

Sprache-Sprechen, Hören, Sehen, Fühlen

Eine medienästhetische Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien

Cornelia Müller

Abstract ›Sprache entgrenzen‹ wird im vorliegenden Beitrag als genuin multimodal verfasstes Sprache-Sprechen und als Sprachgebrauch in audiovisuellen Medien gedacht. Entgrenzung bezieht sich auf den Gegenstand der Sprachwissenschaft der nach wie vor dem Primat des Lautsprachlichen folgt, aber auch auf neue Formen der Sprachgebrauchsforschung. Er plädiert dafür, medienlinguistische Untersuchungen multimodalen Sprachgebrauchs medientheoretisch zu fundieren und skizziert einen transdisziplinären (filmwissenschaftlich-linguistischen) Vorschlag, wie Sprache-Sprechen in audiovisuellen Medien aus einer medienästhetischen Perspektive gefasst werden kann. Ausgangspunkt sind Parallelen von Sprache-Sprechen und audiovisuellen Bildern: Beide werden als Zeit- und Bewegungsfigurationen gedacht, die erst in der verkörperten Erfahrung eines Gegenübers ›Bedeutung‹ gewinnen. Im Kern plädiert der Beitrag dafür, die Multimodalität des Sprachgebrauchs als ästhetische Erfahrung zu beschreiben, die durch die Orchestrierung der audiovisuellen Medien moduliert wird. Damit gerät die Generierung, Modulierung und Mobilisierung von Gefühlen als verkörperte Erfahrung in den Blick der Multimodalitätsforschung. Der Beitrag formuliert einen theoretischen Rahmen und eine methodische Perspektive für eine kulturwissenschaftlich fundierte Kritik am Sprachgebrauch in audiovisuellen Medien.

Keywords Multimodalität; Gestik und Sprache; Temporalität; Embodiment; Ästhetische Erfahrung; Affektivität

1 Einleitung

›Sprache entgrenzen‹? Aus der Sicht einer Erforschung des Sprachgebrauchs bedeutet dies zuallererst, dass das Sprechen einer Sprache ein multimodaler, durch und durch körperlicher Vorgang ist. Es bedeutet auch, die Dichotomie verbal – nonverbal aufzugeben und Sprache als genuin multimodal verfasst anzuerkennen. Die Gestenforschung und die Gebärdensprachforschung haben in den vergangenen Jahrzehnten aufzeigen können, dass Sprache-Sprechen nicht am gesprochenen Wort festzumachen ist (Müller et al. 2013, 2014). Damit gerät die Frage nach der ›Multimodalität des Sprechens‹ in den Blick sprachwissenschaftlicher Forschung.

›Sprache entgrenzen‹ bedeutet im Kontext allseits verfügbarer audiovisueller Medien aber auch, dass die Multimodalität des Sprechens im Alltag des heutigen Sprachgebrauchs immer mehr und immer häufiger über die Bildschirme von Smartphones, Tablets oder Laptops wahrgenommen, geteilt und so zum Movens partikularer Gemeinschaftsbildungen wird, die sich in Form von Klickzahlen oder *Follower*innen* auf Social-Media-Kanälen manifestieren. Audiovisuelle Bilder von Menschen die zueinander, miteinander, übereinander sprechen, sind zum selbstverständlichen und ganz alltäglichen Bestandteil einer globalisierten, durchweg medial strukturierten Alltagswelt geworden. Familienfeste, Reisen, Shoppingtouren werden in einem nie gekannten Ausmaß audiovisuell dokumentiert, die Handykamera ist immer dabei, ob Silvesterfeuerwerk oder Überschwemmungen, es wird aufgezeichnet und oft noch im gleichen Moment geteilt.

Auch für die Sprachforschung bedeutete die Möglichkeit der audiovisuellen Aufzeichnung von natürlichen und experimentellen Situationen des Sprachgebrauchs eine mediale Revolution. Waren von Quintilian bis Wilhelm Wundt Reflektionen über Gesten und ihr Zusammenwirken mit der Sprache auf die teilnehmende Beobachtung angewiesen, machte die Erfindung des Films es möglich, die Flüchtigkeit des Sprache-Sprechens zu überwinden. Nun ließ sich eine Kommunikationssituation wieder und wieder und in unterschiedlichen Geschwindigkeiten anschauen. Das audiovisuelle Medium wurde so zum ›Mikroskop‹ der Erforschung von Körperbewegung, Kommunikation und Interaktion. Ohne audiovisuelle Dokumentationen ist die zeitgenössische Gesten- und Gebärdensprachforschung, die Erforschung von multimodaler Kommunikation und Interaktion, nicht denkbar. Sie spielt daher für die Ausweitung der Perspektive linguistischer Forschung eine entscheidende Rolle.

Sprache-Sprechen begegnet uns im Alltag und in der Forschung nicht nur in großem Umfang, sondern auch in einer Vielfalt verschiedener audiovisueller Formate. So betrachtet, entpuppt sich die Entgrenzung von Sprache als eine alltägliche Praxis des immer schon verkörperten Sprache-Sprechens, in dem die Körperlichkeit miteinander sprechender Personen durch ein körperliches Wahrnehmen audiovisueller Bilder moduliert wird.

Es liegt auf der Hand, dass Film- und Videotechnik ganz neue Möglichkeiten eröffnet haben, das Sprechen und Gestikulieren in ihrer Einheit als gemeinsam artikulierte Körperbewegungen zu beschreiben. Gleichzeitig beobachten wir eine tiefgreifende mediale Durchformung der Alltagswelt. Audiovisuelle Medien prägen die Formen, in denen uns Sprache-Sprechen begegnet, ob auf YouTube oder Instagram, ob in der Kommunikation per Zoom, ob als Serie auf Netflix, als Talkshow im Fernsehen oder als allzeit in einer Mediathek verfügbare Nachrichtensendung. Mit der alltäglichen Verfügbarkeit von Videoaufzeichnungen und der allgegenwärtigen Präsenz audiovisuell festgehaltener Formen des Sprachgebrauchs stellt sich für eine kulturwissenschaftlich orientierte Sprachgebrauchsforschung die drängende Frage nach einer kritischen Reflexion der Bedingungen, von denen Analysen mündlichen Sprachgebrauchs auf der Grundlage audiovisueller Aufzeichnungen auszugehen haben. Eine Erweiterung der sprachwissenschaftlichen Betrachtung der Multimodalität des Sprache-Sprechens um eine kulturwissenschaftlich-medienwissenschaftliche Perspektive erscheint daher dringend erforderlich. Auf diesen Umstand reagiert der vorliegende Beitrag mit einem transdisziplinären Vorschlag, wie Sprache-Sprechen, genauer wie die Multimodalität des Sprechens im Kontext audiovisueller Medien theoretisch wie methodisch zu fassen ist.

Der Beitrag skizziert Eckpunkte einer kulturellen Praxis, in der Sprache-Sprechen zunehmend mit der Zirkulation audiovisueller Bilder verbunden ist. Er macht sich für die These stark, dass es die Gestaltung der audiovisuellen Multimodalität ist, die die körperlich-sprachliche Multimodalität orchestriert und dass es dieses Zusammenspiel ist, das die Wahrnehmung, das Erleben und das affektive Mitschwingen, die affektive Bewegung der Zuschauer*innen an den Bildschirmen gestaltet. Da diese beiden Ausdrucksformen »die Multimodalität des Sprechens« und die »Multimodalität audiovisueller Medien« gewissermaßen in die Zuständigkeitsbereiche verschiedener Fächer fallen, ergibt sich hier die drängende Aufgabe, die fachwissenschaftliche Expertise zu diesen beiden Gegenstandsbereichen ernst zu nehmen und auf dieser Grundlage eine transdisziplinäre Perspektive zu entwickeln. Dieser Versuch

wurde im Kontext verschiedener filmwissenschaftlich-sprachwissenschaftlicher Forschungsprojekte unternommen und am Gegenstand multimodaler Metaphorik in Sprache, Geste und audiovisuellen Medien (Greifenstein et al. 2018; Müller/Kappelhoff 2018) entwickelt.¹ Der kinesisch-kinematographische Ansatz wurde in verschiedenen Forschungsarbeiten am Beispiel verschiedener Gegenstandsbereiche erprobt und fortentwickelt (Greifenstein 2020; Horst 2018; Kindler-Mathôt et al. i. d. Bd.; Scherer 2024; Schmitt 2020).²

Wie sich diese kinesisch-kinematographische Perspektive methodisch umsetzen lässt, dokumentiert im vorliegenden Band der Beitrag »Sprechen in audiovisuellen Medien. Systematik für eine Analyse multimodaler Kommunikation« (Kindler-Mathôt et al. i. d. Bd.). Der vorliegende Beitrag skizziert, wie sich diese kulturwissenschaftliche Methode theoretisch herleiten und begründen lässt.

Dabei werden Parallelen sprachlicher und audiovisueller Multimodalität aufgezeigt, die entscheidend dafür sind, wie Zuschauer*innen das Sprechen in audiovisuellen Medien wahrnehmen. Es geht darum zu verstehen, *was* die Betrachter*innen audiovisueller Bilder vom Sprache-Sprechen zu hören und zu sehen bekommen und wie die Orchestrierung der audiovisuellen Bilder das Fühlen der Zuschauer*innen moduliert.

Im Folgenden gehe ich auf die Eckpunkte dieser medienästhetischen Perspektive ein und stelle drei Aspekte vor, die das Sprache-Sprechen und die audiovisuelle Multimodalität miteinander teilen: (1) Temporalität und Bewegung, (2) Ausdrucksbewegung und Interaffektivität, (3) Interkorporalität und ästhetische Erfahrung. Diese systematischen Parallelen bereiten den Weg nicht nur zur Methodenentwicklung, sondern sie sind vor allem auch Voraussetzung für eine kritische Analyse der Wirkmacht von Sprache-Sprechen als ein Hören, Sehen, Fühlen, kurzum ein Erfahren audiovisueller Bilder.

1 Eine ausführliche Darstellung der filmtheoretischen Grundlagen findet sich in »Kognition und Reflexion: Zur Theorie filmischen Denkens« (Kappelhoff 2018) sowie in »Cinematic Metaphor: Experience – Affectivity – Temporality« (Müller/Kappelhoff 2018), hier auch in ihren Bezügen zur Körpergeste und verkörpertem Sprechen.

2 Beispiele sind: deutsche und polnische Wahlwerbespots (Horst 2018), Hollywood Kino, wie die Screwball Comedy (Greifenstein 2020) oder Hitchcock's *Spellbound* (Müller/Kappelhoff 2018; Schmitt 2020) sowie Deutscher Spielfilm (Schmitt 2020), Politiksendungen, Nachrichten, Werbespots und Musikvideos (Müller/Kappelhoff 2018; Schmitt 2020; Scherer 2024) und parlamentarische Reden (Kindler-Mathôt et al. i. d. Bd.).

2 Temporalität und Bewegung

Audiovisuelle Bilder entfalten sich, ebenso wie das Sprechen, in der Zeit. Sie existieren nur in der Zeit, genauer noch, nur als Bewegungen, die sich in einer bestimmten zeitlichen Kontur entfalten.

2.1 Sprechen ist Körperbewegung

Denkt man die Sprache vom Sprechen her, dann ist sie zunächst einmal Körperbewegung. Laute sind die hörbaren Effekte der komplexen Bewegungen von Stimmbändern, Zunge, Lippen, die den Luftstrom in unterschiedlicher Weise durch den Artikulationsraum, durch Kehle, Mund, Rachen und Nase strömen lassen. Man könnte auch sagen, Laute entstehen als Folge artikulatorischer Gesten. So verweisen etwa die Gebärdensprachforscher Armstrong, Stokoe und Wilcox auf Ulric Neisser (1976), der das Sprechen als »articulatory gesturing« betrachtet und argumentiert, dass die Perzeption des Sprechens vergleichbar sei mit der Perzeption anderer Gesten (Armstrong/Stokoe/Wilcox 1995: 8)³. Die Abtrennung des Lautlichen von der Körperbewegung ist ein Artefakt, das von der Realität des Sprache-Sprechens absieht.

Lange bevor in der Sprachentwicklung die Lautsprache auf den Plan tritt, kommunizieren Eltern und Säuglinge mit den Bewegungen ihrer Körper. Kommunikation findet in der Ontogenese zunächst einmal als ganzkörperliches Ereignis statt, aus dem sich der »artikulierte« Gebrauch der Lautsprache nach und nach herauschält. Der Säuglingsforscher Daniel Stern beschreibt diese kommunikative Verschränkung der Körper als Tanz (Stern 1985): ein Tanz, dessen Spuren in der Synchronisierung von Körperbewegung von Gesprächsteilnehmer*innen sichtbar werden (Kendon 1970, 1972; Oben/Brône 2015, 2016) und über die die Sprecher*innen ein zuallererst körperliches Miteinander herstellen, das man als körperliches Fundament der Interaktion beschreiben kann (Müller/Bohle 2007). Eine sehr typische Form, wie ein soziales Miteinander durch die räumliche Orientierung der Körper hergestellt

3 »Oddly, much more work has been done to describe speech than to describe signing in gestural terms. The psychologist Ulrich Neisser, for example, has noted that it is possible to describe speech as »articulatory gesturing«, and to treat speech perception as comparable to perceiving gestures of others kinds«. Vgl. auch den Hinweis auf das Sprechen als »subtly interleaved patterns of movements, coordinated across articulators« (Armstrong/Stokoe/Wilcox 1995: 8).

wird, ist die Kreisform. Diese Anordnung schafft im wörtlichen Sinne einen Interaktionsraum. Zu diesem Raum haben die Mitglieder der Gruppe exklusiven Zugang, er hat ein Innen (die Vorderseite der Körper) und ein Außen (die Rückseite der Körper), grenzt die so Zusammenstehenden nach Außen ab, und Neuzugänge erfordern eine Rekonfiguration, eine Erweiterung des Kreises. Das Miteinanderstehen, das körperliche Aufeinander-bezogen-Sein ermöglicht das Miteinander-Sprechen. Kendon hat solche räumlich-körperlichen Konfigurationen schon früh als Formationen und Verhaltenssysteme sozialer Begegnungen beschrieben (Kendon 1990). Eingehender hat er das Zusammenstehen von Personen in einem Kreis untersucht. Kendon fasst diese Kreisformation als einen sozialen Raum, der eine spezifische Form sozialen Verhaltens organisiert, ein »F-formation system«:

An F-formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct, and exclusive access. Such a pattern can be seen in the circle of the free-standing conversational group. Here the participants stand so that they all face inwards to a small space which they cooperate together to sustain and which is not easily accessible to others who may be in the vicinity. The system of behavioral organization by which such a spatial-orientational pattern is established and sustained will be called an F-formation system. (Kendon 1990: 209)

Nimmt man diese grundsätzliche Multimodalität des Sprechens ernst, so verschiebt sich der Gegenstand sprachwissenschaftlicher Betrachtung von einem körperlosen kontext-entbundenen Sprachsystem zu einem kontext-gebundenen verkörpertem – immer zeitlich zu denkenden – Miteinander-Sprechen: »Multimodal Languageing«. So gesehen ist Sprechen eine Körperpraxis, die inter-subjektives Verstehen ermöglicht.⁴ Auf die Perspektive des Verstehens zielt auch Flussers Unterscheidung der epistemischen und der ästhetischen Geste des Sprechens (Flusser 1991: 59). In der Geste des Sprechens verbindet sich der Weltbezug (epistemische Geste) mit der Gerichtetheit auf ein Gegenüber (ästhetische Geste). Dabei ist zu beachten, dass sich Flussers Begriff der ästhetischen Geste auf den alten Begriff der »Aisthesis« als ein Verstehen, nicht auf den umgangssprachlichen Gebrauch von »Ästhetik« als »Schönheit«, bezieht.

4 »Also ist das Denken des Sprechenden eine »adaequatio intellectus ad intellectum«, und seine Absicht ist nicht, irgendeine »objektive« Wahrheit einzufangen, sondern ein inter-subjektives Verstehen zu ermöglichen« (Flusser 1991: 59).

2.2 Gestikulieren und Sprechen als Zeitformen

Denkt man das Sprechen vom Gestikulieren her, dann fällt auf, dass sich beide in der Zeit entfalten: Sprechen und Gestikulieren treten als Zeitformen in Erscheinung, sie sind in dieser Zeitlichkeit miteinander verschränkt und beziehen aus dieser Verschränkung Sinn und Bedeutung. Diese Beobachtung gehört zu den Grundeinsichten der Gestenforschung und zwar in expliziter Abgrenzung vom Credo einer Trennung von Verbalität und Körperbewegung, das ja namensgebend für die Erforschung sogenannter non-verbaler Kommunikation war (Kendon 1972, 1980; McNeill 1985, 1992). Dagegen ist für die Gestenforschung gerade die Verschränkung beider Ausdrucksmodi *Movens* eines neuen Forschungsfeldes.⁵ Ob als Theorie einer Psychologie des Denkens, wie bei McNeill (1992), aus sozial-anthropologischer Sicht bei Kendon (2004) oder aus praxeologischer Perspektive, wie bei Streeck (2009), immer geht es um die Verschränkung des Gestikulierens mit dem Sprechen.

Über die verschiedenen Schulen und Ansätze hinweg besteht große Einhelligkeit darüber, dass die zeitliche Verschränkung der gestischen Bewegung mit dem Redefluss so feinkörnig synchronisiert ist, dass jeder Versuch eine »zur Rede passende Geste« zu planen und auszuführen zu einer signifikanten Asynchronität zwischen sprachlichem und gestischem Ausdruck führt. Diese hochgradige Synchronisierung hat zuerst Kendon in seinen frühen Arbeiten gezeigt (Kendon 1970, 1972). Sein Vorschlag einer linearen Segmentierung von Gesten und ihrer Bedeutung war ein Meilenstein der Gestenforschung. Die Analyse des gestischen Bewegungsverlaufs in *Preparation*, *Stroke* und *Recovery* (Kendon 1972, 1980, 2004) hat sich als Bezugspunkt experimenteller wie qualitativer Forschung fest etabliert (vgl. Bressem/Ladewig 2011). Kendon orientiert sich hier zunächst an der prosodischen Struktur sprachlicher Äußerungen und konnte zeigen, dass im Gestikulieren ebenso wie beim Sprechen komplexe zeitliche Strukturen entstehen können (Kendon 1972). Ruhepositionen der Hände markieren die äußeren Grenzen gestischer Einheiten (*Gestu-*

5 Die Erforschung sprachbezogener Gesten (u.a. auch als Schnittstelle zur Gebärdensprachforschung) hat sich in den vergangenen Jahrzehnten fest als interdisziplinäres Forschungsfeld etabliert. Institutionell ist die Gestenforschung in der »International Society for Gesture Studies (ISGS): <https://www.gesturestudies.com/> verankert. Mit einer internationalen Zeitschrift (<https://benjamins.com/catalog/gest>), einer Buchreihe (<https://benjamins.com/catalog/gs>) und den im zweijährigen Rhythmus stattfindenden internationalen Tagungen haben sich vielbeachtete Orte des wissenschaftlichen Austausches und der Publikation von Forschungsergebnissen herausgebildet.

re Units). Innerhalb der gestischen Einheiten entfalten sich eine oder mehrere gestische Phrasen. Kendon definiert eine gestische Phrase als Abfolge einer Bewegung aus der Ruheposition hin zum Ausführungsort der Geste (Vorbereitungsphase, *Preparation*) und einer Ausführungsphase (*Stroke*). Diese Phrasen können recht komplexe temporale Strukturen ausbilden, etwa durch Wiederholungen oder indem eine Geste nur teilweise zurückgeführt wird (*Partial Retraction*) und sich in Form neuer *Strokes* weiteren Gesten anschließen, ohne dass sich die Hände zurück in eine Ruheposition bewegen (Kendon 2004: Chapter 7).⁶

So hängt etwa im Kontext gesprächsanalytischer Rekonstruktionen multimodaler Interaktionen die gestische Bedeutung von der zeitlichen Positionierung der Geste im Gesprächszug, im *Turn*, ab. Ob die Geste zu Beginn eines Redezuges, als Teil einer Konstruktionseinheit (*Turn-Constructional Unit*, TCU) oder im potenziellen Turn-Übergaberaum (*Transition Relevance Place*, TRP) positioniert ist, bestimmt ihre gesprächsstrukturelle Bedeutung.⁷ Die Frage der Entfaltung von Gesten in der Zeit ist auch wesentlich für die semantische Bedeutung einer Geste. Insbesondere singuläre Gesten⁸ sind semantisch unterspezifiziert – ihre lokale, indexikalisch verankerte Bedeutung erhalten sie wesentlich durch die zeitliche Synchronizität mit sprachlichen Elementen der Äußerung (Müller 2010, 2015, 2016). So kann etwa ein gestisch modelliertes Rechteck die Form eines Aquariums ebenso darstellen wie einen Theorierahmen (Müller 1998, 2010, 2016). Die spezifische, lokale semantische Bedeutung der Geste hängt von ihrer zeitlichen Verschränkung mit der Rede ab. Kurz, wir haben es beim Gestikulieren und Sprechen mit multimodalen Zeitformen

6 Kendon verbindet diese Segmentierung gestischer Bewegungen mit einer Festlegung darauf, welcher Bestandteil des gestischen Bewegungsverlaufs bedeutungsrelevant ist: die Phase der Vorbereitung und der Höhepunkt, der *Stroke*. Anders formuliert: Kendon schneidet den gestischen Bewegungsverlauf nach dem Höhepunkt ab. Ob man Kendon in diesem Punkt folgt oder nicht, es bleibt ein breiter Konsens in der Forschung, dass die zeitliche Koordinierung von Sprechen und Gestikulieren entscheidend für die wechselseitige Konstitution von Bedeutung ist.

7 Ich verweise hier auf Arbeiten von Schegloff (1984), von Streeck/Hartge (1992) oder auch von Schmitt (2005).

8 Singuläre Gesten sind Gesten, die im Moment des Sprechens entstehen. Sie unterscheiden sich von rekurrenten Gesten, die sich durch einen hybriden Charakter auszeichnen und sowohl sedimentierte als auch lokale Bedeutungsaspekte aufweisen und von Emblemen, deren Bedeutung kulturell vollständig festgelegt, also konventionalisiert ist (vgl. Müller 2010, 2018, 2024b). Übergänge zwischen den Typen sind fließend.

zu tun, deren Bedeutung aus der zeitlichen Verschränkung der beiden Modi hervorgeht.

Soweit so gut, aber haben diese Segmentierungen, diese Einheiten auch etwas mit dem Verstehen zu tun? Damit ist nun nicht nur das Verstehen als Form interaktiver Aushandlungsprozesse, sondern Verstehen als Verkörperung, als Bewegungsverstehen angesprochen. Kehren wir nochmal zu Kendon zurück: Er verbindet seine Analyse des gestischen Höhepunkts (*Gesture Stroke*) als bedeutungsvolle Phase gestischer Bewegungen mit einem Rekurs auf die Ausdrucksbewegung als theoretisches Konstrukt, das in der Anthropologie, Soziologie, Tanztheorie, Sprachwissenschaft, Psychologie und Filmtheorie der zwanziger Jahre des letzten Jahrhunderts ausgearbeitet wurde. Man denke hier an Plessner (1982 [1925]), Wundt (1975 [1900–1925]), Bühler (1933, 1934), Simmel (1993 [1901–1908]) und den Tänzer Rudolf von Laban (1956). Kendon bezieht sich mit seinem Vorschlag direkt auf die Bewegungsanalyse des Ausdruckstänzers Laban, die im Übrigen bis heute Grundlage von Tanztherapien und einer immer noch gebräuchlichen Tanznotation ist (Kennedy 2013). So schreibt Kendon:

The phase of the movement excursion closest to its apex is usually recognized as the phase when the ›expression‹ of the gesture, ›whatever it may be, is accomplished. It is the phase of the excursion in which the movement dynamics of ›effort‹ and ›shape‹ are manifested with greatest clarity (see Laban and Lawrence 1947; Dell 1970; Bartineff (sic!) and Lewis 1980). This phase is called the stroke. (Kendon 2004: 112)

Wie diese gestischen Phrasen und Einheiten aber verstanden werden, bleibt weitgehend offen, und so transformieren anthropologische, semiotische oder auch psychologische Ansätze in der Gestenforschung Kendons Bewegungsfigurationen in ›flüchtige‹ aber gleichwohl statische ›Zeichenbilder‹. Gesten werden als Züge in einem Gespräch (Streeck/Hartge 1992), als verbunden mit einer lexikalischen Einheit (Schegloff 1984) oder als Ausdruck mentaler Bilder (McNeill 1992) betrachtet. Obwohl die zeitliche Positionierung der Geste relativ zur Sprache in allen Ansätzen eine wichtige Rolle spielt, bleibt die Temporalität, die Bewegungsentfaltung von Gesten, wenn überhaupt, nur implizit relevant. Typischerweise würde dann eine Geste, die die Stufen einer Treppe nachzeichnet und die überlappend mit dem Adjektiv ›steigend‹ verwendet wird als auf dieses Wort bezogen oder im besten Fall auf die Nominalphrase »steigende Lohnzusatzkosten« bezogene gestische Einheit

betrachtet (vgl. Abb. 1). Was bei dieser Perspektive in den Hintergrund tritt, ist die Verschränkung von Gestikulieren und Sprechen, ist ihr gestalthafter Charakter.⁹ Gestikulieren und Sprechen als multimodale Zeitformen zu betrachten, nimmt dagegen gerade die Zeitlichkeit der Bewegungsfigurationen als multimodal orchestrierte Bewegungsbilder zum Ausgangspunkt (vgl. auch Müller 2024a/b).

2.3 Körperliche und filmische Bewegungsbilder

Betrachtet man den folgenden Ausschnitt aus einer Bundestagsrede von Angela Merkel, so formuliert sie in ihrer Entgegnung auf einen Zwischenruf deutlich affektiv engagiert wie folgt: {und SIE ham KEI}ne antwort (.) auf die frAge-{was man denn tut, (-)um die immer STEIgenden lohnzusatz}kosten zu vermeiden.<

Abb. 1: Transkript ›Steigende Lohnzusatzkosten‹

{und SIE ham KEI}ne antwort (.)auf die frage-

[illegible]

Den wesentlichen sachlichen Punkt ihrer Entgegnung formuliert Merkel sprachlich und gestisch:¹⁰ Die immer steigenden Lohnzusatzkosten zeichnet

- 9 Mondadas Vorschlag, Gesten und Sprache als Formen multimodaler Ressourcen zu be-
trachten, öffnet den Blick auf den gestalthaften Charakter multimodaler Kommunika-
tion und Interaktion (Mondada 2013).
- 10 Die Transkription orientiert sich zum einen an den gesprächsanalytischen Konventionen (GAT2) (Selting et al. 2009) sowie an den Notationskonventionen von Müller (1998: Appendix). Jede Zeile besteht aus drei Teilzeilen. Die Transkription berücksichtigt damit sowohl die lautsprachlichen Bewegungen wie Intonation und Pausen als auch die sichtbaren Bewegungen des Körpers, hier der Handgesten. Die GAT2-Konventionen wurden wie folgt erweitert: Geschweifte Klammern ({ }) markieren die Grenzen der

sie gestisch als Treppenstufen in die Luft – ein Beispiel multimodaler Metaphorik, wie wir sie häufig in der Alltagssprache beobachten können (Müller 2008; Müller/Cienki 2009; Müller/Kappelhoff 2018). Betrachtet man die Geste genauer, so entfaltet sie sich als Abfolge von kleinen bogenförmigen Bewegungen, jede für sich etwas höher und weiter zur rechten Merkmals in die Luft gezeichnet. Erst über die Zeit der Entfaltung entsteht so ein Bild aufsteigender Treppenstufen. Hier von *einer* Geste zu sprechen, die sich auf *ein* Wort bezieht und so eine multimodale Metapher bildet, verwandelt eine dynamische Bewegungsfiguration in ein statisches ›Zeichenbild‹.

Wie aber verstehen Betrachter*innen solche Bewegungsabläufe? Wie kommt es dazu, dass wir eine Folge von Treppenstufen wahrnehmen und nicht einzelne voneinander unabhängige kleine Bögen? Um dies zu erläutern, möchte ich Bezug auf Plessners Konzept des Bewegungsbildes nehmen (Plessner 1982 [1925]). Plessner beschreibt Körperbewegungen als dynamische Gestalten, die sich Beobachter*innen als bildhafte Bewegungsgestalten darstellen. Als Beispiele nennt er unter anderem Handbewegungen wie Greifen und Abwehren, die als Ausgangspunkt gestischer Als-Ob-Handlungen äußerst produktiv sind (Müller 2024a), so produktiv, dass sie Eingang in die Alltagsmetaphorik nicht nur des Deutschen gefunden haben: Das ›Be-greifen‹ eines Sachverhaltes, das ›Ab-wehren‹, das ›Weg-Wischen‹ eines Argumentes speisen sich aus solchen Bewegungsbildern.

Wo immer im Reich des Organischen Bewegungen erscheinen, verlaufen sie nach einheitlichem Rhythmus, zeigen sie eine, wohl auch experimentell nachweisbare, dynamische Gestalt. Sie rollen nicht stückhaft ab, als ob ihre Phasenfolge aus einzelnen Elementen assoziiert worden wäre, bilden kein Zeitmosaik, sondern eine gewisse Ganzheit ist vorgegeben, innerhalb derer die einzelnen Bewegungskurven variierbar sind. [...] Diese Ganzheiten gehören zum Organismus durch sein Verhältnis zur Umwelt, seine Morphologie, seine artspezifischen Instinkte als motorische Kategorien [...]. Infolgedessen sind die Bewegungsgestalten bildhaft, wenn auch über eine gewisse Zeitdauer erstreckt, dem Beobachter gegenwärtig. [...] Greifen, Fliehen, Abwehren, Suchen, aber auch schon die »affektlosen« Formen wie Gehen,

Gesteneinheiten, Strokes sind **fett** markiert. Die Zeile unterhalb zeigt den Phasenverlauf der Gesten nach Kendon (**P**reparation, **S**troke, **R**ecovery, **r**estposition). Die Strokes sind durchnummeriert und werden in der untersten Zeile in ihrer Form und Bewegung beschrieben. Gleichbleibende Formen tragen die gleiche Nummerierung (vgl. S.2).

Fliegen, Schwimmen [...] stellen solche Bewegungsbilder dar. (Plessner 1982 [1925], 77–78)

Merkels Geste entfaltet sich in der Zeit. Sie ist kein statisches Bild einer Treppe, sie ist ein Bewegungsbild (Abb. 2). Es besteht kein kausaler Zusammenhang zwischen den einzelnen *Gesture Strokes*, den einzelnen in die Luft gezeichneten aufsteigenden Bögen, vielmehr entsteht der Zusammenhang erst im Prozess des Betrachtens. Kendon hat darauf hingewiesen, dass sich Gesten grundsätzlich in der Zeit entfalten. Mit Plessner lässt sich nun festhalten, dass Gesten Bewegungsbilder sind, die als Gestalten »dem Beobachter gegenwärtig« sind (Plessner 1982 [1925], 78).

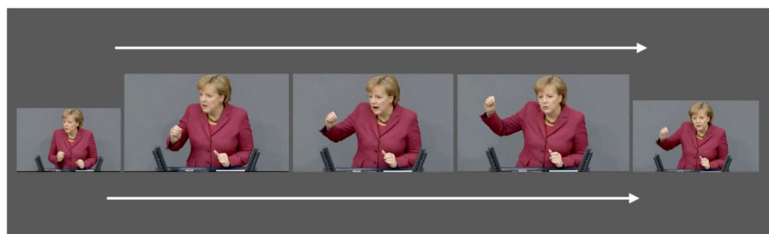
Abb. 2: Merkels Geste als Bewegungsbild



Ebenso wie gestische Bilder entfalten sich auch Filmbilder in der Zeit. Auch hier haben wir es nicht mit einer Abfolge von statischen Zeichenbildern zu tun, sondern mit Bewegungsbildern oder mit Zeitbildern. Diese Bewegungsbilder weisen ein dynamisches Profil, eine zeitliche Struktur auf (Deleuze 2008 a/b; Kappelhoff 2006). Sie sind ein unentwegtes Werden und Vergehen. Abb. 3 illustriert diese zeitliche Dimension audiovisueller Bewegungsbilder für den kleinen Ausschnitt aus der Rede Merkels im Bundestag. Es geht also nicht nur um die Zeit, in der sich die Geste entfaltet, sondern auch um die des audiovisuellen Bildes. Allerdings sind diese Bewegungsbilder keine prae hoc existierenden Einheiten, sie formieren sich erst im Prozess der Wahrnehmung von Zuschauer*innen. Ebenso wie das gestische Bewegungsbild entsteht auch ein audiovi-

suelles Bewegungsbild erst in der Wahrnehmung der Betrachter*in als Wahrnehmungsgestalt.

Abb. 3: Merkels sprachlich-gestische Äußerung als audiovisuellen Bewegungsbild



Die Unterscheidung von (mechanischen) Bewegtbildern und (wahrgenommenen) Bewegungsbildern ist auch Gegenstand der Überlegungen Plessners, der bei seinen Überlegungen zu Körperbewegungen als Bewegungsbildern an das kinematographische Bild gedacht haben mag, wie Kappelhoff ausführt:

Denn diese [die Kinematik, CM] bezieht sich auf den geometrischen Raum; das Bewegungsbild hingegen hat keine räumlichen Koordinaten, sondern meint eine zeitliche Struktur, die sich darstellt als das Diagramm eines Werdens, einer Metamorphose, einer sich verändernden Welt. (Kappelhoff 2004a: 153)

Ebenso wie in der Kommunikation von Angesicht zu Angesicht Sprache, Geste, Blick und Mimik zusammenwirken, entstehen aus dem Zusammenspiel von Kameraführung, Montage, Mise-en-Scène und Sound Bewegungsgestalten, die als Bewegungsbilder in ihren jeweiligen Qualitäten, ihrem spezifischen Rhythmus gesehen und gehört, also unmittelbar körperlich wahrgenommen werden. Die Zuschauer*innen der Aufzeichnung von Merkels Rede erfahren deshalb die audiovisuellen Bilder genauso unmittelbar als Ausdruck, wie das Publikum, das im Bundestag die Rede sah und hörte. Allerdings sind erstere abhängig davon, was ihnen die Regie, die die Aufzeichnungen der Bundestagsreden inszeniert, zu sehen gibt. In unserem Fall gab es keinen Schnitt in die Totale oder ins Publikum. Wir sehen Merkel gleichbleibend in halbnaher Einstellung, zentral im Bildraum positioniert und als herausgehobene Person in Szene gesetzt. Die Zwischenrufe, auf die sie mit ihrer kurzen sprachlich-

gestischem Seitensequenz reagiert, bleiben für die Zuschauer*innen an den Bildschirmen im Hintergrund.

3 Ausdrucksbewegung und Interaffektivität

Wie aber ›fühlt‹ sich Merkels kurze Gegenrede für das parlamentarische Publikum, wie für die Zuschauer*innen an ihren Bildschirmen an? Spielt das überhaupt eine Rolle für das Verstehen? Anders gefragt, lässt sich Sprache-Sprechen als Kommunikation von ›Inhalten‹ und als Vollzug von sprachlichen Handlungen treffend und hinreichend beschreiben? Wagen wir einen Blick zurück an den Beginn des zwanzigsten Jahrhunderts. Ausdruck, Ausdrucksbewegung, Ausdruckstheorie waren in dieser Zeit große Themen in Wissenschaft und Kunst. Sprachtheoretische Reflexionen etwa bei Wilhelm Wundt (Wundt 1975 [1900–1925]) oder bei Karl Bühler (Bühler 1933, 1934) waren Teil dieses breitangelegten kulturellen Diskurses, und ihre sprachtheoretischen Positionen sind wesentlich als Beiträge zu diesem Diskurs angelegt.

3.1 Sprache-Sprechen und Gestikulieren sind multifunktional

In kaum einer Einführung in die Sprach- und Kommunikationswissenschaft fehlt Karl Bühlers Organon-Modell der Sprache (Bühler 1982 [1934]). Dort formuliert Bühler, dass Sprachzeichen zugleich Ausdruck, Darstellung und Appell sind, wobei die drei Funktionen in wechselnden Dominanzen im Sprechen auftreten, aber immer ko-präsent sind.

Bühlers postulierte Gleichberechtigung der drei Funktionen wurde mit wenigen Ausnahmen in der zeitgenössischen Sprachwissenschaft nicht systematisch aufgegriffen (vgl. aber Eschbach 1984, Loenhoff 2017).¹¹ Auch, dass das Organon-Modell Baustein einer theoretischen Trias war, die Sprache-Sprechen aus einer Ausdrucks-, einer Sprach-, und einer Handlungstheorie heraus entwickelt, hat wenig bis gar keine Beachtung gefunden. Der Grund, dass Bühlers Arbeiten in ihrer vor dem zweiten Weltkrieg international hochbeachteten Radikalität und Innovationskraft im Nachkriegsdeutschland nicht den verdienten Widerhall gefunden haben, liegt im Abbruch einer deutschsprachigen wissenschaftlichen Tradition, die mit der nationalsozialistischen

11 Vgl. auch die Arbeiten in der Tradition der Deutschen Ausdruckspsychologie (Krumhuber et al. 2013).

Verfolgung deutschsprachiger jüdischer Intellektueller, zu denen auch Karl Bühler gehörte, einherging (vgl. Jäger 2010). Bühler hat die Ausdruckstheorie und die Sprachtheorie als Monographien fertigstellen können (Bühler 1933, 1934). Im Exil konnte er daran nicht mehr anknüpfen. Dabei geht die Beschäftigung mit Ausdruckstheorien der Sprachreflexion voraus. So gelangt er erst über seine Reflektionen zum Verhältnis von Erleben und Ausdruck zur Sprachtheorie (Jäger 2010). Für den Psychologen Bühler ist Ausdruck ohne Sprechen mit Jemandem nicht denkbar. Sprache-Sprechen als Mitteilung an ein Gegenüber ist Bedingung der Möglichkeit einer Verschmelzung von Erleben und Ausdruck:

[...] für Bühler stellt sich die Einheit von seelischem Erlebnis und Ausdruck nur ein, wenn Ausdruck als »Kundgabe« gedacht wird, als eine kommunikative Handlung im Rahmen einer sozialen Situation, in der die Äußerung an mögliche Adressaten gerichtet werden kann. Der Ausdruck würde nämlich – wie Bühler zeigt – in der Einzelseele nicht zu einer festen Verbindung mit der Vorstellung, bzw. einem seelischen Erlebnis gelangen können, weil beide, das seelische Erlebnis und seine gestische oder mimische Äußerung, sich erst auf der Bühne interaktiven sozialen Handelns herstellen. (Jäger 2010: 8)

Mit dieser Position kritisiert Bühler Wundts individualpsychologische Vorstellungen von Ausdruck, die von einer scharfen Trennung innerer Empfindungen und äußerer Ausdrucksbewegungen ausgehen (vgl. Jäger 2010; Wundt 1975 [1900–1925]). Ausdruck ist für Bühler immer an ein soziales Miteinander gebunden, nur als Mitteilung an jemanden theoretisch sinnvoll zu fassen: »Das Bühlersche Subjekt ist [...] ein soziales Subjekt« (Jäger 2010: 7). Damit ist auch die Sprache sozial gedacht, vom Sprechen aus. Bühler fasst sprachliche Zeichen als »konkrete Schallphänomene«, die nur aus einem Sprechereignis sinnvoll zu denken sind. Dieses Sprache-Sprechen ist an drei Koordinaten gebunden: der eine, der andere, die Dinge (Bühler 1982: 26f.). Bühler denkt die Sprache mithin als kommunikatives Ereignis, als Handlung, als Praxis, die immer mehrere Funktionen zugleich erfüllt.¹² Er entwickelt diese Multifunktionalität aus der Trias des Sprechens als kommunikatives Ereignis: »Dreifach ist die Leistung der menschlichen Sprache [...] Ausdruck, Appell und

12 »Mich dünkt, es sei so etwas wie ein Ariadnefaden, der aus allerhand nur halb begriffenen Verwicklungen herausführt, gefunden, wenn man das Sprechen entschlossen als Handlung (und das ist die volle Praxis im Sinne des Aristoteles) bestimmt.« (Bühler 1934: 56–57)

Darstellung [...].« (Bühler 1982: 28). Die Sprachfunktionen überlagern sich, sie sind gleichzeitig präsent, im konkreten Moment des Sprechens steht jedoch stets die eine oder andere im Vordergrund, ist eine oder die andere dominant (Bühler 1982: 30f; Müller 1998: 20, 90). Meist ist dies die Darstellungsfunktion, der Bezug auf die Welt, der allerdings nur in der Gerichtetheit auf ein Gegenüber zu denken ist und der eben immer auch Ausdruck ist.

Merkels Äußerung zu den steigenden Lohnnebenkosten macht nur Sinn als Antwort auf einen kritischen Zwischenruf aus den Reihen der Parlamentarier. Auf diesen reagiert sie, wenn sie sagt: »Und Sie haben keine Antwort auf die Frage, was man denn tut, um die immer steigenden Lohnnebenkosten zu vermeiden.« Diese Entgegnung ist zugleich ärgerlicher und kämpferischer *Ausdruck*, wie er mit der Zurückweisung des Zwischenrufs als kommunikative Handlung (*Appell* in Bühlers Terminologie) einhergeht und dies mit einer Sachfrage verbindet (*Darstellung* im Sinne Bühlers). Eine vergleichbare Multifunktionalität sieht Bühler auch in den Gesten von Sprecher*innen realisiert.

[...] warum sollen z.B. gestenhaft malende Hände nicht imstande sein, das eine Mal frisch und froh, ein zweites Mal wütend, ein drittes Mal zögernd und erschreckt den Akt des Zeichnens durchzuführen? Und wenn sie es vermögen, so stellen sie dar und drücken aus in eins. (Bühler 1933: 39)

Mit dieser Beobachtung ist Bühler seiner Zeit weit voraus. Er beschreibt hier nämlich nicht mehr und nicht weniger das Potential »gestenhaft malende[r] Hände«, (Gebärden-)Sprache zu werden. Auch Handbewegungen können darstellen und ausdrücken zugleich, und diese sind in der Regel auf ein Gegenüber ausgerichtet.¹³

Wenn nun im Sprache-Sprechen gestische und lautsprachliche Äußerungsformen zusammenkommen, können sie synchron unterschiedliche Bedeutungsfacetten artikulieren. So verbindet Merkel etwa in ihrer Gegenrede eine (für ihre Verhältnisse) ärgerlich-aggressive sprachliche Reaktion mit einer Reihe kleiner, sanft ausgeführter, bogenförmiger Gesten, die hochartikuliert, sehr präzise die Stufen einer Treppe in die Luft zeichnen. Die Gegenrede ist in ihrem affektiven Ausdruck ärgerlich, sanft und präzise zugleich.

13 Zur funktionalen Parallelität von Sprache und Geste und zum Sprachpotential von Gesten vgl. Müller (1998: 87f.).

3.2 Ausdruck und Verstehen

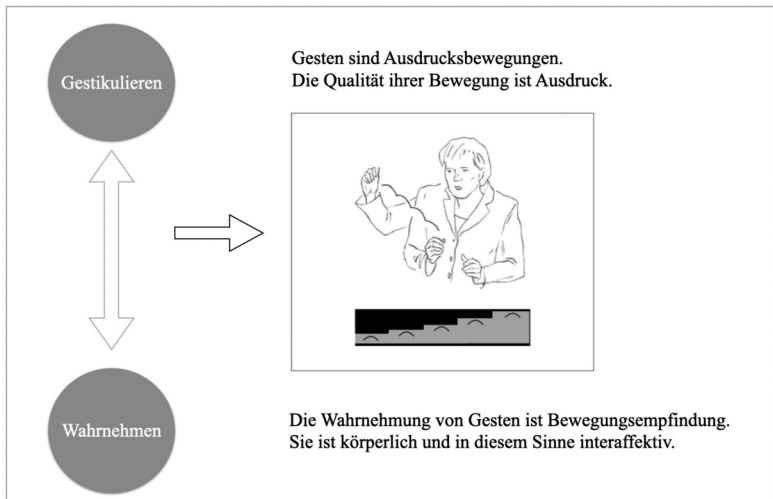
Wir halten mit Bühler fest, Sprache-Sprechen einschließlich der mit dem Sprechen einhergehenden Gesten ist zugleich Ausdruck, Darstellung, Appell. Oder anders formuliert, Sprache-Sprechen ist immer auch Affektausdruck, eingebunden in soziales Miteinander. Für Bühler ist Ausdruck als Verbindung von Empfindung und äußerer Form nur denkbar als kommunikatives Ereignis (Jäger 2010). Nimmt man diesen Punkt ernst, dann stellt sich die Frage nach dem Verstehen. Wie kommt Sprache-Sprechen als ›Ausdruck‹ bei den Zuschauer*innen im Parlament und an den Bildschirmen an? Mit Bezug auf Merleau-Pontys Phänomenologie der Wahrnehmung lautet die Antwort: Das Verstehen ist ein leibliches, ein körperliches. Die Parlamentarier*innen empfinden die Ausdrucksqualität von Merkels Geste unmittelbar körperlich. Sie müssen sie nicht erst als Zeichen dekodieren, das für eine Innerlichkeit steht. Mit Merleau-Ponty ließe sich sagen: Die sanften und zugleich präzise ausgeführten gestischen Bewegungen *sind* Sanftheit, *sind* Präzision. Der ärgerliche Ausdruck der Stimme transportiert nicht einen inneren Zustand nach außen, er *ist* der Ärger. Es gibt weder Sanftheit, noch Präzision, noch Ärger, die als Affekte dem körperlichen Ausdruck vorgängig wären.

Faced with an angry or threatening gesture, I have no need, in order to understand it, to recall the feelings which I myself experienced when I used these gestures on my own account. [...] I do not see anger or a threatening attitude as a psychic fact hidden behind the gesture, I read anger in it. The gesture *does not make me think* of anger, it is anger itself (Merleau-Ponty 2005 [1945]: 214)

Das affektive Verstehen ist also ein leibliches, ein unmittelbar körperliches. Die Körperbewegung ist Affekt, sei es die Tonhöhenbewegung der Stimme, seien es die gestischen Bewegungen (s. Abb. 4).

Betrachtet man Sprache-Sprechen als Ausdrucksbewegung, dann sieht man nicht primär einen Austausch von Gedanken, sondern ein zuallererst körperlich vollzogenes Miteinander, einen interkorporalen Dialog. In diesem entstehen und zirkulieren Affekte, das Sprache-Spreche ist in diesem Sinne grundsätzlich ›interaffektiv‹ (Horst et al. 2014). Fuchs und De Jaegher (2009: 479) bringen diese phänomenologische Position auf den Punkt: »affects are not enclosed in an inner mental sphere to be deciphered from outside but come into existence, change and circulate between self and other in the intercorporeal dialogue«.

Abb. 4: Gesten als Ausdrucksbewegungen



3.3 Multidimensionale Ausdrucksgestalten

Was Zuschauer*innen ebenso wie Teilnehmer*innen an einem Gespräch wahrnehmen, sind dabei nicht aufaddierte Informationen, die sich einzeln in den verschiedenen Ausdrucksmodalitäten, also in Geste, Wort, Mimik, oder als Zoom auf das Gesicht einer Redner*in oder durch einen Schnitt ins Publikum realisieren und die dann gewissermaßen durch bloße Summierung Sinn ergeben. Was wir wahrnehmen, wenn wir einen Film schauen, oder wenn wir unsere Gesprächspartner*innen anblicken, sind multidimensionale Erfahrungsgestalten, die sich im Prozess des Zuhörens, Zusehens, des Hörens und Betrachtens audiovisueller Bilder als fluide aber gestalthafte Figurationen herausbilden (Müller/Kappelhoff 2018). Was auch immer Merkel in dem kleinen Ausschnitt sagt, die sanfte und gleichzeitig präzise Bewegung der Geste, ihre ganze gefasste Körperhaltung setzen einen bedachten, konzentrierten Grundton oder, anders formuliert, sie grundieren das Gesagte affektiv, geben ihm eine besondere affektive Tönung. Sie sind Teil einer Ausdrucksbewegung, die im Sprache-Sprechen entsteht und die sich gleich einer Handgeste als zeitliche Bewegungsfiguration mit Beginn, Höhepunkt und einem Abschluss entfaltet.

Multidimensionale Erfahrungsgestalten als Ausdrucksbewegungen betreffen sowohl die Multimodalität des Sprechens als auch audiovisuelle Bewegungsbilder. Die Bedeutungen stecken in den Bewegungen selbst und sind nicht dahinter verborgene, mittelbare Konzepte. Sprachlich-gestische und filmische Ausdrucksbewegungen werden als multidimensionale Erfahrungsgestalten leiblich, also unmittelbar, von den Zuschauer*innen erfahren und so verstanden. So schreibt Merleau-Ponty zur Wahrnehmung:

My perception is therefore not a sum of visual, tactile, and audible givens. Rather, I perceive in a total way with my whole being; I grasp a unique structure of the thing, a unique way of being, which speaks to all my senses at once. (Merleau-Ponty 1968: 48)

Der Körper der Sprachforscher*innen dient in der Analyse gewissermaßen als ›Instrument‹, die Beschreibungen der empfundenen Affektqualitäten dieser gestalthaften Bewegungsfigurationen sind intersubjektiv überprüfbar.¹⁴

3.4 Filmische und gestische Ausdrucksbewegung

Das Konzept der Ausdrucksbewegung wurde in den ersten Dekaden des zwanzigsten Jahrhunderts in Kunst und Wissenschaft lebhaft und in großer Breite, heute würde man vielleicht sagen ›transdisziplinär‹, diskutiert (Kappelhoff 2004a, 2018; Müller/Kappelhoff 2018). Der Bogen reicht von der Philosophischen Anthropologie, Psychologie, Sprachwissenschaft, bis zur Filmtheorie, Phänomenologie und Philosophischen Ästhetik. Für die Sprachforscher und Psychologen Bühler und Wundt, ebenso für Plessner in seiner philosophischen Anthropologie war es absolut selbstverständlich, die Geste als Ausdrucksbewegung zu fassen. Wie diese theoretisch zu greifen war, stand zur Debatte. Gleichzeitig war für die frühe Filmtheorie das ›Gestische‹ am Film die filmische Ausdrucksbewegung ein wichtiger Gedanke (Balázs 1924; Eisenstein 2010 [1926]; Münsterberg 2002 [1916]). Mit dem Konzept der Ausdrucksbewegung betonten sie (mit Ausnahme Wundts) die Verschränkung von Ausdruck und Erfahrung im menschlichen Verhalten. Diese Verschränkung beschreibt die wechselseitige Affizierung der Körper, sie zielt auf die Unmittelbarkeit der Bewegungserfahrung als verkörpertes, als affektives Empfinden.

14 Zur methodischen Umsetzung vgl. Kindler-Mathôt et al. i. d. Bd.

Überlegungen zur Parallelität von Film und Geste, zu gestischem und filmischem Bewegungsbild und zu ihrem Ausdruckscharakter stehen am Beginn der Filmtheorie. Wenn etwa Balázs (1924) im Ausdruck des Gesichts das Paradigma filmischer Bilder ausmacht, dann bezieht sich dies auf die Zeitlichkeit und die Unmittelbarkeit der Empfindung, die sich mit dem Prozess der Wahrnehmung von Bewegungsbildern verbindet (Kappelhoff 2004a: 154, 2004b). Diese Position steht im Gegensatz zu Ekmans (1972, 2006) Fotografien von Gesichtern, die den Ausdruck universeller Emotionen darstellen sollen. Während Ekmans photographisch festgehaltene Gesichtsausdrücke insinuieren, dass sich ein Gefühl im Gesicht als statischer Ausdruck zeigt, geht Kappelhoff davon aus, dass sich Affekte über die Zeit als sanfte, langsam gemächliche oder explosive, hart konturierte mimische Bewegungen entfalten. Was für das Gesicht gilt, gilt auch für audiovisuelle Bilder. Audiovisuelle Affekte zeigen sich nicht primär in den Gesichtern der Schauspieler*innen, sondern im Rhythmus, in der Dynamik audiovisuellen Bilder.

Die Leinwand ist das Gesicht des Kinos. Das ist, worum es Balázs geht.

Was Balázs an der Bewegung eines Gesichts zu greifen sucht, ist weder der Schein des lebendigen Gesichts noch die Repräsentation des »ganzen Menschen«, sondern die affektive Dimension des Kinos: ein Raum der artifiziellen Affekte, der technisch animierten, künstlichen Seele. (Kappelhoff 2004a: 155)

In ihrer Bewegungsqualität sind Bewegungsbilder Ausdruck. Dies gilt für den mimischen Ausdruck ebenso wie für das Filmbild. Beide sind nicht Repräsentationen ›anderer Zustände‹, ihre affektive Dimension ist die Qualität der Bewegung. Sie sind unmittelbar körperlich erfahrene Ausdrucksbewegungen. Greifenstein (2020: 48) beschreibt deshalb das Verhältnis von Ausdruck und Affekt als nicht arbiträr. Ausdrucksbewegungen begegnen uns als ganzheitliche Gestalten. So argumentiert etwa Plessner am Beispiel einer ängstlich davonlaufenden Katze oder der freudigen Begrüßung eines Hundes, dass die affektive Dimension des Verhaltens in der Qualität der Ganzheit der Bewegung liegt und gerade nicht in der objektiven Beschreibung der Bewegungen.

Daß die Katze mir davonläuft, läßt sich auch objektiv in bloßen Bewegungen darstellen; daß sie mich ängstlich flieht, ist mir überdies in diesen Bewegungen als einheitlicher Charakter mitgegeben. Daß der Hund an mir empor-

springt, ist objektiv konstatierbar; daß er mich freudig begrüßt, ist mir in seinem Gebaren als Richtungsform deutlich. (Plessner 1982 [1925]: 82)

In der Filmtheorie beschreibt das Konzept der Ausdrucksbewegung eine Bewegungsfiguration, die sich als Gestalt im Prozess einer Wahrnehmung herausbildet und deren Ausdrucksqualität als Bewegungsqualität unmittelbar körperlich empfunden wird. Kappelhoff fasst dies als filmische Ausdrucksbewegung (2004a): Spannung, Horror, Trauer, Melancholie, Sehnsucht, Freude vermitteln sich als Bewegungserfahrung, als ästhetische Lust unmittelbar körperlich. Dieser Typus von Erfahrung stellt sich nicht erst ein, nachdem die Puzzlesteinchen der verschiedenen Modalitäten (Dialog, Bildraumgestaltung, Kamerafahrt, Schnittfolge, Musik etc.) kognitiv rekonstruiert und auf eine dargestellte Figur projiziert wurden, der man als Zuschauer*in dann empathisch folgt. Vielmehr handelt es sich um eine unmittelbar empfundene, körperliche Erfahrung von Bewegungsqualitäten, genau in dem Sinne, wie Stern dies in seinen Beobachtungen zur Mutter-Kind-Interaktion als Miteinander-Tanzen zweier Körper beschrieben hat (Stern 1985). Die Leinwand, der Bildschirm tritt an die Stelle des Körpers – sie ist gewissermaßen ein artifizieller Körper (man denke an das von Balázs angeführte Gesicht als Paradigma des Kinos), der das Publikum tatsächlich bewegt – kollektiv erschrecktes Aufschreien oder Tränen der Rührung, die wir im Kino gemeinsam vergießen, zeugen von dieser körperlichen Unmittelbarkeit. Kappelhoff (2018: 129) formuliert den Grundgedanken des Konzepts der Ausdrucksbewegung als: »das Verhältnis zwischen affizierendem und affiziertem Körper«. In diesem Sinne haben wir es sowohl im zwischenmenschlichen Gespräch als auch im Kino mit einer grundsätzlich körperlichen Interaktion zu tun, die interaffektiv ist.

4 Interkorporalität und ästhetische Erfahrung

Wir halten fest, dass Ausdrucksbewegungen, ob audiovisuell oder sprachlich-gestisch, in ihrer Gestalthaftigkeit durch eine je spezifische Bewegungsdynamik zusammengehalten werden. Diese lässt sich in ihren Qualitäten beschreiben: Entfalten sie sich langsam, vorsichtig oder abrupt ansteigend und explosiv, fallen sie danach sanft ab oder hält sich eine Spannung über einen längeren Zeitraum? Immer zeichnen sie sich durch das spezifische Profil und die spezifische Qualität ihrer Bewegungsabläufe aus. Der Ausdruck liegt in der Bewegung selbst, sie ist es, die unmittelbar die Betrachter*in-

nen berührt, mitnimmt, abstößt, in Spannung hält oder diese auflöst. Die Wahrnehmung dieser Bewegung ist körperlich, unmittelbar, ganz gleich ob es sich um die Bewegungen von Gesprächspartner*innen oder um filmische Bilder handelt: Sie ist interkorporal, zwischenleiblich (Fuchs 2017). Gemeint ist einmal mehr die Unmittelbarkeit körperlicher Bewegungserfahrung die, so unsere (linguistisch-filmwissenschaftliche) These, sowohl die Multimodalität des Sprechens als auch die audiovisuelle Multimodalität charakterisiert.

Das bedeutet, wann immer uns ›Sprache-Sprechen‹ in audiovisuellen Medien begegnet, dekodieren wir als Zuschauer*innen und Zuhörer*innen nicht einfach den Inhalt einer sprachlich formulierten Botschaft, vielmehr sehen wir, wie der/die Sprecher*in sich bewegt, ob er/sie recht steif dasteht, unruhig auf- und abschreitet, ob und wie er/sie gestikuliert, wir nehmen wahr, ob gestische Bewegungen im Einklang mit dem Sprechrhythmus stehen oder nicht und mit welcher Dynamik und Qualität sie sich entfalten. Falls noch weitere Personen zu sehen sind, nehmen wir wahr, wie sich das Sprechen und Bewegen der einen mit den Körpern der anderen im Bildraum Anwesenden verschränkt und welche Qualität dieses Zusammenspiel der Körper hat, ob sie harmonisch bestätigend miteinander-schwingen oder sich in harscher Opposition konfrontieren. Die ästhetische Erfahrung dieser audiovisuellen Bilder ist durch und durch körperlich, sie ist Hören, Sehen, Fühlen.

Was wir als Zuschauer*innen an den Bildschirmen von diesen Interaktionen zu sehen bekommen, ist jedoch ganz und gar abhängig von der Gestaltung der audiovisuellen Bildräume. Kameraperspektiven und Schnittfolgen entscheiden noch im neutralen Modus parlamentarischer Dokumentation von Bundestagsdebatten, was genau die demokratische Öffentlichkeit von den Debatten zu sehen bekommt. Gleiches gilt in gewisser Weise auch für die Gestaltung von ethnographischen oder experimentellen Settings, die die Multimodalität des Sprechens in mehr oder minder natürlichen Kontexten untersuchen. Auch hier entscheidet die Position einer oder mehrerer Kameras darüber, was hinterher als ›Datum‹, als ›Material‹ den jeweiligen Analysen zugrunde liegt. Seit Adam Kendons frühen Analysen zur Synchronisation von Sprechen und Körperbewegung ist die Filmkamera zum Mikroskop der Gestenforschung geworden (Kendon 1972). Nun ist es ganz offensichtlich, dass ein Mikroskop eine sehr spezifische Perspektive auf das Untersuchungsobjekt eröffnet und andere Aspekte dafür in den Hintergrund treten. Gestenforscher*innen kennen die Artefakte, die durch das Abspielen von Körperbewegungen in Zeitlupen entstehen, sie sind vertraut mit den Problemen, die durch die extrem hohe Auflösung digitaler Bilder im HD-Format für die Segmentie-

rung von Körperbewegungen entstehen. Wo beim Videobild der achtziger und neunziger Jahre Bewegungs- und Ruhephasen durch unscharfe und scharfe Videobilder zu erkennen waren (Seyfeddinipur 2006), zeigt die heutige digitale Aufnahmequalität eines jeden Mobiltelefons Bewegungsabläufe in einer kristallklaren Bildschärfe, die früher nur bei spektakulären Naturaufnahmen oder bei Aufnahmen von Sportwettkämpfen zu finden waren.

Das bedeutet, dass nicht nur für die Rezeption von Filmkunst gilt, dass die Wahrnehmung der audiovisuellen Bilder zuallererst eine körperliche ist, die Position der Zuschauer*in, ob als Forschende, im Kinosaal oder am Computerbildschirm ist daher entscheidend. Ohne Zuschauer*in kein Film, ohne Gegenüber keine Geste, kein Sprache-Sprechen.

4.1 Die Perspektive der Zuschauer*innen und das ästhetische Erleben

Die Analyse der Multimodalität des Sprechens in audiovisuellen Medien lässt sich nicht ablösen von der Wahrnehmungsposition der Zuschauer*innen. Sie lässt sich auch nicht separieren vom Charakter des ästhetischen Erlebens als einer zeitlich gestalteten Bewegungserfahrung, die die Zuschauer*innen mitnimmt, die sie affektiv in die Entfaltung filmischer Bilder einbindet. Nur so lässt sich die »Position eines verstehenden Subjekts« angeben (Kappelhoff 2018: 144), nur so lässt sich überhaupt rekonstruieren, wie audiovisuelle Bilder wirken, wie sie Zuschauer*innen berühren, wie sie ihre affektive Kraft entfalten:

Erst durch die Rekonstruktion der affektiven Verwicklung des Rezipienten, die Rekonstruktion der Position, aus der heraus sich das ›Gefühl für das Ganze der Situation‹ erschließt, lässt sich überhaupt die Position eines verstehenden Subjektes angeben. Ist dieses doch unauflösbar in seinen körperlichen Empfindungen in den Prozess der Entfaltung des filmischen Bildes verstrickt. (Kappelhoff 2018: 144)

Wie aber lässt sich dieses Empfinden der Zuschauer*innen analytisch einholen, ohne dabei auf Rezipient*innenbefragungen oder ähnliche sozialwissenschaftliche Verfahren zurückzugreifen? Die Position der phänomenologisch inspirierten Filmwissenschaft geht den Weg über eine detaillierte Analyse der audiovisuellen Komposition als einer Zeitform, die ein Wahrnehmungserleben auf der Seite der Zuschauer*innen als zeitliche Erfahrung moduliert. Kappelhoff formuliert diese analytische Perspektive wie folgt:

Deshalb muss die Analyse audiovisueller Bilder immer den Weg über die Rekonstruktion des ästhetischen Erlebens gehen – das meint die Rekonstruktion der je besonderen Wahrnehmungsform des audiovisuellen Bewegungsbildes (d.h. seine Performanz, also die zeitliche Struktur der Entfaltung des Bewegungsbildes in der Rezeption durch den Bildkonsumenten) und seiner inszenatorischen Komposition (d.h. der zeitlichen Figur, als welche die zeitliche Struktur im Prozess der Rezeption zur bedeutungsvollen Gestalt eines filmischen Bewegungsbildes wird). (Kappelhoff 2018: 144)

Methodologisch gilt es in einer medienästhetisch begründeten Perspektive durch eine präzise Analyse filmischer Bewegungsbilder, das verkörperte Erleben der Zuschauer*innen vor den Leinwänden und an den Bildschirmen zu rekonstruieren. In der Wahrnehmung der Zuschauer*innen im Kino, vor dem Fernseher oder vor dem Laptop entfalten sich die audiovisuellen Aufnahmen als Ganzes: Sie ergeben eine zeitliche Gestalt, die nicht im Prozess des Schauens von den Zuschauer*innen in ihre Einzelteile zerlegt wird, sondern die Stimmliches, Gestisches, Kamerabewegung, Bildraumgestaltung, Montage in sich vereint. Damit, so die These, verhält es sich mit der Position der Zuschauer*innen audiovisueller Formate genauso, wie mit der von Teilnehmer*innen an einem Gespräch. Körperliche Ausdrucksbewegungen sind als »Bewegungsgestalten bildhaft, [...], dem Beobachter gegenwärtig« (Plessner 1982 [1925]: 77–78).

Die Perspektive der Zuschauer*innen für die Multimodalität des Sprechens in audiovisuellen Medien ernst zu nehmen, heißt daher, die sich entfaltende Wahrnehmungsgestalt als analytischen Ausgangspunkt anzunehmen. Diese Wahrnehmungsgestalt kann in der weiteren Analyse auf ihre Orchestrierung hin, d.h. mit Bezug auf sprachliche und gestische Figurationen, Bildraum, Kameraeinstellung, Montage etc. beschrieben werden.¹⁵

15 Die einzelnen Analyseschritte dieses Prozesses werden im Beitrag von Kindler-Mathôt et al. (i. d. Bd.) erläutert. Einen Überblick über das gesamte methodische Instrumentarium sowohl für die Multimodalität des Sprechens als auch für die audiovisuelle Multimodalität findet sich im Appendix bei Müller/Kappelhoff (2018).

Abb. 5: Filmische Bewegungsbilder als Ausdrucksbewegung



Kehren wir ein letztes Mal zum Ausschnitt aus der Rede Merkels im Bundestag zurück (Abb. 5). Betrachten wir sie als kompositorische Gestalt, dann zeichnet sich diese durch eine zurückhaltende neutrale Inszenierung aus. Wir sehen eine stabile Kameraposition, die Merkel in der Halbtotalen zeigt, es gibt keinen Schnitt ins Publikum, keine Kamerabewegung, keine Veränderung der Bildraumgestaltung. Merkel wird in einer zentralen Position des Bildraumes gezeigt. In rotem Blazer vor grauem Hintergrund sind ihre Bewegungen gut sichtbar. Die audiovisuelle Inszenierung stellt Merkel in den Mittelpunkt und verleiht ihr durch die Statik Stabilität und den Eindruck von seriöser Neutralität. Vor dieser visuell zurückgenommenen Inszenierung heben sich die sprachliche und gestische Äußerung markant ab. Sie entfalten über die lautliche, gestische Dynamik und die markierte Hinwendung von Merkels Oberkörper zum Publikum eine affektive Dynamik, welche sich gemeinsam mit den kritischen Zwischenrufen aus den Reihen der Parlamentarier*innen als eine plötzlich ansteigende, auffallende Bewegung entfaltet, die im Verlauf von Merkels Gegenrede kontinuierlich heruntergefahren wird. D.h. am Ende der Entgegnung Merkels auf den Zwischenruf ist die Gestalt geschlossen, die affektive Dynamik wieder auf dem Niveau, das sie zuvor hatte.

4.2 Die Poiesis des Filme-Sehens

Audiovisuelle Bilder formen das ästhetische Erleben der Rezipient*innen als spezifische Form ästhetischer Erfahrung. Das ›Verstehen‹ audiovisueller Bil-

der, das Verstehen von Sprache-Sprechen in audiovisuellen Medien ist daher affektiv durchtränkt. Audiovisuelle Bilder modulieren die Erfahrungsräume ihrer Betrachter*innen, gleichzeitig entstehen diese als Bewegungsbilder, als Bewegungsfigurationen, als multidimensionale Gestalten erst in der Erfahrung, im Betrachten der Zuschauenden. Erst im Erleben der Betrachter*innen entstehen die Bewegungsbilder als Wahrnehmungsgestalten (vgl. Abb. 5). Das Sehen und Hören audiovisueller Bilder ist daher ein Tun, eine Produktion der Zuschauenden. Kappelhoff fasst diese »rezeptive Aneignung audiovisueller Bewegungsbilder« als kulturelle Praxis, als eine Poiesis des Filme-Sehens:

Damit wird die Rezeption selbst als eine Form der Produktion, die rezeptive Aneignung audiovisueller Bewegungsbilder als ein genuiner Akt des Herstellens avisiert. Wir sprechen deshalb von der Rezeption audiovisueller Bilder als einer Poiesis des Filme-Sehens. (Kappelhoff 2018: 10)

Das Sehen und Hören audiovisueller Bilder gerät damit als eine Form des Machens in den Blick. Dieses Sehen und Hören wird zwar durch die Gestaltung der audiovisuellen Formate orchestriert, als Bewegungsbilder benötigen sie jedoch die Körper der Zuschauenden. Um Sobchack zu paraphrasieren: Die Zuschauenden »leihen« den Bildern ihre Körper (Sobchack 1992). Das Sehen und Hören audiovisueller Bilder ist darum kein passives Überflutet-Werden, sondern eine mediale Praxis die, wie u. a. der Siegeszug der sozialen Medien zeigt, ein äußerst erfolgreiches Instrument der Gemeinschaftsbildung ist. Von der Wirkmächtigkeit dieses Medienkonsums als kultureller Praxis zeugen etwa die immensen *Follower*innen*-Zahlen von TikTok-Videos der neuen Rechten oder von Schminkanleitungen, die mit Politinformationen oder True-Crime Stories verbunden werden.

In diesem Sinne lässt sich die Poiesis des Filme-Sehens nicht als passive Rezeption audiovisueller Bilder sondern als eine kulturelle Praxis fassen, die sich als geteilte ästhetische Erfahrung in zahllosen Geschmacksgemeinschaften manifestiert – ob dies die Anhänger*innen trashiger Horrorfilme oder Netflix-User, die Fans von Kochvideos, Polit-Talks, Serien, Videospielen, Fußballspielen oder Comedies oder die Millionen *Follower*innen* von Videoclips in den sozialen Medien sind. Es ist diese Praxis des Medienkonsums, die zu völlig neuen transnationalen Gemeinschaftsbildungen führt, wie sie sich nicht nur an globalen *Follower*innen*-Strömen bestimmter *Influencer*innen* nachvollziehen lässt. In den Poetologien der sich ständig verändernden medialen Formen zeigt sich eine kulturelle Praxis, die sich über die präzise Rekonstruktion der

audiovisuellen Orchestrierung als Raum spezifischer ästhetischer Erfahrung rekonstruieren und damit auch kritisch befragen lässt.¹⁶

4.3 Generierung, Modulation und Mobilisierung von Zuschauergefühlen

Mit dem Rekurs auf das Konzept der Ausdrucksbewegung lässt sich eine Medienpraxis beschreiben, die Affekte und Gefühle¹⁷ einer breiten Zuschauer*innenschaft generiert, moduliert und mitunter auch sehr gezielt mobilisiert (Kappelhoff 2008, 2016). Die Poetik der Ausdrucksbewegung betrifft dabei einen grundlegenden Darstellungsmodus westlicher Kunst, der eng verbunden ist mit der Kulturgeschichte des Theaters der Empfindsamkeit und des Melodramas im Film (Kappelhoff 2004a: 182, 2014; Müller/Kappelhoff 2018: 62f.). Die klassische Filmtheorie sieht in der Ausdrucksbewegung eine Form der Poetik, die die Generierung und Modulation von Affekten eines breiten anonymen Publikums erlaubt und die darin eine Form kollektiver Subjektivität sieht. Die Erfahrung intensiver Empfindungen – Horror, Spannung, sentimentale Gefühle – ist in der zeitgenössischen Unterhaltungskultur zum Selbstzweck geworden (Müller/Kappelhoff 2018: 165). So hat die Generierung und Modulation von Gefühlen längst auch alltägliche Formen medialer Gestaltung erreicht, seien es Nachrichtenformate, Reportagen oder die Dokumentation parlamentarischer Debatten. Transdisziplinäre Analysen audiovisueller Multimodalität haben zeigen können, dass in ganz verschiedenen medialen Formaten audiovisuelle Ausdrucksbewegungen Bewegungsbilder formen, die als multidimensionale Erfahrungsgestalten den Wahrnehmungsprozess auch noch so »objektiv« sich präsentierender medialer Formate affektiv grundieren. Die Wirkmacht dieser audiovisuellen Bilder liegt dann in der unmerklichen Erzeugung und Veränderung des affektiven Erlebens der Zuschauer*innen. Was gezeigt wird, entsteht aus dem Wie der audiovisuellen Inszenierung.

Bezogen auf die mediale Inszenierung von Bundestagsreden durch ein Parlamentsfernsehen, wird hier durch eine eher zurückhaltende Inszenie-

16 Gegenstand der Kolleg-Forschungsgruppe »Cinepoetics« an der Freien Universität Berlin waren solche Formen der Poiesis des Filme-Sehens, die über die Rekonstruktion von Poetologien audiovisueller Bilder greifbar wurden: <https://www.cinepoetics.fu-berlin.de/>. Viele dieser Studien sind in der Buchreihe *Cinepoetics* bei De Gruyter im Open Access Format erschienen.

17 In Anlehnung an Damasio (1999) unterscheide ich terminologisch Affekte (»Feeling«) und Gefühle (»Emotion«), wobei sich *Feeling* auf Stimmungen, Atmosphären und *Emotion* auf diskrete Gefühle bezieht (vgl. hierzu auch Müller/Kappelhoff 2018: 55).

rung (statische Einstellung, keine oder wenig Schnitte, keine oder wenige Kamerabewegung) eine Art ›objektiver‹ neutraler Perspektive hergestellt, die im scharfen Gegensatz zu Videoclips auf Plattformen wie TikTok oder Instagram stehen, die mit zusätzlichen visuellen Effekten, Spracheinblendungen, überblendeten Bildeffekten, einer Zerstückelung und Beschleunigung von gesprochener Sprache durch hohe Schnittfrequenzen und Erhöhung der Bildgeschwindigkeit arbeiten – und die sich als sehr kurze Zeitformen (TikTok: 5 Sekunden bis 10 Minuten, Instagram: 3–90 Sekunden) etabliert haben, während die Dokumentation politischer Reden selten kürzer als fünf Minuten ist. Wie sich diese Beschleunigung, Fragmentierung von Sprache-Sprechen auf die Generierung, Modulation und Mobilisierung von Zuschauergefühlen auswirkt bleibt ein drängendes Desiderat kritischer kulturwissenschaftlicher Sprachgebrauchs- und Medienforschung.

5 Zusammenfassung

Audiovisuelle Medien in verschiedensten Ausprägungen strukturieren in stetig steigendem Maße den Gebrauch und die Wahrnehmung von Sprache. Da die Beobachtung sprachlicher Multimodalität eng mit der Möglichkeit der audiovisuellen Fixierung flüchtiger Momente des Sprache-Sprechens verbunden ist, ist der analytische Blick auf sprachliche Multimodalität immer auch ein Blick durch ein audiovisuelles Medium. Dies betrifft offensichtlich nicht nur die Videokameras der Forschung, sondern den sprachlichen Alltag unserer Gesellschaft. Hier hat der vorliegende Beitrag angesetzt und für eine medienästhetische Perspektive auf den Sprachgebrauch in audiovisuellen Medien argumentiert.

Eine medienästhetische Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien einzunehmen, zielt auf eine Beschreibung von Praktiken des Medienkonsums und ihrer Attraktivität ab. Sie verbindet Kappelhoffs Theorie filmischen Denkens, »Die Poiesis des Filme-Sehens«, mit einer Theorie der Multimodalität des Sprechens. Für beide sind Zeitlichkeit, leibliche Verschränkung, Interkorporalität und Interaffektivität als Bewegungserfahrung grundlegend:

Interactive movement as reflexive affection is a key element from which the gestural as well as the cinematic gestalten ultimately emanate. Conceiving of multimodal face-to-face communication as temporally orchestrated and

as inherently interaffective is what connects gestures and spoken language with film. (Müller/Kappelhoff 2018: 155)

Mit anderen Worten, »Sprache-Sprechen, Hören, Sehen, Fühlen, – Erfahren« formuliert eine transdisziplinäre Perspektive, deren theoretischer und methodologischer Angelpunkt das Konzept der körperlich-filmischen Ausdrucksbewegung ist. Es fungiert als theoretischer und methodologischer Bezugspunkt für die Analyse audiovisueller und sprachlicher Multimodalität. Über die Gestaltung audiovisueller Bilder als Ausdrucksbewegungen formen diese eine ästhetische Erfahrung der Zuschauer*innen, die diesem Diskurs audiovisueller Bilder nicht etwa passiv ausgeliefert sind, sondern in deren Wahrnehmung diese sich erst als Bewegungsbilder entfalten. Die Perspektive der Zuschauer*innen für die Multimodalität des Sprechens in audiovisuellen Medien ernst zu nehmen, heißt dann, die Wahrnehmungsgestalt der filmischen Ausdrucksbewegung als analytischen Ausgangspunkt anzunehmen.¹⁸ Diese multidimensionalen Wahrnehmungsgestalten können in der weiteren Analyse auf ihre Orchestrierung hin, d.h. mit Bezug auf sprachliche und gestische Figurationen, Bildraum, Kameraeinstellung, Montage etc. beschrieben werden.

Die Position der Zuschauer*innen einzunehmen hat auch erhebliche Folgen für die ethnographische und experimentelle Erforschung multimodaler Kommunikation und Interaktion, denn es impliziert die Forderung, die Position der Forschenden, der Analytiker*innen an den Bildschirmen systematisch mit in den analytischen Prozess einzubeziehen. Aus medienwissenschaftlicher Sicht gilt es dabei einmal mehr zu beachten, dass »Medien [...] keineswegs bloße Mittel der Kommunikation gegebener Sachverhalte [sind]; vielmehr sind sie technische Erweiterungen menschlicher Wahrnehmung; sie verändern die apriorischen Bedingungen des Verstehens, Urteilens und Imaginierens« (Kappelhoff 2018: 9). Dass dies für die Kamera als analytisches Werkzeug für das Studium multimodaler Interaktionen genauso gilt wie für die Bilder von Handyskameras, die auf *Social Media* gepostet werden, versteht sich von selbst.

Im Zentrum einer medienästhetischen Perspektive auf den Sprachgebrauch in audiovisuellen Medien steht die systematische Unterscheidung

18 »In our research on the poeology of audiovisual images we have developed an analytical model that is comprehensive and cultural-historically oriented: expressive movement (vgl. Kappelhoff 2004a). It captures the affective movement of the viewer as a bodily sensation of the unfolding flow of cinematic movement-images« (Müller/Kappelhoff 2018: 59).

der Multimodalität des Sprechens und der audiovisuellen Multimodalität. Erst diese Unterscheidung ermöglicht es, ihre Gemeinsamkeiten und Verschränkungen als Zeitbilder, Bewegungsbilder sowie den geteilten Modus ästhetischer Erfahrung als Verschränkung von Körpern im Rhythmus und Dynamik der Bewegungserfahrung herauszuarbeiten. Wechselseitiges Verstehen, Intersubjektivität speist sich zuallererst aus der Zwischenleiblichkeit, dem körperlichen Miteinander, das in der Qualität der Bewegung, ob Handgeste oder audiovisuelles Bild, immer schon eine affektive Tönung hat.

In diesem Sinne ist Sprache-Sprechen grundiert in einer affektiven Bewegungserfahrung von Körpern, deren Sprechen Körperbewegung *ist* (Horst et al. 2014). Auch den Konsum audiovisueller Medien kennzeichnet ein Modus ästhetischer Erfahrung, für den die Interkorporalität als interaffektive Erfahrung von Bewegungsbildern grundlegend ist. Als kulturelle Praxis wird dieser Erfahrungsraum durch vielfältige mediale Ökologien gestaltet, die jeweils zu partikulären Gemeinschaftsbildungen auf der Seite der Medienkonsument*innen führen.

Denkt man also den Gegenstand sprachwissenschaftlicher Forschung vom Sprache-Sprechen und den Gegenstand medienlinguistischer Forschung von der audiovisuellen Multimodalität her, offenbaren sich erstaunliche Parallelen, und es entsteht die Möglichkeit eine transdisziplinäre (film- und sprachwissenschaftliche) Theorie und Methode zu erarbeiten, die versucht, der Komplexität des Phänomens und der besonderen Attraktion audiovisueller Medien und des dort eingebetteten multimodalen Sprachgebrauchs etwas näher zu kommen. Diese Perspektive einzunehmen »entgrenzt« den Fokus sprachwissenschaftlicher Betrachtung von einem körperlosen kontext-entbundenen Sprachsystem zu einem kontext-gebundenen, verkörperten, immer zeitlich zu denkenden Sprechen, einem *Languageing* (vgl. Larsen-Freeman 2003), das immer schon multimodal ist. Es »entgrenzt« auch die Perspektive linguistischer Forschung um eine film- und medienwissenschaftliche Perspektive auf den Sprachgebrauch in audiovisuellen Medien. Wie sich dieser Ansatz methodisch umsetzen lässt, dokumentieren wir im vorliegenden Band mit einer medienästhetischen Fallstudie aus einem laufenden Forschungsprojekt zu parlamentarischen Reden (Kindler-Mathôt et al. i. d. Bd.). Für das Sprechen in audiovisuellen Medien bedeutet dies, dass die Begriffe der Multimodalität des Sprechens und der audiovisuellen Multimodalität als analytische Kategorien fungieren, die an verschiedene artikulatorische Modi gebunden sind, die in der Analyse die komplexe dynamische Orchestrierung der sprachlichen und der audiovisuellen Multimodalität greifbar machen. Nimmt man jedoch konse-

quent die Perspektive der Zuschauer*innen ein, dann haben wir es mit einem einzigen, spezifischen Modus ästhetischer Erfahrung zu tun. Ein Sprechen, das uns audiovisuelle Bilder vorführen, wird eben nicht nur gehört, sondern gleichzeitig gesehen und gefühlt. Es zeigt sich als Aspekt einer multidimensionalen Erfahrungsgestalt und ist in diesem Sinne ein Modus.

6 Kulturwissenschaftliche Erforschung des Sprachgebrauchs als Medienkritik?

Eine kulturwissenschaftliche Sprachgebrauchsforschung tut gut daran, ihre eigenen historischen Bezüge nicht aus dem Blick zu verlieren. Wie die Bezugnahme auf Bühlers Theorie gezeigt hat, können altbekannte Grundannahmen in einer interdisziplinär orientierten wissenschaftshistorischen Kontextualisierung neu betrachtet große Aktualität gewinnen.¹⁹ Auf diese Weise erschließt sich eine kulturhistorische Tiefe und Weite sprachtheoretischer Reflexionen, die aktuellen Diskursen um *Embodiment*, Interkorporalität, um Sprache als kulturelle Praxis nicht nur eigene kulturhistorische Bezüge und Kontexte aufzeigt, sondern auch die grundsätzliche Historizität geistes- und kulturwissenschaftlicher Diskurse unterstreicht (vgl. etwa Meyer/Streeck/Jordan 2017).

Sprachgebrauchsforschung kulturwissenschaftlich gedacht versucht sich dann daran, die disziplinäre Engführung der Linguistik als Fach hin auf andere kulturwissenschaftliche Fächer zu öffnen – und knüpft damit unter anderem an die eigene philologische Tradition der Sprachwissenschaft an, in der Sprachbetrachtung grundsätzlich verbunden war mit Literaturwissenschaft und historischer Sprach- und Literaturforschung. Sie geht über die philologische Tradition hinaus, wenn sie kulturwissenschaftliche Sprachforschung über die Fächergrenzen öffnet. In diesem Sinne verstehe ich das Credo dieses Bandes – Sprache entgrenzen auch als ein Plädoyer dafür Sprache und Sprachwissenschaft historisch und kulturwissenschaftlich zu denken.

Im Lichte der gesellschaftlichen, sozialen, politischen, kulturellen Bedeutung ihres Gegenstandsbereiches ›Sprache‹ hat sich die Linguistik früh hin

19 So argumentiert etwa Jäger (2010: 5): »Vor allem das lange Ausblenden ausdrucks- und zeichentheoretischer Fragestellungen aus dem sprach- und zeichentheoretischen Diskurs nach 1945, die gegenwärtig etwa im Horizont der Gesture-Forschung wieder aufgegriffen und entfaltet werden, hat wohl hier eine wesentliche Ursache.«

zu anderen Disziplinen geöffnet. Die Bindestrich-Linguistiken entstanden, und Neugründungen erfreuen sich weiterhin großer Beliebtheit: Sozio-, Psycho-, Historio-, Diskurs-, Politico-, und nicht zuletzt die Medien-Linguistik. Diese Bindestrich-Linguistiken haben immer auch die Grenzen der Sprache und mitunter auch des Sprache-Sprechens neu ausgelotet, haben auf neue gesellschaftliche und wissenschaftliche Entwicklungen reagiert und deutlich gemacht, dass die Linguistik hier Wesentliches beitragen kann. Eine Erweiterung der Sprachgebrauchsforschung um einen transdisziplinären (filmwissenschaftlich-linguistischen) Zugang zur Multimodalität des Sprechens in audiovisuellen Medien, wie ich sie in diesem Beitrag vorgeschlagen habe, stellt einen Versuch dar, eine kulturwissenschaftliche Öffnung der Sprachgebrauchsforschung theoretisch und methodisch zu leisten.

Gerade für medienlinguistische Forschung scheint es dabei von außerordentlicher Bedeutung zu sein, medienwissenschaftliche Expertise mit in den Blick zu nehmen. Denn ein tieferes Verständnis der Besonderheiten audiovisueller Medien eröffnet neue Möglichkeiten kritischer Reflexionen des Sprachgebrauchs in audiovisuellen Medien. Es bedeutet, die Forderung Marshall McLuhans »the medium is the message« (McLuhan 1964) noch einmal in den Blick zu nehmen.²⁰ Vielleicht gerade, weil diese mittlerweile zum selbstverständlichen Teil eines *Common Sense* wissenschaftlicher und alltagsweltlicher Medienreflexion geworden ist, und weil zugleich eine sich explosionsartig diversifizierende Medienlandschaft die Gesellschaft vor enorme Herausforderungen stellt – zum Teil mit erheblichen sozialen und politischen Effekten.

Nun wäre es zu kurz gegriffen, diese Explosion lediglich als Form unendlicher Vervielfältigung audiovisueller Repräsentationen des Sprachgebrauchs in verschiedenen Kontexten zu beschreiben. Lassen sich doch selbst auf Sachinformationen spezialisierte mediale Formate wie Nachrichten und Reportagen nicht auf die schlichte Kommunikation von Sachverhalten reduzieren – auch nicht, wenn etwa in Interviews oder Talkshows Sprecher*innen unmittelbar zur Rede kommen. Immer haben wir es mit der Ko-Präsenz von Ausdruck, Darstellung und sozialer Adressierung des Sprechens und Gestikulierens, immer haben wir es mit einer spezifischen medialen Ökologie, einer medialen Orchestrierung dieses Sprache-Sprechens zu tun, die sich etwa im selbstgedrehten YouTube- oder TikTok-Video vom Interview im Rahmen der Tages-themen in ihrer medialen Verfasstheit deutlich unterscheiden.

20 Vgl. auch Müller/Kappelhoff 2018, »Introduction«.

Angesichts dieser medialen Diversifizierung audiovisueller Formate gewinnt Walter Benjamins Klassiker »Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit« einmal mehr bedeutende Aktualität (Kappelhoff 2018: Kap.1). Mit Benjamin gerät nämlich die Frage in den Blick, wie diese verschiedenen medialen Ökologien das Sehen, Hören, Fühlen ihrer Rezipient*innen modellieren, was sie ihnen buchstäblich zu sehen, zu hören und zu fühlen geben: »Nimmt man den Gedanken ernst, dass unser Sinnesapparat ein Produkt der Geschichte der Medien menschlicher Wahrnehmung ist, dann ist das Sehen, Hören, Fühlen zunächst eine Frage des Mediengebrauchs« (Kappelhoff 2018: 10). Damit verbindet sich die Frage nach der Verfasstheit audiovisueller Formate mit der Praxis eines Medienkonsums, in dessen Vollzug das Sprechen auf YouTube oder in den abendlichen TV-Nachrichten den Rezipient*innen nicht einfach begegnet, sondern als Form des Mediengebrauchs die Wahrnehmung, das Denken und Fühlen der Zuschauer*innen und Zuhörer*innen rahmt, strukturiert und modelliert. Zugleich haben wir es »in der Aneignung audiovisueller Bewegtbilder« mit einer kulturellen Praxis zu tun »in der Menschen ihre subjektive Wahrnehmungswirklichkeit als Teil einer gemeinsam geteilten Welt zu fassen suchen« (Kappelhoff 2018: 10).

Als kulturelle Praxis zeigt sich diese Rezeption audiovisueller Bilder in den verschiedensten, häufig global formierten, Geschmacksgemeinschaften deren Zahlen sich in Klicks und zum Teil in Millionen von *Follower*innen* manifestieren. Diese neuen Gebrauchskontexte stellen (nicht nur) die linguistische Forschung vor erhebliche Herausforderungen. Sie verlangen nach einer theoretischen und analytischen Durchdringung dieser Veränderungen. Sie erfordern nicht zuletzt die Arbeit an einer Kompetenz zur Medienkritik.

Die hier vorgestellte medienästhetische Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien unterbreitet hierzu einen transdisziplinären Vorschlag. Wenn sie das Sprache-Sprechen als multimodales Ereignis als Fühlen, Sehen, Empfinden der Betrachter*innen adressiert, geht es ihr um die unmerklichen Modulationen der Zuschauerempfindungen mit ihren gemeinschaftsbildenden Effekten. Es geht darum, diese theoretisch und empirisch zu greifen, es geht um Grundlagen für eine kritische Sprachgebrauchs- und Medienforschung.

Literaturangaben

- Armstrong, David. F./William C. Stokoe/Sherman E. Wilcox (1995): *Gesture and the nature of language*, Cambridge: Cambridge University Press.
- Balázs, Bela (1924): *Der sichtbare Mensch oder die Kultur des Films*, Leipzig: Deutsch-Österreichischer Verlag.
- Bressem, Jana/Silva H. Ladewig (2011): Rethinking gesture phases: Articulatory features of gestural movement? In: *Semiotica* 184:1/4, 53–91. <https://doi.org/10.1515/semi.2011.022>
- Bühler, Karl (1968 [1933]): *Ausdruckstheorie. Das System an der Geschichte aufgezeigt*, Jena: Fischer.
- Bühler, Karl (1982 [1934]): *Sprachtheorie. Die Darstellungsfunktion der Sprache*, Jena: Fischer.
- Damasio, Antonio R. (1999): *The Feeling of What Happens. Body and Emotion in the Making of Consciousness*, New York: Harcourt Brace.
- Deleuze, Gilles (2008a): *Cinema 1. The Movement Image*, London: Continuum.
- Deleuze, Gilles (2008b): *Cinema 2. The Time Image*, London: Continuum.
- Eisenstein, Sergej M. (2010 [1926]): Bela Forgets the Scissors. In: Taylor, Richard (Hg.), *Writings, 1922–1934. Sergei Eisenstein Selected Works*, London: I.B. Tauris, 77–81.
- Ekman, Paul (1972): *Emotions in the Human Face*, New York: Pergamon.
- Ekman, Paul (2006): *Darwin and Facial Expression: A Century of Research in Review*, Cambridge: Malor Books.
- Eschbach, Achim (Hg.) (1984): *Bühler-Studien. Band 1 und 2*, Frankfurt a.M.: Suhrkamp.
- Flusser, Vilem (1991): *Gesten. Versuch einer Phänomenologie*, Düsseldorf: Bollmann.
- Fuchs, Thomas (2017): Intercorporeality: Emerging socialities in interaction. In: Meyer, Christian/Jürgen Streeck/Scott Jordan (Hg.), *Intercorporeality and interaffectivity*, Oxford: Oxford University Press, 3–25.
- Fuchs, Thomas/Hanne De Jaegher (2009): Enactive intersubjectivity: Participatory sense-making and mutual incorporation. In: *Phenomenology and the Cognitive Sciences* 8, 465–486.
- Greifenstein, Sarah (2020): *Tempi der Bewegung – Modi des Gefühls. Expressivität, heitere Affekte und die Screwball Comedy*, Berlin/Boston: De Gruyter.
- Greifenstein, Sarah/Dorothea Horst/Thoma Scherer/Christina Schmitt/Hermann Kappelhoff/Cornelia Müller (Hg.) (2018): *Cinematic Metaphor in*

- Perspective. Reflections on a Transdisciplinary Framework*, Berlin/Boston: De Gruyter Mouton.
- 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 Bressemer (Hg.), *Body – Language – Communication. Handbücher zur Sprach- und Kommunikationswissenschaft*, Berlin/Boston: De Gruyter Mouton, 2112–2125.
- Jäger, Ludwig (2010): Sprache als Organon. Karl Bühlers Beitrag zur Begründung der modernen Sprachwissenschaft. In: *Sprache und Literatur* 41:105, 3–17.
- Kappelhoff, Hermann (2004a): *Matrix der Gefühle. Das Kino, das Melodrama und das Theater der Empfindsamkeit*, Berlin: Vorwerk 8.
- Kappelhoff, Hermann (2004b): Unerreichbar, unberührbar, zu spät. Das Gesicht als kinematografische Form der Erfahrung. In: *montage/av*, 13:2, 29–53.
- Kappelhoff, Hermann (2006): Dauer der Empfindung. Von einer spezifischen Bewegungsdimension im Kino. In: Bischof, Margrit/Claudia Feest/Claudia Rosiny (Hg.), *E_motion*, Münster: LIT, 205–219.
- 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 Gemeinnutz: Hollywood zwischen Krieg und Demokratie*, Berlin/Boston: De Gruyter.
- Kappelhoff, Hermann (2018): *Kognition und Reflexion. Zur Theorie filmischen Denkens*, Berlin/Boston: Walter de Gruyter.
- Kendon, Adam (1970): Movement coordination in social interaction: Some examples described. In: *Acta Psychologica* 32, 101–125. [https://doi.org/10.1016/0001-6918\(70\)90094-6](https://doi.org/10.1016/0001-6918(70)90094-6).
- Kendon, Adam (1972): Some relationships between body motion and speech. An analysis of an example. In: Siegman, Aron Wolfe/Benjamin Pope (Hg.), *Studies in Dyadic Communication*, New York: Pergamon Press, 177–210.
- Kendon, Adam (1980): Gesture and Speech: Two aspects of the process of utterance. In: Kendon, Adam (Hg.), *Nonverbal Communication and Language*, The Hague: Mouton, 207–227.

- Kendon, Adam (1990): Spatial organization in social encounters. The F-formation system. In: Kendon, Adam (Hg.), *Conducting interaction. Patterns of behavior in focused encounters*, Cambridge: Cambridge University Press, 209–238.
- Kendon, Adam (2004): *Gesture: Visible action as utterance* (1. Aufl.), Cambridge: Cambridge University Press.
- Kennedy, Adam (2013): Laban based analysis and notation of body movement. 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. Handbücher zur Sprach- und Kommunikationswissenschaft: Bd. 38.1*, Berlin/Boston: De Gruyter Mouton, 941–957.
- Krumhuber, Eva/Susanne Kaiser/Kappas Arvid/Klaus Scherer (2013): Body and speech as expression of inner states. 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. Handbücher zur Sprach- und Kommunikationswissenschaft: Bd. 38.1*, Berlin/Boston: De Gruyter Mouton, 551–564.
- Laban, von Rudolf (1956): *Principles of Dance and Movement Notation*, London: Macdonald & Evans.
- Laban, von Rudolf/Frederick C. Lawrence (1947): *Effort: Economy in body movement*, London: MacDonald and Evans.
- Larsen-Freeman, Diane (2003): *Teaching language. From grammar to grammaring*, Boston, MA: Heinle/Cengage.
- Loenhoff, Jens (2017): Intercorporeality as a Foundational Dimension of Human Communication. In: Meyer, Christian/Jürgen Streeck/J. Scott Jordan (Hg.), *Intercorporeality: emerging socialities in interaction*, Oxford: Oxford University Press, 25–49.
- McLuhan, Marshall (1964): *Understanding Media. The Extensions of Man*, New York: New American Library.
- McNeill, David (1985): So you think gestures are nonverbal? In: *Psychological Review* 92:3, 350–371.
- McNeill, David (1992): *Hand and mind: What gestures reveal about thought*, Chicago: University of Chicago Press.
- Meyer, Christian/Jürgen Streeck/J. Scott Jordan (Hg.) (2017): *Intercorporeality: emerging socialities in interaction*, Oxford: Oxford University Press.
- Merleau-Ponty, Maurice (1968 [1964]): *The Visible and the Invisible*, Evanston: Northwestern University Press.

- Merleau-Ponty, Maurice (2005 [1945]): *Phenomenology of Perception*, London: Routledge.
- Mondada, Lorenza (2013): Multimodal interaction. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Sedinha Tefßendorf (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. Handbücher zur Sprach- und Kommunikationswissenschaft: Bd. 38.1*, Berlin/Boston: De Gruyter Mouton, 577–589.
- Müller, Cornelia (1998): *Redebegleitende Gesten. Kulturgeschichte, Theorie, Sprachvergleich*, Berlin: Berlin Verlag.
- Müller, Cornelia (2008): What gestures reveal about the nature of metaphor. In: Cienki, Alan/Cornelia Müller (Hg.), *Metaphor and gesture*, Amsterdam/Philadelphia: John Benjamins, 219–245.
- Müller, Cornelia (2010): Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. In: *Sprache und Gestik. Sonderheft der Zeitschrift Sprache und Literatur* 41:1, 37–68.
- Müller, Cornelia (2015): Using gestures with speech: Variable cognitive-semantic and pragmatic relations. In: *Vestnik of Moscow State Linguistic University* 6:717, 452–466.
- Müller, Cornelia (2016): From mimesis to meaning: A systematics of gestural mimesis for concrete and abstract referential gestures. In: Zlatev, Jordan/Goran Sonesson/Piotr Konderak (Hg.), *Meaning, Mind and Communication. Explorations in Cognitive Semiotics*, Frankfurt a.M.: Peter Lang, 211–226.
- Müller, Cornelia (2018): Gesture and sign: Cataclysmic break or dynamic relations? In: *Frontiers in Psychology* 9:1651.
- Müller, Cornelia (2024a): Mimesis as as-if action. In: Zywiczynski, 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/Ulrike Bohle (2007): Das Fundament fokussierter Interaktion: Zur Vorbereitung und Herstellung von Interaktionsräumen durch körperliche Koordination. In: Schmitt, Reinhold (Hg.), *Koordination: Analysen zur multimodalen Interaktion*, Tübingen: Narr, 129–165.
- Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Sedinha Tefßendorf (Hg.) (2013): *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. Handbücher zur*

- Sprach- und Kommunikationswissenschaft: Bd. 38.1*, Berlin/Boston: De Gruyter Mouton.
- Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressemer (Hg.) (2014): *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. Handbücher zur Sprach- und Kommunikationswissenschaft: Bd. 38.2*, Berlin/Boston: De Gruyter Mouton.
- Müller, Cornelia/Alan Cienki (2009): When speech and gesture come together. Forms of multimodal metaphor in the use of spoken language. In: Forceville, Charles J./Eduardo Urios-Aparisi (Hg.), *Multimodal Metaphor*, Berlin/Boston: De Gruyter Mouton, 299–332.
- Müller, Cornelia/Hermann Kappelhoff (2018): *Cinematic Metaphor. Experience – Affectivity – Temporality*, Berlin/Boston: De Gruyter Mouton.
- Münsterberg, Hugo (2002 [1916]): The Photoplay – A Psychological Study. In: Langdale, Allan (Hg.), *Hugo Münsterberg on Film. The Photoplay – A Psychological Study and Other Writings*, London: Routledge, 45–162.
- Neisser, Ulric (1976): *Cognition and Reality: Principles and Implications of Cognitive Psychology*, San Francisco: W.H. Freeman & Co.
- Oben, Bert/Geert Brône (2015): What You See Is What You Do. On the Relationship Between Gaze and Gesture in Multimodal Alignment. In: *Language and Cognition* 7:4, 546–562.
- Oben, Bert/Geert Brône (2016): Explaining Interactive Alignment. A Multimodal and Multifactorial Account. In: *Journal of Pragmatics* 104, 32–51.
- Plessner, Helmuth (1982 [1957]): Ausdruck und menschliche Natur. In: Dux, Günter/Odo Marquard/Elisabeth Ströker (Hrsg), *Helmuth Plessner. Gesammelte Schriften* 10,7, Frankfurt a.M.: Suhrkamp.
- Plessner, Helmuth (1982 [1925]): Die Deutung des mimischen Ausdrucks. Ein Beitrag zur Lehre vom Bewußtsein des anderen Ichs [in Zusammenarbeit mit Buytendijk, F. J. J.]. In: Dux, Günter/Odo Marquard/Elisabeth Ströker (Hrsg), *Helmuth Plessner. Gesammelte Schriften*, Frankfurt a.M.: Suhrkamp, 67–129.
- Schegloff, Emmanuel A. (1984): On some gestures' relation to talk. In: Atkinson, Maxwell J./John Heritage (Hg.) *Structures of Social Action. Studies in Conversation Analysis*, Cambridge: Cambridge University Press, 266–296.
- Scherer, Thomas J. (2024): *Inszenierungen zeitgenössischer Propaganda. Kampagnenfilme im Dienste des Gemeinwohls*, Berlin/Boston: De Gruyter.
- Schmitt, Chistina (2020): *Wahrnehmen, fühlen, verstehen, Metaphorisieren und audiovisuelle Bilder*, Berlin/Boston: De Gruyter.

- Schmitt, Reinhold (2005): Zur multimodalen Struktur von turn-taking. In: *Gesprächsforschung – Online Zeitschrift zur verbalen Interaktion* 6, 17–61.
- Selting, Margret/Peter Auer/Dagmar Barth-Weingarten/Jörg Bergmann/Pia Bergmann/Karin Birkner/Elizabeth Couper-Kuhlen/Arnulf Deppermann/Peter Gilles/Susanne Günthner/Martin Hartung/Friederike Kern/Christine Mertzluft/Christian Meyer/Miriam Morek/Frank Oberzaucher/Jörg Peters/Uta Quasthoff/Wilfried Schütte/Anja Stukenbrock/Susanne Uhmann (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, 353–402. Abrufbar unter: <https://www.gespraechsforschung-online.de/fileadmin/dateien/heft2009/px-gat2.pdf>
- Seyfeddinipur, Mandana (2006): *Speech-dysfluencies. Interrupting speech and gesture*, Nijmegen: Radboud University.
- Simmel, Georg (1993): Die ästhetische Bedeutung des Gesichts. In: Kramme, Rüdiger/Alessandro Cavalli (Hg.), *Aufsätze und Abhandlungen, 1901–1908*. 1. Aufl., Bd. 1, Berlin: Suhrkamp, 36–42.
- Stern, Daniel N. (1985): *The Interpersonal World of the Infant. A View from Psychoanalysis and Developmental Psychology*, London: Routledge.
- Sobchack, Vivian (1992): *The Address of the Eye. A Phenomenology of Film Experience*, Princeton: Princeton University Press.
- Streeck, Jürgen (2009): *Gesturecraft. The manufacture of meaning*, Amsterdam/Philadelphia: John Benjamins.
- Streeck, Jürgen/UlrikeHartge (1992): The contextualization of language. In: Auer, Peter/Aldo Di Luzio (Hg.), *Previews: Gestures at the transition place*, Amsterdam/Philadelphia: John Benjamins, 135–158.
- Wundt, Wilhelm M. (1975 [1900–1925]): *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte: Vol. 2. Die Sprache*, Alfred Kröner Verlag.

Cornelia Müller, Prof. Dr.
 Europa-Universität Viadrina
 Große Scharnstraße 59
 15230 Frankfurt (Oder)
 cmueller@europa-uni.de

