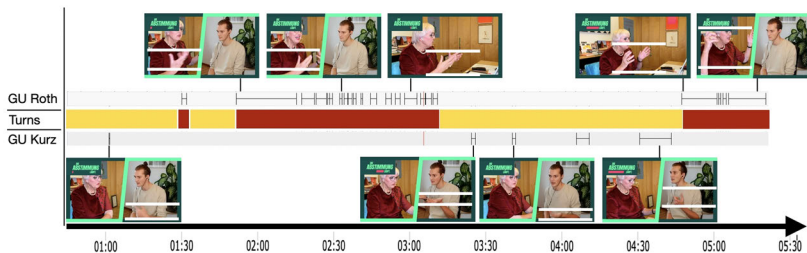
Abb. 5: Modulation einer einseitig dominierenden Hin und Her Bewegung.



Untereinander stehen jeweils Videostills der mit dem Turn-Taking einhergehenden Einstellungen. Pfeile in rot (Roth) und gelb (Kurz) visualisieren die jeweiligen Turns, also die Ballwechsel. Eine orange Verbindungslinie tritt dann auf, wenn ein Ball weitergespielt und nicht fallen gelassen wird (vgl. weiter unten: dynamische Anschlussfähigkeit). Ausschnitte des Transkripts verdeutlichen die Sprachhandlungen, die als Überschriften neben den nummerierten Turns aufgeführt sind. Dabei treten drei Aspekte zu Tage, die die Bewegungsqualität greifbar machen.

1.: Die *Fokussierung der Aufmerksamkeit ist wechselseitig* zwischen den jeweils sprechenden Personen. Während eines Turns ist mehr Bewegung auf entweder der rechten oder linken Seite des Bildraums erfahrbar. Diese ist audiovisuell, da sie zum einen durch sprachliche, hörbare »Bewegung«, zum anderen durch sichtbare Bewegung im Bild aktiviert wird. Dabei rücken Handgesten in den Fokus, die auf Grund der Kameraeinstellung und Perspektive gut sichtbar sind. Annotiert wurden die gestischen Einheiten (GU), die von Ruhepositionen der Hände abgegrenzt werden. Ohne auf eine detaillierte Gestenanalyse (zur Methodologie vgl. Müller 2024) einzugehen, wird deutlich, dass Roth mehr und tendenziell größere bzw. raumeinnehmendere Handgesten ausführt als Kurz. Zum Ende des Gesprächs nehmen sie außerdem an Intensität zu, sowohl in der Bewegungsqualität (ruckhafte, akzentuierte Schläge) als auch in der Größe (Armbewegungen durch den gesamten Bildraum, in der Abbildung markiert durch weiße Linien). Durch den Split Screen und auch die Einzeldarstellung der Personen auf der rechten oder der linken Seite im Bild wird die Aufmerksamkeit zusätzlich gelenkt. Der Blick wandert so je nach Sprecher*in von rechts nach links.

Abb. 6: Annotation der Gesteneinheiten (GU) im zeitlichen Verlauf des »Gesprächs« mit exemplarischen Stills.



2.: Das bereits beschriebene audiovisuell induzierte Turn-Taking erfolgt zwar zügig, dennoch sind die Übergänge *nicht durchgehend dynamisch anschlussfähig*. So nimmt nur Kurz zweimal den vorangehenden Turn dynamisch auf und führt ihn hin weiter, einmal als Erfüllung der Paar-Sequenz (s. Abb. 5, Turn 3) und einmal mittels eines metakommunikativen Kommentars zur vorangehenden Äußerung (s. Abb. 5, Turn 5). Solche expliziten Anknüpfungen finden bei Roth nicht statt. Sie bildet immer neue, unabhängige Einheiten, die, abgesehen von der Orientierung am Thema »Selbstbestimmung«, nicht in direktem Bezug zum vorherigen Turn stehen.

3.: Thematisch kann ein *Aushandlungsprozess* rund um das Thema »Selbstbestimmung« verfolgt werden. Die Turns stellen, vor allem ab Turn 3, Positionierungshandlungen dar. Bei Roth findet eine Steigung dieser Positionierung statt,

die in Turn 6 ihren Höhepunkt findet, welcher in prosodisch (und gestisch) beinah überbetonten, kurzen Aussagephrasen als eine Art Abschlussplädoyer formuliert wird (=selbstbestimmung verwirklicht FREIHEIT. (.) !SELBST! (.) bestimmung verwirklicht auch- die !WÜRDE! jedes menschen; [...]). Daraufhin endet das Gespräch sofort, da sich qua Montage ein grünes Bild einschiebt. Roth behält damit das letzte Wort. Im Vergleich dazu, und mit Blick auf die dynamische Anschlussfähigkeit der Turns, positioniert sich auch Kurz zunächst in Turn 3, schwächt dies jedoch in Bezug auf Roths darauf folgenden Turn (Nr. 4) deutlich ab (*das möchte ich vielleicht da noch vorne ANgestellt haben; das darf man nicht verWECHseln;=wenn ich mich da ein bisschen eingeschränkt fühle ist das was völlig anderes als wenn man dafür kämpfen muss dass zum beispiel paraGRAphen abgeschafft werden.*). Während Roth somit konsequent einen Standpunkt formuliert, von dem sie nicht abweicht, sondern diesen zum Ende hin noch deutlich prosodisch hervorhebt, ordnet sich Kurz im Verlauf diesem unter.

Das sich im Verlauf steigende Dominanzgefühl der linken Bildseite wird folglich durch dreierlei Aspekte moduliert:

- 1) durch die wiederholte Selbstwahl ohne dynamischen Anschluss sowie durch die Selbstbeantwortung der eigenen Frage, die als Irritationsmoment markiert wurde;
- 2) durch die Intensivierung der Bewegungserfahrung (sprachlich und gestisch), die die Aufmerksamkeit fokussiert;
- 3) durch die Verstärkung der sprachlichen Positionierungen in Bezug auf das Thema des Gesprächs.

Alle drei Aspekte modulieren den Erfahrungsraum. Die Verschmelzung von sprachlicher und audiovisueller Multimodalität (vgl. Müller im Erscheinen; Kindler-Mathôt im Erscheinen) formt das Bewegungsbild, und damit die Zuschauer*innenerfahrung, eine unmittelbar affektive Erfahrung von Dominanz als gefühlter Ausdrucksqualität. Es entfaltet sich eine mediale Wirklichkeit, die nicht der vorfilmischen Realität entspricht, über die, ausgehend vom Material, nur wenige Aussagen getroffen werden können. Der einzige Anhaltspunkt hierfür ist die Videokonferenz, die aber nur am Rande Teil des Erfahrungsraums wird, sowie die Irritationsmomente, die diesen immer wieder erschüttern und so, zumindest für kurze Momente, die Medialität hinterfragbar machen.

4. Diskussion

Welche Rolle spielt nun die Zuschauer*innenerfahrung medialer Wirklichkeiten vor dem Hintergrund zeitgenössischen Medienkonsums? Welche Implikationen gehen damit einher? Eingangs wurde bereits die zunehmende Digitalisierung

angesprochen, aus welcher immer neue mediale Formen hervorgehen. Seien es politische Reden im Bundestagsfernsehen, positive Affirmationen von Instagram Influencer*innen oder populistische YouTube-Shorts: Im Konsum werden diese alle zu Zuschauer*innenerfahrung neuer, medialer Wirklichkeiten. Eingebettet in spezifische Interfaces und thematische Kontexte gehen sie immer mit einer grundsätzlichen Affizierung der Zuschauer*innen einher. Hier schließt sich die Diskussion um die Wirkmacht von Medien an, insbesondere mit Blick auf die vermeintliche »Objektivität« (Luginbühl/Schneider 2020: 60) von Massenmedien. Aus medienästhetischer Perspektive zeigt sich, dass es die affektive Grundierung ist, die die affektiven Bewertungen und Haltungen der Zuschauer*innen gegenüber moralisch-politischen Sachverhalten strukturieren. Scherer (2024) bezeichnet dies auch als Formen zeitgenössischer Propaganda. Wenn vor diesem Hintergrund mediale Wirklichkeiten nicht mehr hinterfragbar sind und mit vorfilmischer Realität gleichgesetzt werden, wenn der Abstand verschwindet und wir als Zuschauer*innen der »Illusion« einer scheinbar objektiv beobachtenden Kamera aufsitzen, werden die Potenziale der medialen Wirklichkeiten offenbar. Die Modulation von politischer Bildung, Wissensgenerierung sowie Meinungen und Bewertungen, sind dann abhängig von der Medienkompetenz der Zuschauer*innen – nicht nur in Bezug auf die Überprüfung inhaltlicher Richtigkeit, sondern vor allem mit Blick auf die Fähigkeit, die Medialität zu hinterfragen und etwa Irritationsmomente aufzudecken. Sonst werden Meinungen zu Fakten und Deep Fakes zu Realität. Es bleibt also nicht bei einem einfachen Übertragungsprozess, sondern im Konsum werden neue soziale Ordnung und Realitäten, neue Praktiken im Sinne einer Poesis des Filme-Sehens hervorgebracht (vgl. Kappelhoff 2018). McLuhan und Fiore bringen dies mit Blick auf die historische Entwicklung im Zusammenspiel von Medien und Gesellschaft auf den Punkt:

Gesellschaften wurden schon immer stärker von der Natur der Medien geprägt, mit denen Menschen kommunizieren, als vom Inhalt der Kommunikation. [...] Ohne Kenntnis der Wirkungsweise von Medien ist es unmöglich, Prozesse des sozialen und kulturellen Wandels zu verstehen. (2016: 8)

Als Praktik des Medienkonsums stehen die audiovisuellen Bilder damit im direkten Zusammenhang mit dem (demokratischen) Willensbildungsprozess. Vor diesem Hintergrund plädiert der Beitrag für eine Stärkung der Medienkompetenz, die insbesondere die charakteristische Medienspezifik audiovisueller Bilder als Bewegungsbilder vor dem Hintergrund von Verstehensprozessen berücksichtigt.

5. Schluss

Der Artikel hat aus einer medienästhetischen Perspektive heraus die Modulation medialer Wirklichkeit in der Zuschauer*innenerfahrung audiovisueller Bilder untersucht. Dabei wurde betont, dass diese Wirklichkeit weder bloßes Abbild der vorfilmischen Realität ist, noch sich auf der Ebene eines repräsentierten Narrativs entfaltet, sondern vielmehr als Bewegungsbilder mit spezifischen Dynamiken aus dem Zusammenspiel von sprachlicher und audiovisueller Multimodalität hervorgeht. Im Moment des Sehens und Hörens werden sie zu affektiv grundierten Bedeutungsgestalten, die nicht kognitiv dekodiert, sondern unmittelbar körperlich erfahrbar werden. Dies wurde am Beispiel eines Videobeitrags der digitalen Bundesdelegiertenkonferenz der Partei BÜNDNIS 90/DIE GRÜNEN 2020 illustriert. Das hier erfahrene Gespräch ist nicht nur durch die Inszenierung geformt, sondern – wie in der Analyse des Turn-Takings gezeigt – geht erst aus ihr hervor.

Obwohl in der vorfilmischen Realität als Videokonferenz stattgefunden, entfaltet sich vielmehr die Dynamik eines Gesprächs von AzA, dessen mediale Bedingungen (Wechselseitigkeit und Kopräsenz durch Anwesenheit) durch die audiovisuelle Inszenierung eines geteilten Interaktionsraums aktiviert werden. Nur in Irritationsmomenten wird die Erfahrung gestört und somit die Medialität für die Zuschauer*innen hinterfragbar. Aus dem Zusammenspiel sprachlicher und audiovisueller Multimodalität geht eine komplexe Gestalt – ein Erfahrungsraum medialer Wirklichkeit – hervor, womit sowohl affektive als auch pragmatische Aushandlungsprozesse in der Zuschauer*innenerfahrung aktiviert und relevant gesetzt werden. Die Erfahrung dieser medialen Wirklichkeiten geht immer mit Prozessen des sozialen und kulturellen Wandels einher, womit die gesellschaftliche Relevanz dieser komplexen Gestalten im Kontext zeitgenössischen Medienkonsums deutlich zu Tage tritt.

Literatur

- Auer, Peter (2017): »Anfang und Ende fokussierter Interaktion: Eine Einführung« in: InLiSt 59, S. 1–56.
- Auer, Peter (2020): »Die Struktur von Redebeiträgen und die Organisation des Sprecherwechsels«, in: Karin Birkner/Peter Auer/Angelika Bauer/Helga Kotthoff (Hg.), Einführung in die Konversationsanalyse, Berlin/Boston: De Gruyter, S. 106–235.
- Bakels, Jan-Hendrik (2017): Audiovisuelle Rhythmen. Filmmusik, Bewegungskomposition und dynamische Affizierung des Zuschauers, Berlin/Boston: De Gruyter.

- Bühler, Karl (1933): *Ausdruckstheorie. Das System an der Geschichte aufgezeigt*, Jena: Gustav Fischer.
- Deleuze, Gilles (1989): *Das Bewegungs-Bild: Kino 1*, Frankfurt a. M.: Suhrkamp.
- Deleuze, Gilles (1990): *Das Zeit-Bild: Kino 2*, Frankfurt a. M.: Suhrkamp.
- Gehrau, Volker (2008): »Rezeptionsforschung«, in: Uwe Sander/Friederike Von Gross/Kai-Uwe Hugger (Hg.), *Handbuch Medienpädagogik*, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 341–345.
- Gehrau, Volker (2017): *Die Beobachtung als Methode in der Kommunikations- und Medienwissenschaft*, 2. Aufl., Stuttgart: utb GmbH.
- Greifenstein, Sarah (2020): *Tempi der Bewegung – Modi des Gefühls. Expressivität, heitere Affekte und die Screwball Comedy*, Berlin/Boston: De Gruyter.
- Haddington, Pentti/Tuire Oittinen (2022): »Interactional Spaces in Stationary, So-bile, Video-Mediated and Virtual Encounters« in: Andreas H. Jucker/Heiko Hausendorf (Hg.), *Pragmatics of Space*, Berlin/Boston: De Gruyter Mouton, S. 317–361.
- Henne, Helmut/Helmut Rehbock (1979): *Einführung in die Gesprächsanalyse*, Berlin/New York: Walter de Gruyter.
- Hickethier, Knut (2012): *Film- und Fernsehanalyse*, 5., aktualisierte und erweiterte Aufl., Stuttgart: J.B. Metzler.
- Holly, Werner (2015): »Sprache, Bild, Text. Visualität und Intermedialität von Sprache«, in: Ludwig Eichinger (Hg.), *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*, Berlin/Boston: De Gruyter, S. 71–92.
- Horst, Dorothea (2018): *Meaning-Making and Political Campaign Advertising: A Cognitive-Linguistic and Film-Analytical Perspective on Audiovisual Figurativity*, Berlin/Boston: De Gruyter.
- Kappelhoff, Hermann (2004): *Matrix der Gefühle. Das Kino, das Melodram und das Theater der Empfindsamkeit*, Berlin: Vorwerk 8.
- Kappelhoff, Hermann (2008): »Zuschauergefühl. Die Inszenierung der Empfindung im dunklen Raum des Kinos«, in: Beate Söntgen/Geraldine Spiekermann (Hg.), *Tränen*, München: Wilhelm Funk, S. 195–206.
- Kappelhoff, Hermann (2018): *Kognition und Reflexion. Zur Theorie filmischen Denkens*, Berlin/Boston: De Gruyter.
- Kappelhoff, Hermann/Jan-Hendrik Bakels (2011): »Das Zuschauergefühl. Möglichkeiten qualitativer Medienanalyse«, in: *Zeitschrift für Medienwissenschaft* 5:2, S. 78–96.
- 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, S. 121–153.
- Kendon, Adam (1990): »Spatial Organization in Social Encounters. The F-Formation System«, in: Adam Kendon (Hg.), *Conducting Interaction. Patterns of Behavior in Focused Encounters*, Cambridge: Cambridge University Press, S. 209–238.

- Kindler-Mathôt, Clara/Cornelia Müller/Jana Junge/Katerina Papadopoulou (im Erscheinen): »Sprechen in audiovisuellen Medien: Systematik für eine medienästhetische Analyse multimodaler Kommunikation«, in: Rita Vallentin/Dorothea Horst (Hg.), *Sprache entgrenzen. Beiträge zu einer kulturwissenschaftlichen Linguistik*, Bielefeld: transcript.
- Koch, Gertrud (2003): »Nachstellungen – Film und historischer Moment«, in: Eva Hohenberger/Judith Keilbach (Hg.), *Die Gegenwart der Vergangenheit. Dokumentarfilm, Fernsehen und Geschichte, Vorwerk 8*, S. 216–229.
- Krämer, Sybille (1998): »Über die Kommunikation im Internet. Überlegungen zur telematischen Interaktion«, in: *Netzdiskurs – Das Internet und der Strukturwandel von Kommunikation und Öffentlichkeit 67/97*, S. 11–22.
- Lanwer, Jens Philipp (2019a): »Blended joint-attention in medial-vermittelter Interaktion«, in: Konstanze Marx/Axel Schmidt (Hg.), *Interaktion und Medien. Interaktionsanalytische Zugänge zu medienvermittelter Kommunikation*, Heidelberg: Winter, S. 99–124.
- Lanwer, Jens Philipp (2019b): »Erzählen im virtuellen Interaktionsraum«, in: *Networx 84:1.0*, S. 4–44.
- Luginbühl, Martin (2019): »Mediale Durchformung. Fernsehinteraktion und Fernsehmundlichkeit«, in: Konstanze Marx/Axel Schmidt (Hg.), *Interaktion und Medien. Interaktionsanalytische Zugänge zu medienvermittelter Kommunikation*, Heidelberg: Universitätsverlag Winter, S. 125–146.
- Luginbühl, Martin/Jan Georg Schneider (2020): »Medial Shaping from the Outset: On the Mediality of the Second Presidential Debate, 2016«, in: *Journal für Medienlinguistik/Journal for Media Linguistics Vol. 3:1*, S. 57–93.
- Marx, Konstanze/Axel Schmidt (Hg.) (2019): *Interaktion und Medien. Interaktionsanalytische Zugänge zu medienvermittelter Kommunikation*, Heidelberg: Winter.
- McLuhan, Marshall/Quentin Fiore (2016): *Das Medium ist die Massage. Ein Inventar medialer Effekte*, 4. Aufl., Stuttgart: Tropen.
- Meier, Christoph (2000): »Neue Medien – neue Kommunikationsformen? Strukturmerkmale von Videokonferenzen«, in: Werner Kallmeyer (Hg.), *Sprache und neue Medien*, Berlin/New York: Walter de Gruyter, S. 195–221.
- Müller, Cornelia/Ulrike Bohle (2007): »Das Fundament fokussierter Interaktion. Zur Vorbereitung und Herstellung von Interaktionsräumen durch körperliche Koordination«, in: Reinhold Schmitt (Hg.), *Koordination. Analysen zur multimodalen Interaktion*, Tübingen: Narr, S. 129–165.
- Müller, Cornelia/Hermann Kappelhoff (2018): *Cinematic Metaphor. Experience – Affectivity – Temporality*, Berlin/Boston: De Gruyter Mouton.
- Müller, Cornelia (2024): »A Toolbox of Methods for Gesture Analysis«, in: Alan Cienki (Hg.), *Handbook of Gesture Studies*, Cambridge: Cambridge University Press, S. 182–216.

- Müller, Cornelia (im Erscheinen): »Sprache-Sprechen, hören, sehen, fühlen. Eine medienästhetische Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien«, in: Rita Vallentin/Dorothea Horst (Hg.): Sprache entgrenzen. Beiträge zu einer kulturwissenschaftlichen Linguistik. Bielefeld: transcript.
- Plessner, Helmut (1982): »Die Deutung des mimischen Ausdrucks. Ein Beitrag zur Lehre vom Bewußtseins des anderen Ichs« [in Zusammenarbeit mit J. J. Buijten-dijk], in: Philosophischer Anzeiger 1, S. 72–126.
- Sacks, Harvey/Emanuel A. Schegloff/Gail Jefferson (1974): »A Simplest Systematics for the Organization of Turn-Taking for Conversation«, in: Language 50:4, S. 696–735.
- 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: Cornelia Müller/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressemer (Hg.), Body – Language – Communication. Handbücher zur Sprach- und Kommunikationswissenschaft, Berlin/Boston: De Gruyter Mouton, S. 2081–2092.
- Schmitt, Reinhold (2005): »Zur multimodalen Struktur von *turn-taking*«, in: Gesprächsforschung – Online Zeitschrift zur verbalen Interaktion 6, S. 17–61.
- 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 Mertzluft/Miriam Morek/Frank Oberzaucher/Jörg Peters/Uta Quasthoff/Wilfried Schütte/Anja Stukenbrock/Susanne Uhmann (2009): »Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)«, in: Gesprächsforschung 10, S. 353–410.
- Sobchack, Vivian (1992): The address of the eye. A Phenomenology of Film Experience, Princeton: Princeton University Press.
- Stern, Daniel (2011): Formen der Vitalität. Die Erforschung dynamischen Erlebens in Psychotherapie, Entwicklungspsychologie und den Künsten, Frankfurt a. M.: Brandes & Apsel.
- Wulff, Hans J. (1991): »Split Screen: Erste Überlegungen zur semantischen Analyse des filmischen Mehrfachbildes«, in: Kodikas/Code 14:3-4, S. 281–290.
- Wundt, Wilhelm (1975): Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte, 3. Aufl., Band 1: Die Sprache, Aalen: Scientia Verlag.
- 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, S. 1–29.

