

»Language is (not) obvious« Asynchrone Bewegtbilder von Geste und Wort

Einleitung

Die natürlichen Bewegungen des Körpers - und vor allem die Gesten - begleiten die verbale Sprache in jeder Sprechsituation. In audiovisuellen Bewegtbildern, besonders solchen die für die Dissemination von Wissen produziert werden, steht heute oft der Sprecher mit seinem sichtbaren Körper im Mittelpunkt. Insbesondere wenn diese Videos online verbreitet werden, tritt dabei häufig Asynchronität zwischen Bild- und Tonspur auf.

Im vorliegenden Beitrag wird anhand dreier Videobeispiele aufgezeigt, wie asynchrone Versätze von Geste und Wort eine Bedeutungsveränderung erzeugen. Als Beispiele dienen erstens Interview-Videos mit Slavoj Žižek, einem geübten Redner, der deutlich ausgeprägte Gesten verwendet. Zweitens wird Bezug auf eine Aufnahme eines Interviews mit Michael Tomasello genommen. In diesem Beispiel wird durch den Bild- und Ton-Versatz eine autoperformative Repräsentation seiner Aussage generiert. Tomasello spricht über die kontextuelle Relevanz der Zeigegeste, während sein Zeigen ironischerweise temporal versetzt zu der verbalen Sprache geschieht.

Im Sinne einer Verkörperung des Wissens können solche temporalen Verschiebungen die Bedeutung der Sprache sowohl verändern indem sie zu unbeabsichtigten Rekombinationen von Geste und Wort führen.

Hintergrund

Die redebegleitenden Gesten erzeugen in ihrer visuellen Performanz zusätzliche Aspekte der Sprache, die sie ergänzen, erweitern oder sogar ersetzen können (McNeill & Duncan 2002; McNeill 2012).¹ Gesten sind Teil einer Rückkopplung zwischen Körper, Umwelt und Handlung und widerspiegeln in ihrer kommunikativen Expression den momentanen Zustand des Individuums. Sie sind in ihrer Dynamik ein offenes Fenster in die mentale Repräsentation des Sprechers (McNeill 1992; Goldin-Meadow et al. 1993; Kendon 2004) und sind damit ein bedeutungsvoller Bestandteil der menschlichen Sprache (Clark 1996; 2003; Spencer/Özyürek 2010). Gesten sind verkörperte Aktionen, die aktiv als Teil des kognitiven Prozesses in der Sprache und im Denken involviert sind (Clark 2013; Pouw u.a. 2014) und werden damit als Beweis dafür angesehen, dass Wissen selbst verkörpert ist (Gibbs, 2006a; McNeill 2005; Núñez 2005; Hostetter/Alibali, 2008; Alibali/Nathan 2012).

Mit den Gesten befindet sich zudem der gesamte Körper des Sprechers in einer wenn auch häufig subtilen, so doch kontinuierlichen Bewegung. Vor allem die sogenannten ›beat-gesten‹, zu Deutsch ›Schlaggesten‹, welche die verbale Sprache in einem wiederkehrenden, rhythmischen Muster strukturieren, übertragen sich mit ihrer Dynamik auf den gesamten Körper des Sprechers. Der Sprecher zeigt damit gleichzeitig mit seinen Worten die Sprache als bildsprachlich und rhythmisch

¹ Redebegleitende ›speech-framed‹ Gesten können in vier Dimensionen eingeteilt werden: deiktische, ikonische, metaphorische sowie beat-Gesten. Die vier Dimensionen sind von der semiotischen Kategorisierung von C.S. Peirce inspiriert. Alle vier Gestikulationen sind speech-framed gestures des ›Kendon Continuum‹ (McNeill 2006).

verkörperte.

Innerhalb der Theorien der Verkörperung ist die Asynchronität in Bewegtbildern ein interessantes Thema, weil in der asynchronen Repräsentation eines Sprechers Kontraste und Differenzen entstehen.² Bild und Ton werden aus ihrer anthropologischen Einheit getrennt und, in einer Synchrese neu zusammengefügt. Die Synchrese ist ein Begriff von Michel Chion (1994) und definiert als »the forging of an immediate and necessary relationship between something one sees and something one hears at the same time [...]« (ebd.: 224).³ Mit der Synchrese verweist Chion vor allem auf Filme mit einem künstlerischen Hintergrund wie diejenigen der Avantgarde, die absichtlich Bild und Ton auf ungewöhnliche Weise rekombinieren. Es geht hier jedoch nicht um eine Kritik an intentional eingesetzter Asynchronität als Stilmittel, wie sie beispielshalber Jean-Luc Godard mit *2 ou 3 choses que je sais d'elle*, Chris Marker mit *Sans Soleil* oder Wim Wenders mit seinem Debutfilm *What's up Tiger Lilly* eingesetzt haben. Der Begriff der Synchrese ist auch für die Analyse von Formaten die eher einen dokumentarischen Charakter aufweisen äußerst interessant da auch in diesen ein determiniertes Schmieden von Gehörtem und Gesehenem geschieht. Deshalb werden im Folgenden die Konsequenzen von asynchronem Bild und Ton auf die Gesten eines sichtbaren Sprechers und seine verkörperte Sprache in Formaten untersucht, die eine Dissemination von Wissen in einem wissenschaftlichen, didaktischen oder politischen Kontext mit sich bringen.

Eine Interdependenz des Gehörten (verbale Sprache) und gleichzeitig Gesehenen (Gesten und Körper des Sprechers) gilt unabhängig davon, ob es sich um audiovisuelle Bewegtbilder oder natürliche Audiovision handelt. Wenn im Folgenden die Begriffe der *natürlichen Audiovision* verwendet werden, so verweisen diese auf das Sehen und das Hören in einem anthropologischen Verständnis, sprich ohne Kamera und Monitor.

Die kognitive, mentale und körperliche Disposition für die multimodale audiovisuelle Wahrnehmung und Verarbeitung von Reizen variiert zwar je nach der individuellen Konstitution des Einzelnen und gilt unabhängig davon, ob es sich um die Verarbeitung von Reizen aus audiovisuellen Bewegtbildern oder um solche aus der natürlichen audiovisuellen Sprachwahrnehmung handelt.⁴

Diese Ähnlichkeit in der Rezeption und Verarbeitung von Reizen aus der natürlichen Audiovision mit audiovisuellen Bewegtbildern wurde in zahlreichen Untersuchungen aus den kognitiven Filmstudien, der Wahrnehmungspsychologie und den Neurowissenschaften aufgezeigt (Anderson 1996; Cutting 2005; Zacks/Magliano 2011; Smith 2012).⁵

² Der Begriff der Synchronität welcher hier verwendet wird, ist nicht äquivalent mit dem Jung'schen Begriff der ›Synchronizität‹. Synchronizität beschreibt ein psychoanalytisches Paradigma von metaphysischen Ereignissen, welche zur gleichen Zeit geschehen, zum Beispiel bleibt eine Uhr stehen, während der Großvater stirbt. Der Begriff der Synchronität ist ein technischer Begriff aus dem Film und besteht aus ›synchron‹ besteht aus ›syn‹: zusammen, zugleich und ›chronos‹: die Zeit - und wird im audiovisuellen Bewegtbild als Synonym für »gleichzeitig, zeitgleich und übereinstimmend« verwendet (vgl. Rudloff 1997). ›Asynchron‹ oder die Asynchronität bezeichnet alle nicht-zeitlich zusammenfallenden Begebenheiten welche im eigentlichen Sinne einen Ursprung und Ziel gemeinsam haben.

³ »La synchrèse (mot que nous forçons en combinant 'synchronisme' et 'synthèse') est la soudure irrésistible et spontanée qui se produit entre un phénomène sonore et un phénomène visuel ponctuel lorsque ceux-ci tombent en même temps, cela indépendamment de toute logique rationnelle« (Chion, 1990: 55).

⁴ ›Multimodal‹ beschreibt, »dass unserer Sprachwahrnehmung durch Informationen aus einer Reihe anderer Sinne beeinflusst werden kann« (Goldstein 2007: 324).

⁵ Es ist wichtig hierbei zu erwähnen, dass beinahe alle empirischen Studien, welche die Geste und Wort in der natürlichen Audiovision heute untersuchen, audiovisuelle Bewegtbilder als Untersuchungsinstrument verwenden.

Die Inszenierung, Produktion und Wiedergabe von audiovisuellen Bewegtbildern selbst unterscheidet sich jedoch von der natürlichen Audiovision hinsichtlich vieler Parameter. Neben der Montage, die seit Beginn der Bewegtbild Medien immer wieder eine zentrale Position im Diskurs einnimmt, ist vor allem die medienspezifische Eigenschaft der Temporalität und der Bild-Ton Rekombination zu nennen, die in der natürlichen Audiovision nicht in einer vergleichbaren Eigenart vorkommen kann. Eine Asynchronität von Bild und Ton kann beispielsweise durch einen temporalen Versatz von Bild und Ton (willkürlich oder unwillkürlich), Nachsynchronisation (Dubbing) oder aber auch durch Rekombination von Bild und Ton erzeugt werden. Asynchrone Bild-Ton Verhältnisse von Gesten eines sichtbaren Sprechers existieren in dieser Weise in der natürlichen Audiovision nicht. Die Gesten besitzen einzigartige, semiotische Eigenschaften mit einem Bedeutungsgehalt, welcher erst in einer synchron ausgeführten Geste-Wort Expression gänzlich zum Vorschein kommen kann. Es könnte gesagt werden, dass erst, wenn der Sprecher mit seinem Körper sichtbar ist, das gesamte Spektrum der Aussage in seiner *verkörperten* Sprache gezeigt wird. Diese Aussage ist unabhängig davon gültig, ob es sich um natürliche Audiovision oder eine Repräsentation im audiovisuellen Bewegtbild handelt. Gleichzeitig entsteht dadurch jedoch ein Problem für das audiovisuelle Bewegtbild⁶. Denn gerade wenn ein Sprecher im Bewegtbild sichtbar ist und durch fehlerhaftes Vorgehen ein zeitlicher Bild- und Ton-Versatz eintritt, so wird die Bedeutung des Körpers durch die asynchrone Wiedergabe von der verbalen Sprache getrennt. Der relevante, verkörperte Bestandteil der Sprache wird nicht mehr im Gesamtkontext der Sprache sichtbar sondern wird getrennt davon wahrnehmbar. Insofern stellen sich folgende Fragen: Wie verhalten sich die Darstellung des Körpers - und spezifisch der Gesten - eines sichtbaren Sprechers zur verbalen Sprache, wenn diese in audiovisuellen Bewegtbildern mit einem temporalen Versatz von Bild und Ton gezeigt werden? Inwiefern wird durch die durch die Synchronisation von asynchronen Bewegtbildern eine Bedeutungsveränderung erzeugt?

Asynchronität von Geste-Wort bei sichtbaren Sprechern

Bild und Ton bedingen sich im audiovisuellen Bewegtbild gegenseitig; das, was gesehen wird, beeinflusst das, was gehört wird und vice versa. Spätestens als McGurk & McDonald (1976) die crossmodale Abhängigkeit der audiovisuellen Sprachwahrnehmung effektiv aufzeigten, wurde klar, dass das Zusammenspiel von Sehen und Hören im Bewegtbild-Medium komplexen Verarbeitungs-Prozessen unterliegt. Selbst wenn willentlich versucht wird das Gesagte nicht mit dem simultan Gehörten zusammenzuführen, kann eine unwillkürliche Verbindung aus visuellen und auditiven Reizen nicht unterbunden werden. Vielmehr ergänzt sich das Gehörte mit dem Gesehenen trotz des expliziten *Wissens* der ursprünglichen Getrenntheit der Reize zu einer Einheit.

⁶ Gerade heute, wo weltweit unzählige Bewegtbild-Konsumenten ebenso Bewegtbild-Produzenten sind, finden sich vermehrt digitale, asynchrone Bewegtbilder. Auf der Statistik-Plattform von ›Youtube‹ wird angegeben, dass 2014, rund 100 Stunden Videomaterial pro Minute hochgeladen wird, pro Monat die Plattform von mehr als 1 Milliarde Nutzer besucht wird (wieviele davon wirklich Menschen oder Cyborgs und Bots sind ist nicht ersichtlich) und in einem Jahr über 6 Milliarden Stunden Video abgespielt werden (siehe Youtube Statistik 2014). Es besteht Grund zur Annahme, dass die meisten Asynchronitäten unabsichtlich produziert worden sind und diese aus Produktionsfehlern resultieren (wie zeitliche Spurverschiebungen in der Montage, Render-Probleme, Kompressions- und Formatfehler etc.). Auf diesen Aspekt kann im Rahmen dieser Arbeit nicht weiter eingegangen werden. Literatur zu technischen Fehlerquellen von Asynchronität findet sich jedoch reichlich.

Michel Chion definiert die Zirkulation von Sehen und Hören in Bewegtbildern folgendermaßen: »We never see the same thing when we also hear; we don't hear the same thing when we see as well« (Chion 1994:19). Chion spricht von einer determinierten Abhängigkeit der Wahrnehmung, einer multimodalen Integration des Auditiven und Visuellen, welche unabdingbar und untrennbar in die Bedeutung welche man auch als Amalgamierung bezeichnen könnte. Wird seine Aussage auf die Sprachwahrnehmung eines sichtbaren Sprechers bezogen, so würde dies die Geste und das Wort betreffen. Folglich könnte Chion paraphrasierend gesagt werden: Wir sehen niemals die gleiche Geste, wenn wir die Worte dazu hören; wir hören ebenfalls nicht die gleichen Worte, wenn wir die Geste dazu sehen. In der Asynchronität von Bewegtbildern existiert eine beinahe unbegrenzte Anzahl an Möglichkeiten für eine Rekombination von Geste und Wort.

Beispiel Eins: Geste-Wort-Schere

Wird nun die anthropologische Bedeutung von Geste und Wort im Bewegtbild durch Asynchronität getrennt, so könnte von einer *Geste-Wort-Schere* gesprochen werden. Die Geste-Wort-Schere ist ein von der Ton-Bild-Schere abgeleiteter Begriff.

Die Diskussion über die Ton-Bild-Schere, oder auch Text-Bild-Schere, im TV-Nachrichtengenre und deren Wirkung auf die inhaltsbasierte Wahrnehmung von Informationen findet ihre Anfänge in den 1970er Jahren bei Bernward Wember (1976). Sie beschreibt folgendes Verhältnis: Sind Bild und Ton von einem Versatz betroffen, so dass das Bild eine inhaltlich verschiedenartige Quelle als der Ton besitzt, wird von einer Ton-Bild-Schere gesprochen. Um Nachrichtensendungen zu *bebildern*, wird zum gesprochenen Text inhaltlich passendes Bildmaterial gesucht. Wenn anschließend die Off-Stimme des Nachrichtensprechers über eine gewaltsame Revolution spricht und gleichzeitig lachende Menschen zu sehen sind, die in einem Vorgarten sitzen und Wein trinken, trennt sich die Aussage von bildlichem und hörbarem Inhalt.

Betrifft dies einen sichtbaren Sprecher, so beschreibt die Geste-Wort-Schere die resultierende Bedeutungsdifferenz zwischen verkörperter Geste und gesprochenem Wort.

Erscheint eine solche Geste-Wort-Schere im Bewegtbild, wird unabhängig von dem Wissen um den *fehlerhaften* Bild- und Ton-Versatz dennoch, wenn auch nicht immer durchgehend, das Bild mit dem Ton kombiniert. Denn das unmittelbare kognitive Prozessieren der Bilder und Töne kann nicht isoliert vollzogen werden, alsbald die Reize simultan auftreten (vgl. McGurk Effect).

Eine solche Geste-Wort-Schere kann gleichermaßen, auch wenn Geste und Wort ursprünglich aus derselben Aufnahme entstanden sind, anhand eines zeitlichen Versatzes in einem punktuellen, kontradiktorischen Auseinanderklaffen der Bedeutungsinhalte von Geste und Wort resultieren. Punktuell und nicht sequenziell, da zwar das asynchrone Verhalten von Visuellem und Auditivem in der Sequenzabfolge erkannt wird, sie jedoch trotzdem in einer synchretischen Konstruktion von Bild und Ton verschmelzen.

Ein Beispiel für eine Geste-Wort-Schere ist in Abbildung 1 sichtbar. Die zwei verschiedenen Standbilder zeigen den Philosophen Slavoj Žižek aus dem dokumentarischen Video-Interview *Don't act. Just think*, welches von der Organisation *Big Think* auf *Youtube* veröffentlicht worden ist (Žižek

2012).⁷

Žižek spricht im Zeitraum des Standbildes über Occupy Wallstreet:

»Okay, we heard your story protest, horrible, big banks depriving us of billions, hundreds thousand of billions of common peoples money, okay. But, what do you really want, what should replace *the system*« (1:13min - 1:28min /Don't Act. Just Think).



Abb. 1: Žižek, »the system«. Das linke Standbild zeigt die synchrone Fassung. Das rechte Standbild zeigt die asynchrone Version, die Geste-Wort-Schere. Quelle: Youtube, Big Think.

Die Untertitel visualisieren die im Zeitraum des Standbild gesprochenen Wörter: *the system* und wurden im Nachhinein eingefügt. Žižek ist in einer *Halbnahen* Einstellung in Richtung Kamerabild zu sehen und sein Blick ist direkt in die Kamera gerichtet.⁸ Das linke Standbild zeigt die synchrone Originalfassung, das rechte Standbild eine asynchrone Bild-Ton Rekombination. Die Standbilder sind im Moment des gestischen *strokes* erstellt in welchem der Bedeutungsgehalt der Geste am deutlichsten erkennbar ist (Kendon 2004; McNeill 2008).⁹ In Abb. 1 erzeugt *the system* ein unterschiedliches Resultat, je nachdem, ob das linke Bild oder das rechte Bild gesehen wird. Während die Körpergeste von Žižek in der synchronen Fassung (links) einen eher komplexen und schwer fassbaren Begriff eines Systems darstellt, so verweist die asynchrone Geste (rechts) auf eine betonte klare Vorstellung des Systems. Die eigentliche Bedeutung über die Art und Weise, wie Žižek über das System denkt, spricht und nicht zuletzt bildsprachlich visualisiert, ändert sich durch eine asynchrone Geste-Wort-Schere fundamental.

Die redebegleitenden Gesten besitzen in ihrer gleichursprünglichen Entstehung mit den Worten dieselbe Grundidee, bringen jedoch nicht notwendigerweise den gleichen Aspekt davon zum Ausdruck. Gesten und Worte vermitteln in ihrer Expression einen Bestandteil dessen, was im Denkprozess entsteht: Während die gesprochene Sprache einer eher konventionalisierten Form der

⁷ Alle Videobeispiel der Arbeit sind offen zugänglich (siehe Reutemann 2014).

⁸ In audiovisuellen Bewegtbildern besteht eine klassische Bild-Kadrierung (Cadre) von sichtbaren (on-screen) Sprechern zumeist in Variationen aus *Halbtotale*, *Halbnahen* oder *Nahen* Einstellungen. Die *Halbtotale* und *Halbnahe* bilden den Oberkörper des Sprechers gänzlich ab, wohingegen die *Nah*e meistens den oberen Torso bis zur Kopfspitze des Sprechers zeigen. Die ersten beiden Einstellungen haben gemeinsam, dass sie redebegleitende Gesten größtenteils abbilden. Vor allem die *Nah*e ist als eine Einstellung zu nennen, in welcher die Bewegungsdynamiken stärker hervortreten, da der Körper des Sprechers *näher* abgebildet und sich damit der Radius der Bewegungen in der Cadre vergrößern. Andererseits sind die Gesten in der *Nahen* nur fragmentarisch sichtbar.

⁹ Siehe dazu nächstes Kapitel.

Grammatik, Semantikstruktur oder Betonung folgt, zeigen die Gesten idiosynchratische und imaginäre - bildsprachliche – Formen (McNeill 1992). Vielleicht gerade weil Gesten nicht einer standardisierten Form von Regeln und Satzbau folgen, drücken sie damit das Gedachte in einer »anderen Sprache« aus. Cornelia Müller (2008) erläutert in einem Interview, wie die Bildhaftigkeit von Gesten beispielsweise bei der Beschreibung einer Spinne zustande kommen kann. Mit der Geste kann die Person zeigen, wie eine Spinne sich bewegt und Variationen wie »die Geschwindigkeit der Spinne, ihre Laufrichtung, ihre Vielbeinigkeit, ihre Größe, den ebenen oder unebenen Untergrund [darstellen: J.R.]. Spannen wir die Finger stark an, können wir sogar noch ausdrücken, dass der Anblick des Tieres uns nicht behagt hat« (ebd.).

Der Rhythmus, die Bewegungsdynamik, das Tempo und die Form werden damit in der Bewegung des Sprechenden verkörpert dargestellt und zeigen imaginistische, bildsprachliche Zeichen. Sobald diese visuellen Informationen bei der Sprachwahrnehmung vorhanden sind, können diese nicht mehr *weggedacht* werden.

In der natürlichen Audiovision existiert eine asynchrone Geste-Wort Darstellung äußerst selten. Eine Nicht-Übereinstimmung von Geste und Wort tritt primär bei Kindern auf, welche in einer Phase des Lernprozesses stehen (Church/Goldin-Meadow 1986; Goldin-Meadow u.a. 1993 Goldin-Meadow, Susan 1997). Hierbei kann die Geste die richtige Antwort andeuten, die verbale Sprache jedoch eine unvollständige Aussage mitteilen. Der *mismatch* von Geste und Wort wird dabei als ein Übergangszustand (*Transition State*) angesehen, in welchem die kognitive Entwicklung der Person zum Verständnis der Thematik noch nicht abgeschlossen ist und sich die Unsicherheit in der Diskrepanz oder Mehrdeutigkeit von Geste und Wort zeigt.

Eine widersprüchliche Darstellung von Geste und Wort kann in der natürlichen Audiovision auch als Indiz für Betrug (Ekman/Friesen 1976) oder Mangel an Glaubwürdigkeit gedeutet werden: »Wenn [umgekehrt] Gebärde und Miene mit der Rede in Widerspruch steht, wir also Trauriges mit heiterer Miene sagen oder etwas mit Kopfschütteln bekräftigen, so dürfte gewiss den Worten nicht nur alle Nachdruck, sondern sogar die (schlichte) Glaubwürdigkeit fehlen« (Quintilianus 2011 [35-60 v.Chr.]: 635).

Die redegleitenden Gesten sind fast immer synchron zur verbalen Sprache (McNeill 2012; Church, Spencer & Holcombe 2014). Eine spezifische Eigenschaft von redegleitenden Gesten ist, dass diese in der Phase des *strokes*, also im Moment, in welcher ihre Bedeutung gänzlich gezeigt wird, sich temporal und semantisch in einer Einheit mit der verbalen Sprache befinden (Kendon 2004; Loehr 2007; McNeill 2008).

Wenn der Sprecher in der natürlichen Audiovision sichtbar und hörbar ist, so entsteht ein temporär kausaler Zusammenhang und eine sich gegenseitige, kontextuelle Ergänzung von Visuellem und Auditivem, oder Geste und Wort. Fehlt ein Bestandteil der Information, zum Beispiel durch akustische Probleme, so wird diese Lücke bei Unverständlichkeiten der Sprache geschlossen. Dieser natürliche Ergänzungseffekt tritt bereits bei kleinen Verständnislücken eines Wortes in der auditiven Wahrnehmung auf. Obwohl ein Teil des Wortes nicht richtig verstanden wird, erschließt sich im kognitiven Verarbeitungsprozess eine kontextuelle Ergänzung. Dieser natürliche Effekt der

Phonemergänzung (Warren 1970) erweitert das gehörte Wort passend zum situativen Kontext und konstruiert dessen Bedeutungsgehalt anhand der verfügbaren Reize des perzeptuellen, motorischen sowie mentalen Sets (Zimbardo/Gerrig 2008: 155).

Zusammenfassend kann an diesem Punkt gesagt werden, dass vor allem redebegleitende Gesten in einer engen Verbindung mit der verbalen Sprache stehen und ein wichtiger Bestandteil der Sprache sind, ein offenes Fenster in die Gedanken sozusagen.

Entsteht im audiovisuellen Bewegtbild ein Zeitversatz, bleibt der ergänzende Zugang zu den verkörperten, gestischen Denkprozessen verschlossen. Die ursprüngliche Information, welche durch die Gesten des Sprechers in der audiovisuellen Repräsentation mitgetragen wird, entzieht sich durch asynchrone Bild- und Ton-Versätze dem Zusammenhang der gesprochenen Sprache. Anstatt dessen zeigt die Geste bildhaft eine Bedeutung, die nicht mehr mit dem Wort in seinem ursprünglichen Zusammenhang steht, es entsteht eine Geste-Wort-Schere. Es kann vermutet werden, dass diese Geste-Wort-Schere nicht in einer isolierten Darstellung von Geste und Wort verbleibt, sondern dass, ähnlich dem Effekt der natürlichen *Phonemergänzung*, eine kontextuell-situative Bedeutung aus dem asynchronem Bild und Ton erschlossen wird.

Beispiel Zwei: Rekontextualisierte Zeigegeste

In Zeigegesten manifestiert sich ein Kernelement einer verkörperten Kognition. Die Verwendung der Zeigegeste lässt vermuten, dass der Sprecher die Umwelt als Teil seines kognitiven Systems integriert und einbettet (Alibali u.a. 2014).

Die Zeigegeste »zeigt etwas und sie weist zugleich den Körper vor, der sich zeigt« (Boehm 2010: 25). Durch die körperliche Deixis visualisiert der Sprecher eine performative Differenz, in der die »verhaltene Haltung«, der basso continuo des Redners mit den von ihm ausgehenden Gebärden ein Wechselspiel vollführt. Gerade bei Zeigegesten kann Asynchronität durch die damit verbundene Kontextverschiebung den Bedeutungsgehalt stark verändern.

Die im Folgenden beschriebene Szene aus einem Video-Interview von Michael Tomasello birgt ein *wortwörtliches* (und bildhaftes) Beispiel dafür, wie durch Asynchronität die Zeigegeste von den Wörtern getrennt wird, und eine Bedeutungsveränderung geschieht. Zudem wird in diesem Interview, in welchem Tomasello über die Relevanz der Zeigegeste spricht, eine unabsichtliche Selbst-Referenz erzeugt, und damit seine Argumentation *in situ*, autoperformativ bestätigt.

Das Video wurde 2008 vom Sender 3Sat aufgezeichnet und 2013 von einem privaten User auf Youtube veröffentlicht. Durch ein technisch nicht uninteressantes *Fehlverhalten* besitzt das Video einen über die Dauer von 51:13 Minuten stetig zunehmenden Bild- und Ton-Versatz. Während sich die Asynchronität zu Beginn des Videos auf wenige Millisekunden beschränkt, beträgt der Versatz am Ende des Videos beachtliche sechs Sekunden, in denen die Gesten vor den Wörtern sichtbar sind.

Michael Tomasello hat wichtige Erkenntnisse zum Gebiet der Gestenforschung beigetragen und erzählt in dem TV-Interview über seine Forschungstätigkeit. Um seine Argumentation zu stützen, erklärt Tomasello mit seinen Händen die Bedeutung von Zeigegesten und deren Verlust bei einer

fehlenden Kontextualisierung, während in der Wiedergabe seine Zeigegesten durch den Bild- und Ton-Versatz rekontextualisiert sind:

»Communication is something that clearly differentiates us from our nearest relatives. Language (Zeigefinger in Richtung Kamerabild) is obvious. [...] all the sudden it's, ähm (Zeigefinger kurz in Richtung Kamerabild) meaningful that we're, we have the common ground, that we both like this kind of bicycle or something and I point to it and you know what I mean (Zeigefinger in Richtung Kamerabild). But if I point outside of any context (Blick auf Zeigefinger, Zeigefinger nach oben gerichtet) what so ever, it means nothing. If I just point for no reason, it's absolutely meaningless. There is no information in the finger. You look over there and you see something but you don't know exactly where I am pointing to and you don't know exactly why I'm pointing to it« (Tomasello 2008a).



Abb. 2: Tomasello, »all the sudden«, Bild- und Ton-Versatz: 5.7sec. Quelle: Youtube, privater User.

Der Versatz des Videos ist an dieser Stelle 5.7sec. Tomasello deutet mit seinem Zeigefinger viermal; dreimal in den rechten, unteren Bereich des Kamerabildes, einmal mit dem Finger nach oben. Die Zeigegeste erscheint bei »[...] all the sudden it's, ähm (Zeigefinger auf „ähm“ kurz in Richtung Kamerabild) meaningful [...]« verfrüht und das erwähnte »meaningful« ist nicht erkennbar, da die Geste unspezifisch zum Wort „ähm“ in Richtung Kamera gerichtet und gleich wieder zurückgezogen wird.

Bei der nächsten Position erscheint die Zeigegeste nicht gänzlich bedeutungslos: »[...] that we both like this kind of bicycle or something and I point to it and you know what I mean. (Zeigefinger in Richtung Kamerabild).« Die Zeigegeste kommt mit einer Verzögerung am Ende des Satzes und es hat den Anschein, als ob Tomasello mit »you know« die Zuschauer direkt ansprechen möchte.

Für diese Arbeit ist insbesondere die folgende Stelle von grossem Interesse: »[...] but if I point outside of any context (Blick auf Zeigefinger, Zeigefinger nach oben gerichtet) what so ever, it means nothing [...]«. Die Geste übernimmt hier die Aussage »outside of any context«, steht jedoch in der synchronen Fassung für »there is no information in the finger«.



Abb. 3: Tomasello, »outside of any context«, Bild- und Ton-Versatz: 5.7sec, Quelle: Youtube, privater User.

Die Bedeutung der nach oben gerichteten Zeigegeste bezieht sich somit in der asynchronen Fassung auf ein kontextloses Zeigen. Tomasello richtet gleichzeitig seinen Blick auf den Finger. Die nach oben gerichtete Zeigegeste steht damit durch die Gleichzeitigkeit der Worte in einer neuen Bedeutung – dem kontextlosen Zeigen. Die ursprüngliche Bedeutung der Geste steht jedoch für die Informationslosigkeit des Fingers selbst, wenn dieser nicht kontextuell eingebunden ist. Da der temporale Versatz jedoch in einer Gleichzeitigkeit von Geste und Wort zusammen passt, wird in einer Synchrese eine neue Bedeutung geschmiedet.

Der ursprüngliche Bedeutungsgehalt der gestischen Zeigefunktion verliert gerade hier, wenn die Zeigegeste als intentionale Form der Kommunikation erklärt werden will, ihren Bedeutungskontext zum gesprochenen Wort. »[...] Language (Zeigefinger in Richtung Kamerabild) is obvious« (ebd.) stimmt demnach nur bedingt, denn die Sprache ist bei Asynchronität alles andere als offensichtlich. Tomasello's Kernaussage »Schauen sie einfach, wohin ich zeige, und Sie werden sehen, was ich meine« (Tomasello 2008b) wird in diesem Beispiel in eine rekursive Aussage verwandelt, da die betonte, zeitgebundene Relevanz der Zeigegeste sogleich durch den medialen Bild- und Ton-Versatz nicht mehr sinngemäß *zeigt*.

Rhythmus des Sprechers

Die beat-Gesten¹⁰ dienen weniger einer Bedeutungsergänzung oder Erweiterung, sondern begleiten in einem rhythmischen Muster die Dynamik der verbalen Sprache. Die Bewegungsmuster von Beat-Gesten sind durch mehr oder weniger redundante hin und her oder auf und ab Bewegungen von einer oder beiden Händen gekennzeichnet und markieren wie ein visueller Leuchtstift den Sprechrhythmus. Mikroanalysen von beat-Gesten zeigen, dass diese in einem strukturierten Rhythmus einheitlich mit den gesprochenen Wörtern geschlagen werden (Leonard/Cummins 2011). Eine zeitliche Verzögerung der Schlaggeste, welche nicht im natürlichen Sprechrhythmus liegt, wirkt nicht nur von der Seite eines Betrachters seltsam a-rhythmisch (Treffner, Peter & Kleidon 2008), sondern ist vor allem körperlich für den Sprecher spürbar fremd. Es ist für einen ungeübten Rhetoriker beinahe eine Sache der Unmöglichkeit, einen Gestenschlag a-rhythmisch zu seinen gesprochenen Worten zu koordinieren. Bereits ein reguläres Schlagen der flachen Hand auf die Oberschenkel, das sich a-rhythmisch zu einem gehörten Schlag verhält, erfordert eine hohe

¹⁰ Die Bezeichnung der beat-Gesten wurde aus dem früheren Begriff »Batons« entwickelt (siehe dazu: Efron 1941; Ekman/Friesen 1969).

kognitive, mentale und motorische Kontrolle über die Körperbewegungen. Asynchronität zwischen der verbalen Sprache und einer intentionalen, kontrollierten Schlaggeste die z.B. ein Argumente untermauern und so die Wichtigkeit und Richtigkeit des Gesagten betonen soll, tritt in der natürlichen Audiovision häufig bei einer Lüge oder Täuschung auf. Sie kann daher beim Rezipienten ein Gefühl von Misstrauen erzeugen (Ekman & Friesen 1976).

Die Gesten und die verbale Sprache sind in eine kontinuierliche Bewegung des Körpers eingebettet, und seien diese von nur minimaler Natur. Nebst den redebegleitenden Gesten unterliegen die Kopfbewegungen (Kendon 1972; McClave 2000), die gesprochenen Wörter und sogar das Blinzeln der Augen einer rhythmischen Beziehung und werden aufeinander abgestimmt und miteinander synchronisiert (Condon/Ogston 1966; Kendon 1972; Condon 1986; Loehr 2007). Der rhythmische Puls der Sprache variiert individuell nach Person, emotionalem Zustand, Artikulation, Intensität, Aktivierungszustand, Tätigkeit, etc. und ist phasenweise regelmäßig. Der Körper des Sprechers befindet sich damit in einem »more or regular rhythmic pulse which is, like the semiotic content, expressed through both verbal and nonverbal channels« (Tuite 1993). Nicht nur die Finger oder Hände des Redners, sondern die gesamte Körperlichkeit befindet sich damit in einem kontinuierlichen, rhythmischen Puls mit der gesprochenen Sprache. Der römische Rhetoriker Quintilianus schreibt in seiner Überlieferung von der *Ausbildung des Redners* über Cicero und dessen Wertschätzung des körperlichen Rhythmus:

»[...] denn auch die Bewegung des ganzen Körpers macht etwas aus, sogar so viel, dass Cicero meint, sie spiele eine größere Rolle als selbst die Hände. Er sagt nämlich im Orator:
·Kein Geplapper der Finger, keine Fingerspitzen, die den Rhythmus schlagen, eher soll der Redner mit dem ganzen Rumpf sich seinen Rhythmus geben· [...]« (Quintilianus 2011 [35-60 v.Chr.]: 653f.).

Beispiel Drei: A-rhythmische Bewegungen des Körpers

Dieser verkörperte Rhythmus wird in den folgenden zwei Beispielen (Abb. 4 und Abb. 5) sichtbar. Sie dienen einer differenzierten Hervorhebung der Bewegungen des Körpers (visual highlighting¹¹) und machen den kontinuierlichen, rhythmischen Puls des Körpers zur verbalen Sprache differenzierter erkennbar. Abb. 4, zeigt eine 7 x 7 Feld-Multiplikation des Bildes¹². Die Spalte links außen ist synchron zu den gesprochenen Wörtern und jede weitere folgende Spalte besitzt einen asynchronen Bild-Zeitversatz, der von links nach rechts um +500 Millisekunden pro Spalte ansteigt. Beim Abspielen des Videos werden die Verläufe der Gestenbewegungen (hier beat-Geste), gleich einem Musterverlauf sichtbar. Die Multiplikation des Bildfeldes ermöglicht damit eine simultane Betrachtung der Gestenbewegung mit verschiedenen Zeitversätzen.

¹¹ »Visual highlighting« bedeutet hier, dass die Originalvideos in der Postproduktion mit spezifischen Techniken der Visualisierung bearbeitet worden sind, damit der Rhythmus der körperlichen Bewegungen differenzierter erkennbar wird. Es wurden hierbei Techniken der Bild-Überlagerung, Multiplikation und Outlines verwendet.

¹² Videobeispiele siehe Reutemann (2014).



Abb. 4: Žižek, *Rhythmischer Puls der Schlag-Geste, +500 Millisekunden Bild- und Ton-Versatz pro Spalte*. Originalquelle: Youtube, Big Think (bearbeitet).

Abb. 5 zeigt den Körperumriss des Sprechers. Von Žižek sind keine spezifischen Merkmale bis auf die körperlichen Bewegungen mehr erkennbar. Der Körperumriss ist in Linien mehrfach multipliziert und in unterschiedlichen Farbstufen und Versätzen gleichzeitig sichtbar. Die Bewegungen von Žižek hinterlassen damit eine Farblinien-Verlaufsspur. Die weiße Linie ist synchron zur Audiospur. Der Zeitversatz zur Audiospur ist mit der Helligkeit der blauen Farbe gekennzeichnet. Je dunkler das Blau, desto asynchroner ist das Bild zur Audiospur. Jede noch so minimale Bewegung, und sei es nur ein im Original nicht feststellbares Zucken seiner Schulter wird durch die Verlaufsspur erkennbar.



Abb. 5: Žižek, *Verlaufsspur der Bewegungen*. Originalquelle: Youtube, Big Think (bearbeitet).

Die Linie steht über die gesamte Verlaufszeit des Videos niemals still. Durch die Asynchronität entsteht eine a-rhythmische Distanz zwischen Körper und Wort. Der Rhythmus des Sprechers, welcher sich von den Füßen, über den Oberkörper, dem Kopf bis hin zu den Fingerspitzen des Körper zeigt, ist nicht mehr taktgleich mit der akustischen Stimme, dem Satzbau und der Intonation – schlichtweg allem. Der sichtbare Körper des Sprechers wird in einer verkörperten Darstellung isoliert vom auditiven Wort gezeigt und verliert damit die ursprünglich rhythmisch exakt

synchronisierte Bewegungseinheit von Körper und gesprochener Sprache.

Verkörperte Synchronie aus der Asynchronität

Eine Asynchronität von Bild und Ton wird deutlicher bemerkt und als *irritierender* empfunden, wenn der Ton dem Bild vorausseilt, als umgekehrt (Welch u.a. 1986). Die Erkennung von Asynchronität ist bei Sprachszenen schwieriger als bei distinkten Geräuschen, wie z.B. die eines Hammerschlages. Die gehörten Worte können bis zu 250 Millisekunden rückverschoben zum Bewegtbild stehen, bis eine Asynchronität bewusst bemerkt wird (Grant u.a. 2004). Die Begründung dafür wird durch die biologischen Grundlagen der menschlichen Wahrnehmung gegeben: Aufgrund der unterschiedlichen Geschwindigkeit von Schall und Licht ist es in der Natur häufiger, dass der Ton dem Bild folgt als andersrum (Rudloff 1997). Hinzu kommt, dass der Verarbeitungsprozess des Gehörten in der Aufnahme, Prozessierung und Interpretation von Sinnesreizen grundsätzlich schneller sei als der visuelle Sinn (Daurer 2013). Die Effekte von Asynchronität werden in früheren Studien vor allem mit den Begriffen irritierend oder unbewusst störend beschrieben. Gemäss Reeves und Voelker (1993) sind asynchrone Bilder und Töne von sichtbaren Sprechern »unconsciously disturbing [...] and causes negative evaluation of speakers and material with little or no conscious awareness of the effect« (ebd.:4). Wird davon ausgegangen, dass eine Perzeption vom sichtbaren Sprecher in einer verkörperten Mimesis verarbeitet wird, so könnte diese negative Evaluation des Sprechers auf einen Konflikt von unabhängiger Geste und Wort zurückzuführen sein.

Die Beobachtung von körperlichen Bewegungen ist augenblicklich ein substantieller Bestandteil dessen, was mitgeteilt wird, und dessen was im kognitiven Prozess mit den inneren und äußeren Stimulis zu einer kausalen Einheit in der Verarbeitung prozessiert wird. In einer mimetischen Perzeption können Bewegungen des Gegenübers mit unseren eigenen koordiniert werden. Der Rhythmus eines Sprechers wird relativ leicht vom Hörer adaptiert und die Sprechweise und der Ausdruck passt sich in kurzer Zeit in einer *interactional synchrony* an (McClave 1994; Cummins/Port 1998).

In Bewegtbildern besteht eine »continuity between the physiological and affective responses of my own body and the appearances and disappearances, the mutations and the perdurances, of the bodies and images on the screen« (Shaviro 1994).

Mittlerweile existieren zahlreiche Studien welche den Einfluss von Sprechern auf den Zuhörer aufzeigen (Goldin-Meadow u.a.1999; Spencer u.a. 2004; Wu/Coulson 2005; Özyürek u.a. 2007; Holle/Gunter 2007; Wagner Cook/Tannenhaus 2009; Alibali u.a. 2014). Untersuchungen über die unbewusste Imitation eines Gegenübers dokumentieren, dass Expressionen spontan nachgeahmt werden. Dabei kann das Sehen einer Geste bereits eine Auswirkung auf das weitere Verhalten des Gegenüber haben (Wagner Cook/Tannenhaus 2009). Die zum Teil unbewusste Reaktion des Rezipienten auf Tätigkeiten, Emotionen und Bewegungen des sichtbaren Sprechers zeigt sich bereits beim Blinkeln der Augen welche in einer Frequenz mit einer Verzögerung von 250-500 Millisekunden auf den Augenaufschlag des Publikums überträgt (Nakano/Kitazawa 2010). Der innere, ultradiane Bewegungsrhythmus des Menschen wird durch die Wahrnehmung von äußeren

Rhythmen in ihren Abläufen beeinflusst. Ein schneller Rhythmus kann die inneren Rhythmen von Kreislauf, Atmung und den Sympathikus beschleunigen (Kreidler 1980). Die kontinuierliche, rhythmische Kinetik des sichtbaren Körpers könnte somit, wie Rhythmen der Musik, Einfluss auf die inneren Aktivitäten des Körpers des Rezipienten nehmen (Fraisie 1982; Mikunda 2002). Beat-Gesten können so die Gehirnaktivität von Zuhörern modulieren (Hubbard et al. 2009).

Gleichermaßen vermögen die Atemgeräusche eines Akteurs in Bewegtbildern einen Effekt auf die Atmung des Betrachters auszulösen (Chion 1994). Steigt die Atemfrequenz des Sprechers, so ändert sich damit seine Bewegungsdynamik. Eine gesteigerte Atmungsaktivität aktiviert eine hormonelle Veränderung, wie einen erhöhten Adrenalin Spiegel, wobei sich dieser Zustand wiederum auf die Wahrnehmung, Erinnerung und Aufmerksamkeit auswirkt (Gallagher 2012). Damit ist eine Veränderung der Atmungsaktivität nicht einfach ein Faktum, sondern „etwas, das meine Erfahrung bestimmt. Der rasende Herzschlag, der steigende Adrenalin Spiegel, »[...] - all das sind keine neutralen, objektiven Tatsachen, die zu einer Beschreibung des Körpers gehören, sondern Prozesse, die mein Leib durchlebt; Prozesse, die unsere Wahrnehmung und unser Denken im Moment ihrer Erfahrung entscheidend bestimmen« (Gallagher 2005).

Alle diese Studien deuten darauf hin dass die Perzeption von sichtbaren Sprechern, sei es in der natürlichen Audiovision oder im audiovisuellen Bewegtbild, einen Effekt auf die verkörperte Wahrnehmung haben kann. Die Geste und das Wort werden in einer mimetischen Perzeption in einem somatischen, mentalen und neuronalen Prozess gemeinsam verarbeitet. Asynchronität greift in diesen komplexen, körperlichen Wahrnehmungsprozess ein.

Die Effekte von Asynchronität bei einem sichtbaren Sprecher sind damit nicht nur *irritierend* oder *unbewusst störend*, sondern es besteht Grund zur Annahme, dass diese Irritation oder Störung aus einer Fehlinterpretation des Bedeutungszusammenhangs von Körper und Wort entsteht.

Die versetzte Repräsentation der Gesten zur gesprochenen Sprache wird jedoch nicht isoliert voneinander verarbeitet. Sondern die inkongruenten Reize von Körper und Wort werden vielmehr, ähnlich wie bei der Phonemergänzung, in der Wahrnehmung des Rezipienten in einer Synchrese neu zusammengefügt. Das Schmieden der Synchrese könnte im Unterschied zur natürlichen Audiovision darin bestehen, dass die natürliche Sprachwahrnehmung aus Gesten und Worten besteht, die stets eine gewisse Kongruenz aufweisen, während bei einer Synchrese im asynchronen audiovisuellen Bewegtbild eine Konstruktion von Geste und Wort aus inkongruenten Gegebenheiten geschieht. Es wird zwar in einer anthropologisch determinierten Konstruktion von Kausalität versucht, in einer mimetischen Perzeption aus den inkongruenten Reizen des Bewegtbildes eine Einheit des Körpers und der verbalen Sprache zu schmieden; diese erhalten jedoch eine unvermeidbare Bedeutungsveränderung im Sinne einer Geste-Wort-Schere wie bei Žižek's *the system*, und bei der Zeigegeste Tomasello's oder einen Bedeutungsverlust des Körperhythmus', wie er in der Multiplikation und den Outlines von Žižek sichtbar wird.

Zusammenfassung

Durch die kognitive Verarbeitung und die zwanghafte Neigung, Bedeutung zu erzeugen, entsteht eine Synchronie aus der Asynchronität. Der ursprüngliche Bedeutungsgehalt der Sprachkommunikation kann damit jedoch nicht rekonstruiert werden. Folglich verliert der Sprecher im asynchronen Bewegtbild nicht nur seinen Bedeutungszusammenhang, sondern das, was durch den Körper übermittelt werden sollte wird in einer neuen, aus inkongruenten Informationen bestehenden Beobachtung synchronisch mit den gehörten Worten zu einer verzerrten Bedeutung zusammengefügt.

Zwischen dem medial-asynchronen Bild und Ton und der kognitiven Prozessierung der Reize wird versucht, eine synchronische, kausale Bedeutungsinterpretation zu erstellen. Der Versatz kann dabei entweder, wie bei Žižek's *the system* oder Tomasello's Zeigegeste, in einer Geste-Wort-Schere zu einer Bedeutungsveränderung führen oder es kann durch den Versatz ein Bedeutungsverlust von Rhythmus und Körper entstehen, wie bei dem zweiten Beispiel mit Žižek. Es resultiert ein Bedeutungskonflikt von Gesehenem und Gehörtem, da durch die *verkörperte* Kognition der Widerspruch aktiv erlebt wird, während gleichzeitig eine kausale Synchronie geschmiedet wird. Gebrochene Bild und Ton Synchronität und die daraus folgenden synchronischen Konstruktionen sind ein vielseitiger Forschungsgegenstand für Gestenforschung, Theorien der Verkörperung und Medienwissenschaften, der in seiner Komplexität in dieser Arbeit nur angeschnitten werden konnte. Gerade die Erforschung des Abweichenden erlaubt Differenzen und Kontraste zu erkennen, die in der regulären Form der audiovisuellen Wahrnehmung im Verborgenen bleiben.

Literatur

Alibali, Martha W./Boncoddio, Rebecca/Hostetter, Autumn B (2014). An Embodied Perspective, in: Zdravko Radman (Hg.), *The Routledge Handbook of Embodied Cognition*, Cambridge, S.150.

Alibali, Martha W./ Nathan. Mitchell J. (2012), Embodiment in Mathematics Teaching and Learning: Evidence From Learners' and Teachers' Gestures, in: *Journal of the Learning Sciences* 21, no. 2, S.247–286.

Anderson, Joseph D. (1996), *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*, Southern Illinois University Press.

Boehm, Gottfried (2010), *Wie Bilder Sinn Erzeugen*, Berlin, S. 25.

Chion Michel (1999), *The Voice in the Cinema*, New York.

Chion, Michel (1994), *Audio-Vision*, New York.

Chion, Michel (1990), *L'Audio-Vision*, Paris.

Church, R. Brechinridge, & Goldin-Meadow, Susan (1986). The mismatch between gesture and speech as an index of transitional knowledge. In: *Cognition* 23, S. 43–71.

Church, R. Breckinridge/Spencer, Kelly/Holcombe, David (2014), Temporal Synchrony Between Speech, Action and Gesture During Language Production, in: *Language, Cognition and Neuroscience* 29, no. 3, S. 345-354.

Clark, Andy (2013), Gesture as thought, in: *The Hand, an Organ of the Mind: What the Manual Tells the Mental*, in: Zdravko Radman (Hg.), *The Routledge Handbook of Embodied Cognition*, Cambridge,

S.255–268.

Clark, Andy (1999), Where brain, body, and world collide. In: *Journal of Cognitive Systems Research*, S.15–17.

Clark, Herbert H (1996), *Using language*. Cambridge.

Cummins, Fred/Port, Robert (1998), Rhythmic constraints on stress timing in English, in: *Journal of Phonetics*, 26(2), S. 145-171.

Condon, William S. (1986), Communication: Rhythm and structure. In: Evans/Clynes (Hg.): *Rhythm in Psychological, Linguistic and Musical Processes* Springfield, Ill., S.55–77.

Condon, William S./Ogston, Richard (1966), Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease*, 143, S. 338-347.

Cutting, James (2005), Perceiving Scenes in Film and in the World, in *Moving Image Theory: Ecological Considerations*, Joseph D. Anderson/Barbara Fischer Anderson (Hg.) Southern Illinois University Press, S. 9-27.

Daurer, Gerhard (2013), Audiovisuelle Wahrnehmung, 07.02.2014, In: <http://www.see-this-sound.at/drucken/33>.

Ekman, Paul/Friesen, Wallace (1976), Body movement and voice pitch in deceptive interaction, In: *Semiotica*, 16, S. 23-27.

Ekman, Paul/Friesen, Wallace (1969), The repertoire of nonverbal behavior: Categories, origins, usage, and coding. in: *Semiotica*, 1, S.49-98.

Efron, David (1941), *Gesture and Environment*. New York. Republished 1972: *Gesture, Race, and Culture*. Den Haag.

Fraisse, Paul (1982), Rhythm and Tempo, in: *The Psychology of Music*, New York, S. 149-180.

Gallagher, Shaun (2012), Kognitionswissenschaften – Leiblichkeit und Embodiment, in: Alloa, Emmanuel, Thomas Bedorf, Christian Grüny, Tobias Nikolaus Klass, *Leiblichkeit, Geschichte und Aktualität eines Konzepts*, Tübingen, S. 321-333.

Gallagher, Shaun (2005), *How the Body Shapes the Mind*. Oxford.

Gibbs, Raymond W. Jr. (2006), *Embodiment and cognitive science*. Cambridge.

Goldin-Meadow, Susan/ Kim, San/ Singer, Melissa (1999), What the teacher's hands tell the student's mind about math, in: *Journal of Educational Psychology* 91, no. 4. S. 720.

Goldin-Meadow, Susan (1997), When gestures and words speak differently. *Current Directions, in Psychological Science*, 6(5), S. 138-143.

Goldin-Meadow, Susan/Wagner Alibali, Martha/Breckinridge Church, R. (1993), Transitions in Concept Acquisition: Using the Hand to Read the Mind. In: *Psychological Review & American Psychological Association Inc.* 100. S. 279–97.

Goldstein, E. Bruce (2007), *Wahrnehmungspsychologie* (7 ed.). Berlin Heidelberg.

Grant, Ken W/ van Wassenhove, Virginie/Poeppl, David (2004), *Detection of Auditory (Cross Spectral) and Auditory-Visual (Cross Modal) Synchrony*, S. 1–20.

Hostetter, Autumn B., and Alibali, Matha W. (2008). Visible embodiment: gestures as simulated action. In: *Psychonomic Bulletin & Review* 15, no. 3, S.495–514.

- Holle, Henning/ Gunter, Thomas C. (2007), The Role of Iconic Gestures in Speech Disambiguation: ERP Evidence, in: *Journal of Cognitive Neuroscience* 19, no. 7, S. 1175–92.
- Kendon, Adam (2004). *Gesture*, Cambridge: Cambridge University Press.
- Kendon, Adam (1972), Some relationships between body motion and speech., in: Aaron Siegman/Benjamin Pope (Hg.), *Studies in dyadic communication* 7, New York, S. 177-210.
- Kreitler, Hans (1980), *Psychologie der Kunst*, Stuttgart.
- Leonard, Thomas/Cummins, Fred, The Temporal Relation Between Beat Gestures and Speech, in: *Language and Cognitive Processes* 26, no. 10, S.1457-1471.
- Loehr, Daniel (2007), Aspects of Rhythm in Gesture and Speech, in: *Gesture* 7, no. 2, S.179-214.
- Mikunda, Christian (2002). *Kino Spüren*. Wien.
- McClave, Evelyn Z (2000), Linguistic Functions of Head Movements in the Context of Speech, in: *Journal of Pragmatics* 32, no. 7, S. 855-878.
- McClave, Evelyn Z. (1994), Gestural Beats: the Rhythm Hypothesis, in: *Journal of Psycholinguistic Research*, Vol. 23, No.1, S. 45-66.
- McGurk, Harry/MacDonald, John (1976) Hearing Lips and Seeing Voices, in: *Nature*, 264, S. 746–748.
- McNeill, David (1992) *Hand and Mind*. Chicago.
- McNeill, David (2005), *Gesture and thought*. Chicago.
- McNeill, David (2008), Gesture: a Psycholinguistic Approach, in: *Psycholinguistic Section, the Encyclopedia of Language and Linguistics*, February 14, S. 1–15.
- McNeill, David (2012), *How Language Began: Gesture and Speech in Human Evolution*. Vol. 8, Cambridge, New York, Melbourne, Madrid, Cape Town.
- Müller, Cornelia (2008): „Grammatik der Gesten“. Interview vom 30.3.2008 der Neuen Zürcher Zeitung, 22.1.2015, In: <http://www.nzz.ch/aktuell/startseite/grammatik-der-gesten-1>.
- Nakano, Tamami/ Kitazawa, Shigeru (2010), Eyeblink Entrainment at Breakpoints of Speech, in: *Experimental Brain Research* 205, no. 4, S. 577–81.
- Nuñez, Rafael (2005), Do real numbers really move? Language, thought, and gesture: The embodied cognitive foundations of mathematics. In Ida/Pfeifer/Steels/Kuniyoshi (Hg.), *Embodied artificial intelligence*, Berlin, S. 54-73.
- Özyürek, Asli/Willems, Roel M./ Kita, Sotaro/Hagoort, Peter (2007), On-Line Integration of Semantic Information From Speech and Gesture: Insights From Event-Related Brain Potentials, in: *Cognitive Neuroscience, Journal of* 19, no. 4, S. 605–16.
- Pouw, Wim/de Nooijer, Jacqueline A./van Gog, Tamara/Zwaan, Rolf A./Paas, Fred (2014), Toward a More Embedded/Extended Perspective on the Cognitive Function of Gestures. In: *Frontiers in Psychology* 5. Article 359, S.1-14.
- Quintilianus, Marcus Fabius (2011) [35-60 v.Chr.], *Ausbildung des Redners*, Helmuth Rahn (Hg.). Darmstadt.

- Reeves, Bryon/ Voelker, David (1993), *Effects of Audio-Video Asynchrony on Viewer's Memory, Evaluation of Content and Detection Ability*, S. 1–22.
- Reutemann, Jeanine (2014), Videobeispiele zur Publikation von Slawoj Žižek und Michael Tomasello, 15.04.2015, In: <http://www.audiovisualresearch.org/gesture-and-speech/case-studies>.
- Rudloff, Ingo (1997), *Untersuchungen Zur Wahrgenommenen Synchronität Von Bild Und Ton Bei Film Und Fernsehen*, Bochum.
- Shaviro, Steven (1994), *The Cinematic Body*, Minneapolis.
- Smith, Tim J. (2012), The Attentional Theory of Continuity Editing, *Projections* 1, S. 1-27.
- Spencer, Kelly D./Özyürek, Asli/Maris, Eric (2010), Two Sides of the Same Coin Speech and Gesture Mutually Interact to Enhance Comprehension." In: *Psychological Science*, S. 260-267.
- Spencer, Kelly, D./ Kravitz, Corinne/Hopkins, Michael (2004), Neural Correlates of Bimodal Speech and Gesture Comprehension, in: *Brain and Language* 89, no. 1, S.253-60.
- Slawoj, Žižek (2012), Originalvideo des Interviews von Slawoj Žižek, Don't act. Just think, von der Organisation Big Think, 06.02.2014, in: <https://www.youtube.com/watch?v=IqR6uaVqWsQ>.
- Tomasello, Michael (2008a), Ein Interview mit 3Sat von einem privaten User veröffentlicht, 13.07.2014, In: <https://www.youtube.com/watch?v=2A5t8FcWGas>.
- Tomasello, Michael (2008b) *Die Ursprünge Der Menschlichen Kommunikation*. Frankfurt/M.
- Treffner, Paul/Peter, Mira/Kleidon, Mark (2008), Gestures and Phases: the Dynamics of Speech-Hand Communication, in: *Ecological Psychology* 20, no. 1, S. 32-64.
- Tuite, Kevin (1993), The Production of Gesture, in: *Semiotica*, 1993, S. 83-105.
- Wagner Cook, Susan/Tanenhaus, Michael K (2009), Embodied Communication: Speakers' Gestures Affect Listeners' Actions, in: *Cognition* 113, no. 1, S. 98–104.
- Warren, Richard M. (1970), Perceptual Restoration of Missing Speech Sounds, in: *Science* 167, no. 3917, S. 392-393.
- Welch, Robert B./Dutton Hunt, Lance/ Warren, David H (1986), Contributions of Audition and Vision to Temporal Rate Perception, in: *Perception & Psychophysics*, 39 (4), S. 294–300.
- Wember, Bernward (1976), *Wie informiert das Fernsehen?* München.
- Wu, Ying Choon/Coulson, Seana (2005), Meaningful Gestures: Electrophysiological Indices of Iconic Gesture Comprehension, in: *Psychophysiology* 42, no. 6, S. 654–67.
- YouTube Statistik, Datenanalyse der Benutzer, 12.03.14, <https://www.youtube.com/yt/press/de/statistics.html>.
- Zacks, Jeffrey M./Magliano, Joseph, Film (2011), Narrative, and Cognitive Neuroscience, in *Art and the Senses*, Francesca Bacci/David Melcher (Hg.), New York, S. 435-54.
- Zimbardo, Philip George/Gerrig, Richard J. (2008), *Psychologie*, München.