

## 2 Theoretische Grundlagen, Forschungsüberblick und Datenerhebung

### 2.1 Theoretischer Hintergrund

Das folgende Kapitel widmet sich den theoretischen und methodischen Grundlagen dieser Arbeit. Dabei geht es zunächst darum, aufzuzeigen, wie eine Probe definiert werden kann und worin sie sich gegenüber Aufführungen abgrenzt (Kap. 2.1.1). Außerdem wird das institutionelle Interaktionssetting der Orchesterprobe eingehender reflektiert, ein prototypischer Ablauf einer Orchesterprobe skizziert sowie die Überlappung von mehreren zeitlichen Dimensionen in Orchesterproben veranschaulicht (Kap. 2.1.2). In einem zweiten Schritt gilt es, die vorliegende Arbeit im aktuellen Forschungszusammenhang zu verorten (Kap. 2.1.3 bis 2.1.7). Zu den theoretischen Ausgangspunkten zählen die ethnomethodologische Konversationsanalyse, Sprecherwechsel-Systeme (bzw. „*Turn-taking*-Prozeduren“ aus konversationsanalytischer Perspektive) und ihre Veräußerlichung in Orchesterproben, die Weiterentwicklung der ‚klassischen‘ Konversationsanalyse zur multimodalen Konversationsanalyse, die Gegenüberstellung von Dirigiergesten und sprach- oder redebegleitenden Gesten sowie Höflichkeit und Face.<sup>3</sup> Die Kombination verschiedener Theorieansätze soll der Komplexität der Interaktion in Orchesterproben Rechnung tragen. Diese Komplexität zeichnet sich dadurch aus, dass in der Orchesterprobe unterschiedliche Modalitäten aufeinandertreffen (etwa Sprache, Musik, Gestik usw.) und dass es einen ständigen Wechsel zwischen Besprechungssteilen (der/die Dirigent:in spricht, die Musiker:innen hören zu) und Spielteilen (die Musiker:innen spielen, der/die Dirigent:in hört zu und dirigiert) gibt. Nicht zuletzt muss sich der/die Dirigent:in laufend mit der Frage auseinandersetzen, ob und wie er:sie gesichtsbedrohende Handlungen ausführen oder unterlassen soll. In der vorliegenden Arbeit soll – der methodischen Vorgehensweise der Konversationsanalyse folgend – eine Brücke zwischen den unterschiedlichen theoretischen Ansätzen geschlagen werden. Anschließend werden Studien besprochen, die sich der Kommunikation in Orchesterproben aus unterschiedlichen theoretischen Richtungen genähert haben und auf deren Ergebnisse diese Arbeit zurückgreifen kann (Kap. 2.2). Der letzte Teil des Kapitels reflektiert schließlich den methodischen Zugang zum Datenmaterial, befasst sich mit der Datenerhebung und stellt die empirische Basis der Studie vor (2.3).

---

<sup>3</sup> Mit den Themen Mehrsprachigkeit und Instruktionen beschäftigen sich die einleitenden theoretischen Abschnitte in den jeweiligen empirischen Kapiteln dazu (Kap. 3 und Kap. 5).

### 2.1.1 Was sind Proben?

Die Probe bewegt sich [...] in einem Spannungsfeld  
zwischen Prüfen und Versuchen,  
zwischen dem Erproben von Möglichkeiten und der Aufführung einer Leistung,  
zwischen einer individuellen Praxis und dem Hervorbringen einer szenischen Darstellung.  
Annemarie Matzke<sup>4</sup>

Der Begriff der Probe leitet sich von lat. *proba* ab, mit der Bedeutung eines Versuchs, eines Experiments, einer Prüfung oder auch einer Untersuchung. Dieses Prüfen und Versuchen sind in dem einleitenden Zitat von Matzke zu finden, die sich aus theaterwissenschaftlicher Perspektive mit Proben auseinandersetzt. Laut Matzke (2012: 20) kann eine Probe als ein Prozess des Suchens beschrieben werden, in dem unterschiedliche Möglichkeiten ausprobiert werden. Dieses Experimentieren mit unterschiedlichen Möglichkeiten bringt Wiederholungen mit sich: Im Theater und auch im Orchester gibt es Praktiken der Wiederholung, die zur Aufführungsreife eines Theaterstücks oder eines Orchesterwerks führen. Proben können damit als Prozesse der Generierung, Sicherung und Verfügbarmachung von Wissen aufgefasst werden (Matzke 2011: 133). Das, was in der Probe geprobt wird, soll möglichst genau und in jeder Aufführung wiederholbar sein. Außerdem haben Proben immer etwas mit Arbeit und mit Leistung zu tun, die Künstler:innen erbringen müssen. *Duden online* definiert eine Probe als „vorbereitende Arbeit (der Künstler[innen]) vor einer Aufführung“ und auch nach dem Theaterwissenschaftler Jens Roselt (2011) ist die Probenphase ein räumlich und zeitlich kontingenter Arbeitsbereich. In der Probe wird an etwas gearbeitet, es wird etwas produziert, es wird beobachtet und interagiert. Damit wird der Blick auf den Akt und die Ästhetik des Produzierens und des Leistens gerichtet und weniger auf das mögliche Ergebnis dieses Prozesses (vgl. Matzke 2012: 18). Matzke (2012: 100–103) differenziert außerdem zwischen Spielvorlage, Probe, Inszenierung und Aufführung. In einer Probe werden Inszenierungen anhand von Spielvorlagen erarbeitet. Vorlagen werden in der Probe zur Anschauung gebracht, umgekehrt modifizieren Proben auch immer bestehende Vorlagen. Eine Probe ist nach Matzke (2012: 105) „immer auch eine Aufführung: Jemand spielt, während ein anderer zuschaut“. In Proben werden also Teile der Aufführung bereits vorweggenommen. Diese Auffassung eines wechselseitigen Bezugs zwischen Probe und Aufführung vertritt auch Roselt (2011: 35): „Die Probe weist hin auf die Aufführung, die aus ihr hervorgehen soll, wie in dieser gegenwärtig bleibt, dass sie das [...] Ergebnis eines Prozesses ist.“

---

<sup>4</sup> Vgl. Matzke (2012: 98).

Neben dieser theaterwissenschaftlich ausgelegten Herangehensweise an Probe und Aufführung ist für die vorliegende Studie auch das Verständnis von Proben aus soziologischer Perspektive relevant. Der Soziologe und Wegbereiter der Interaktionsanalyse Erving Goffman vergleicht bereits in einem seiner frühesten Werke, „Wir alle spielen Theater“ (1959), menschliche Interaktion mit dem Theater. Wenn Personen miteinander interagieren, spielen sie Rollen, die sie zuvor auf der Hinterbühne (ein Ort, der für die anderen Interaktionsbeteiligten nicht zugänglich ist) eingeübt haben und nun auf der Vorderbühne, in der stattfindenden Kommunikationssituation, präsentieren. Diese Rollen können je nach Interaktionspartner:in und Interaktionssituation variieren, z. B. unterscheidet sich die Rolle, die jemand einnimmt, wenn er:sie mit seinem/ihrer Vorgesetzten spricht, von der, in die jemand schlüpft, wenn er:sie mit Freund:innen ein Gespräch führt. Das heißt, soziales Handeln wird in dieser soziologischen Perspektive mit Schauspiel gleichgesetzt, das auch ein gewisses Proben nach sich zieht. Allerdings kann der/die Einzelne nicht steuern, wie soziales Handeln und Interagieren verlaufen wird, unabhängig davon, wie gut er:sie seine/ihrer Rolle vorbereitet und darstellt. Den Unterschied zwischen Bühne und Leben definiert Goffman (1959: 3) folgendermaßen: „Auf der Bühne werden Dinge vorgetäuscht. Im Leben hingegen werden höchstwahrscheinlich Dinge dargestellt, die echt, dabei aber nur unzureichend erprobt sind.“

In seiner Rahmenanalyse (1977) setzt sich Goffman eingehender mit Proben auseinander und beschreibt sie als Modulationen. Wenn etwa eine Hochzeitszeremonie im Rahmen einer Probe ausgeführt wird, so wird diese Zeremonie nachgebildet, sie repräsentiert die Handlung in darstellender Funktion. Laut Goffman (1963) findet in diesem Fall fokussierte, zentrierte Interaktion statt, d. h. es handelt sich um eine Interaktion mit mehreren Beteiligten, die auf ein bestimmtes Ziel ausgerichtet ist. Goffman (1963) unterscheidet innerhalb einer sozialen Situation (*social situation*), d. h. der Umgebung, in der ein Individuum aufgrund wechselseitiger Wahrnehmung Teil einer ko-präsenten Zusammenkunft (*gathering*) wird, zwei kommunikative Verhaltensweisen: die nicht-fokussierte (*unfocused*) sowie die fokussierte (*focused*) Interaktion. Eine nicht-fokussierte Interaktion zeichnet sich durch das „management of sheer and mere copresence“ (Goffman 1963: 24) aus, wobei Beteiligte nicht direkt miteinander kommunizieren (z. B. wenn sich mehrere Personen im Warteraum beim Arzt aufhalten). Dahingegen beschreibt die fokussierte Interaktion eine Kommunikationssituation, in der zwei oder mehrere Beteiligte gemeinsam anwesend sind und – meist unter Einbezug einer Gesprächsaktivität – einen gemeinsamen Aufmerksamkeitsfokus, ein gemeinsames (handlungsorientiertes) Ziel teilen. Am Beginn einer fokussierten Interaktion steht immer die Herstellung eines sozialen Arrangements (*encounter*) zwischen den Beteiligten durch eine körperliche Hinwendung (vgl. „bodily addressing one another“ bei Goffman 1981: 130).

Auch Goffman unterscheidet zwischen Probe und Aufführung: Er versteht Proben als Demonstrationen oder Vorführungen, d. h. als Ausführung einer normalen Tätigkeit zum Zweck der Vorführung. Um ein Beispiel Goffmans (1977: 79) zu nennen: Ein Vertreter führt vor, wie ein Staubsauger den Schmutz aufsaugt, den er vorher zu Demonstrationszwecken auf den Boden gestreut hat. Das heißt, die entsprechende Handlung wird nur zur Probe ausgeführt und nicht im Rahmen einer tatsächlichen Aufführung. Eine Aufführung zeichnet sich nach Goffman (1977: 143–175) dadurch aus, dass ihr ein Publikum beiwohnt, das nicht Teil der Probe und des Probierens ist. Proben können demnach in dieser soziologischen und rahmenanalytischen Perspektive als Vorgänge bezeichnet werden, „die sinntypologisch in einer spezifischen Weise ausgewiesen sind, nämlich als Aktivitäten, die zum Zwecke des Einübens, Vorführens und Demonstrierens ausgeführt werden“ (Krug et al. 2020: 162).

Der Fokus der vorliegenden Arbeit liegt auf professionellen Proben im künstlerischen Arbeitskontext des Orchesters. Diese Proben wurden offiziell anberaumt und sind verpflichtend für alle relevanten Beteiligten an der jeweiligen Produktion. Orchesterproben reihen sich – Goffman folgend – unter fokussierte Interaktion ein, da die daran Beteiligten auf ein bestimmtes Ziel hinarbeiten (ein Stück so lange proben, bis es aufführungsreif ist), dabei miteinander kommunizieren und ihr Handeln darüber hinaus mit gewissen Pflichten verknüpft ist (z. B. ist der/die Dirigent:in in der Orchesterprobe verantwortlich für die künstlerische Leitung eines musikalischen Projekts).

### **2.1.2 Wie sieht eine Orchesterprobe aus?**

From the moment the conductor steps on the podium  
a special world is in the process of being constructed.

Robert R. Faulkner<sup>5</sup>

#### **2.1.2.1 Die Orchesterprobe als institutionelles Interaktionssetting**

Die Orchesterprobe stellt den Arbeitsplatz für Dirigent:in und Musiker:innen dar und kann damit als Schauplatz institutioneller Kommunikation und Interaktion bezeichnet werden. Nach Drew & Heritage (1992: 22–25) zeichnet sich institutionelle Interaktion durch folgende Faktoren aus (vgl. auch Heritage 2005):

- a) Sie ist auf ein bestimmtes, vorab festgelegtes Ziel ausgerichtet; dieses Ziel ist mit den institutionsrelevanten Rollen verknüpft, z. B. verfolgt die Kommuni-

---

<sup>5</sup> Vgl. Faulkner (1973: 149).

kation im Klassenzimmer zwischen Lehrer:in und Schüler:in einen pädagogisch-didaktischen Zweck.

- b) Sie findet auf Basis eines spezifischen Sprecherwechselsystems statt, das gewisse Asymmetrien und Restriktionen hinsichtlich der diskursiven Rechte und Pflichten, d. h. der Rederechtsverteilung und der Äußerungskonstruktion sowie -produktion, nach sich ziehen kann. In einem Interview beispielsweise stellt der/die Interviewer:in Fragen, der/die Interviewte gibt Antworten.
- c) Institutionelle Interaktion richtet sich an aktivitätsspezifischen Inferenzen (*inferential frameworks*) aus, die sich beispielsweise in der Anwendung einer bestimmten Fachsprache oder in einer spezifischen thematischen Orientierung äußern können.

Auch in der Orchesterprobe lässt sich diese Spezifik institutioneller Interaktion beobachten. Dort erarbeiten der/die Dirigent:in und die Musiker:innen gemeinsam ein musikalisches Werk, das zu einem späteren Zeitpunkt vor einem Publikum aufgeführt wird. Dabei sind die jeweiligen Funktionsrollen, jene des/der Dirigent:in und jene der Musiker:innen, vorab festgelegt. Mit diesen komplementären Rollen gehen bestimmte Erwartungen hinsichtlich kategoriengebundener Aktivitäten und rollenspezifischer Äußerungstypen einher: Der/die Dirigent:in dirigiert, gibt Feedback und strukturiert die Probe; die Musiker:innen spielen und setzen die Anweisungen des/der Dirigent:in auf ihren Instrumenten um. Das heißt, der/die Dirigent:in hat andere diskursive Rechte und Pflichten als die Musiker:innen.<sup>6</sup> Diese Rechte und Pflichten sind vor allem die Äußerungsproduktion und -konstruktion sowie das Sprecherwechselsystem betreffend asymmetrisch verteilt. Der/die Dirigent:in ist in der Probe derjenige/diejenige, der/die das Rederecht inne hat. Er:sie kann die Musik an jedem erdenklichen Punkt unterbrechen und zu einer Besprechung der Musik ansetzen. In solchen Besprechungsphasen kann er:sie entweder das gesamte Orchester als Kollektiv ansprechen, spezifische Instrumentengruppen adressieren oder sich auch nur mit einzelnen Musiker:innen austauschen. Außerdem weist er:sie nach Abschluss einer Besprechungsphase durch verbale und körperliche Zeichen den (musikalischen) Turn erneut den Musiker:innen zu. Damit strukturiert der/die Dirigent:in gleichzeitig eine Probe und entscheidet, wann gespielt und wann besprochen wird.

Neben dem/der Dirigent:in kommt auch dem/der Konzertmeister:in in der Orchesterprobe eine spezifische Rolle beim Spielen und beim Sprechen zu. Der/die Konzertmeister:in ist der/die am ersten Notenpult außen sitzende Stimmführer:in der Gruppe der 1. Violinen. In der Hierarchie eines Orchesters folgt er:sie

---

<sup>6</sup> Vgl. auch das Kapitel zu Sprecherwechselsystemen (Kap. 2.1.4).

gleich nach dem/der Dirigent:in. Seine/ihr zentrale Bedeutung zeigt sich zum einen dadurch, dass er:sie zu Beginn einer Probe – und auch eines Konzerts – das Orchester nach dem eingestrichenen A der Oboe stimmt. Zum anderen erweist sich sein/ihr sozialer Status darin, dass er:sie als Primus inter Pares aller Streichstimmführer:innen, der Solobläser:innen, der Soloharfe und der Solopauke während der Probe sowohl mit dem/der Dirigent:in als auch mit den anderen Orchester-solist:innen interagiert. Außerdem hat er:sie Mitspracherecht, wenn es um Phrasierung, Stricharten, Legatobindungen, Artikulation, Lautstärke und Fingersätze geht (vgl. Mahling 1996). Das heißt, die diskursiven Rechte und Pflichten des/der Konzertmeister:in sind vergleichbar mit jenen des/der Dirigent:in. Der/die Konzertmeister:in kann sich in der Probe zu Wort melden, das Rederecht ergreifen und mitbestimmen, wie bestimmte Stellen im Stück gespielt werden sollen.

Die Asymmetrie zwischen Dirigent:in und Musiker:innen spiegelt sich auch in den Artefakten wider. Die Interaktion zwischen Dirigent:in und Musiker:innen in der Probe – und auch dem/der Konzertmeister:in in bestimmten Kommunikations-situationen – ist immer an der Partitur und den respektiven Teilen in den Noten der Musiker:innen ausgerichtet. Die Partitur steht im Fokus der Arbeit und ist unumgänglich für die Organisation der gemeinsamen Arbeit. Es findet eine Metakommunikation statt, die sowohl das in der Partitur festgeschriebene als auch das Gespielte umfasst. Der/die Dirigent:in ist dabei derjenige/diejenige, der/die in der Partitur alle Stimmen gebündelt vor sich hat. Im Gegensatz dazu arbeiten/spielen die Musiker:innen nur mit ihren Noten, in denen ihre jeweilige Stimme vermerkt ist.<sup>7</sup> Dadurch hat der/die Dirigent:in epistemische Autorität hinsichtlich des Gesamtklangs sowie der Verteilung und des Einsetzens der einzelnen Stimmen.

Nicht zuletzt interagieren die Beteiligten an einer Orchesterprobe auf der Basis von Aktivitätsspezifischen Inferenzen, die an unterschiedliche Faktoren geknüpft sind, etwa die thematische Orientierung (das gemeinsame Musikmachen), die jeweiligen Verantwortlichkeiten (der/die Dirigent:in dirigiert und weist an, die Musiker:innen spielen und setzen die Anweisungen um), die Rederechtsverteilung (der/die Dirigent:in spricht, die Musiker:innen sprechen, wenn sie vom/von der Dirigent:in dazu aufgefordert werden) sowie die Anwendung einer bestimmten Fachsprache (z. B. Italienisch für die Musik-Terminologie).

Allerdings sind diese Dimensionen institutioneller Interaktion nicht von vornherein durch den Kontext festgeschrieben, sondern die Handlungen der Beteiligten bringen in reflexiver Weise diesen spezifischen institutionellen Kontext hervor, in welchem sie selbst als sinnvoll und angemessen wahrgenommen werden. Kommu-

---

<sup>7</sup> Allerdings konnte in einigen Proben beobachtet werden, dass auch Musiker:innen die Partitur aufliegen hatten.

nikative Handlungen werden vom jeweiligen Kontext beeinflusst (*context shaped*) und konstituieren diesen fortlaufend mit (*context renewing*) (vgl. Duranti/Goodwin 1992). In der Orchesterprobe reproduzieren der/die Dirigent:in und die Musiker:innen ständig die institutionelle Interaktionssituation, indem sie die Vorgaben des Kontexts ausführen. Es können allerdings Abweichungen eintreten, etwa wenn ein/e Musiker:in selbst die Redeinitiative ergreift. Tauchen solche Abweichungen vermehrt auf, kann dies auch zu einer Änderung der Regeln führen. Das heißt, wenn sich Musiker:innen in Proben öfters zu Wort melden ohne vom/von der Dirigent:in dazu aufgefordert zu werden, kann dieses interaktionale Verhalten Teil des Kontexts werden und diesen mit herstellen.

### 2.1.2.2 Der/die Dirigent:in als Fokusperson

Wie bereits weiter oben angeführt, stellt der/die Dirigent:in in der Orchesterprobe die Bezugsperson dar, an der die Musiker:innen ihr Handeln kontinuierlich ausrichten. Der/die Dirigent:in ist „the focal participant in the ensemble's effort“ (Faulkner 1973: 149) bzw. eine „Fokusperson“ (Schmitt/Deppermann 2007: 110), die ihr eigenes und das Verhalten der Musiker:innen koordiniert und damit über eine gewisse Autorität gegenüber den Musiker:innen verfügt. Dieses Autoritätsverständnis in der Orchesterprobe hängt nach Schütz (1964, 1972: 147–149) mit dem Wissen über gemeinsames Musizieren (*making music together*) zusammen. Demnach gibt es Akteur:innen, denen eine größere Autorität in ihren Entscheidungen und in ihrem Wissen über musikalische Interpretation (*epistemic authority*) zugeschrieben wird als anderen (im Fall der Orchesterprobe dem/der Dirigent:in). Nach Schütz (1964, 1972) kann Autorität als Wissen darüber beschrieben werden, wie gehandelt (*how to act*) bzw. wie gespielt werden soll (*how to play*).

Dieses asymmetrische Verhältnis spiegelt sich auch in der Anordnung der Beteiligten an einer Orchesterprobe wider. Der/die Dirigent:in positioniert sich vor den Musiker:innen, meistens zudem erhöht auf einem kleinen Podest. Er:sie ist die einzige Person, die vorne steht, während die Musiker:innen in einem Halbkreis um ihn/sie angeordnet sind.<sup>8</sup> Diese räumliche Anordnung kann als *F-formation* beschrieben werden. Nach Kendon (1990: 209) formiert sich eine solche Interaktionskonstellation immer dann, „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“. Mondada (2007b) hebt den Beitrag Kendons hinsichtlich Aufmerksamkeitsorganisation in der Interaktion hervor:

---

<sup>8</sup> Für die Anordnung der Teilnehmer:innen und damit einhergehend das spezifische Teilnehmer:innen-Format in einer Orchesterprobe vgl. auch Kap. 3.7.2.3.

Er [Kendon] betrachtet räumliche Positionierung und Körperausrichtung als Ressource der Teilnehmer zur Bildung eines Arrangements, das für einen gemeinsamen Aufmerksamkeitsfokus und den Eintritt in eine als solche anerkannte gemeinsame Aktivität geeignet ist. Dieses Arrangement verändert sich oder löst sich auf, wenn die Aktivität sich ändert oder beendet wird. (Mondada (2007b: 59–60)

In der Orchesterprobe verfolgt die Anordnung der Beteiligten den Zweck, dass sich die Aufmerksamkeit für die Aktivität des Musikmachens optimal ausrichten und darüber hinaus auch gesteuert werden kann. Die Musiker:innen können ihre Aufmerksamkeit den Noten zuwenden und gleichzeitig das Handeln des/der Dirigent:in – wie Schmitt & Deppermann (2007: 110) es für die Fokuspersönlichkeit beschreiben – „kontinuierlich auf potenzielle Implikationen für ihr eigenes Handeln beobachten und ausdeuten“. Ebenso kann der/die Dirigent:in aus seiner/ihrer Position heraus die Gruppe der Orchestermusiker:innen führen und anleiten und sich dabei an der Partitur orientieren. Dieser doppelte Aufmerksamkeitsfokus auf räumlich-struktureller und auf personaler Ebene bringt allerdings auch eine gewisse Schwierigkeit mit sich, da der/die Dirigent:in und die Musiker:innen in der Lage sein müssen, unterschiedliche Subjekte und Objekte fast gleichzeitig zu fixieren.

### **2.1.2.3 Ablauf einer Orchesterprobe**

Ein Orchester probt für ein musikalisches Projekt zumeist eine Woche, am Ende der Woche gibt es eine oder auch mehrere Aufführungen. Proben von Symphonieorchestern finden an einem zuvor definierten Ort statt (einer Probebühne oder auch dem späteren Aufführungsort) und dauern für gewöhnlich etwa acht Stunden. Häufig entspricht die Orchesterprobe daher einem Arbeitstag. Im Folgenden soll ein grober Ablauf der Orchesterprobe skizziert werden.

Zu Beginn einer Probe werden die dafür notwendigen materiellen und physiologischen Bedingungen hergestellt. Dafür nimmt jede/r Musiker:in seinen/ihren Platz ein, um sein/ihr Instrument auszupacken und sich ein- bzw. warmzuspielen. Der/die Konzertmeister:in gibt einen verbalen oder gestischen Hinweis, der von den Musiker:innen als Zeichen verstanden wird, dass es nun an der Zeit ist, die Instrumente zu stimmen, um später beim Spielen einzelner Noten dieselbe Tonhöhe zu produzieren. Der/die Dirigent:in kann während dieser Phase bereits im Raum sein oder auch erst später hinzustossen. Dieser Teil der Probe, in dem die Musiker:innen auch untereinander sprechen können, kann als Prä-Phase bezeichnet werden. Für die vorliegende Studie ist die eigentliche (Spiel-)Probe relevant, die dann beginnt, sobald der/die Dirigent:in seinen/ihren Platz vor dem Orchester einnimmt, die Anwesenden begrüßt und das zu spielende Stück ankündigt. In einem zweiten Schritt weist der/die Dirigent:in zum Spielen an, er:sie beginnt zu dirigieren, die Musiker:innen beginnen zu spielen. Handelt es sich um die erste

Probe eines Stücks, so wird es in der Regel in einem ersten Durchgang ohne Unterbrechung und komplett gespielt; es gibt keine korrigierenden Anweisungen des/der Dirigent:in. Ab dem zweiten Durchgang wird die Musik immer wieder durch den/die Dirigent:in unterbrochen. Er:sie identifiziert Probleme, die jeweils auch in einzelne Teile zerlegt werden können. Diese Teile werden anhand korrigierender Anweisungen des/der Dirigent:in eingeübt, teilweise auch nur mit einzelnen Instrumentengruppen oder einzelnen Musiker:innen. Im Anschluss werden mehrere Problemstellen gemeinsam geprobt oder Passagen gespielt, in denen die Probleme vorkommen. Das heißt, in dieser Phase wird die Aufführung des Stücks zerteilt, verlangsamt oder wiederholt praktiziert. Dadurch entsteht ein Muster, das wie folgt aussieht: Spiel/Musik – Unterbrechung – Anweisung – Wiedereinstieg – Spiel/Musik usw. Diesem Muster folgend wechseln sich Spielaktivitäten mit Dirigieraktivitäten ab, dazwischen gibt es Transitionen und Besprechungsphasen. Am Ende der Probe löst sich der kollektive Zusammenhang wieder auf. Der/die Dirigent:in beendet die Probe durch ein verbales Zeichen und die Musiker:innen packen ihre Sachen zusammen. In dieser Post-Phase kann es – wie vor der eigentlichen Probe – ebenfalls zu einem verbalen Austausch zwischen den Musiker:innen kommen.<sup>9</sup>

#### **2.1.2.4 Zeitlichkeit in der Orchesterprobe**

Das in der Orchesterprobe sequenziell verlaufende Muster, in dem sich Spielphasen mit Besprechungsphasen abwechseln, hat auch eine Auswirkung auf ihre Zeitlichkeit. In Orchesterproben können drei unterschiedliche Zeit-Dimensionen unterschieden werden (vgl. Abb. 1):

- a) Die Interaktions- oder Realzeit, die sich aus Sprech- und Spielzeiten zusammensetzt;
- b) die Partiturzeit, die in gleiche Teile bzw. Takte und in einzelne Stimmen der Musiker:innen unterteilt ist (die Musiker:innen spielen nicht immer alle gleichzeitig);
- c) die Spielzeit, die, wenn das Stück ohne Unterbrechung gespielt wird (erster Durchgang oder auch im Konzert), mit der Partiturzeit koinzidiert oder, wenn das Stück mit Unterbrechungen geprobt wird (zweiter oder weitere Durchgänge), in einer Zickzacklinie gegenüber der Partiturzeit verläuft.

---

<sup>9</sup> Allerdings können Musiker:innen auch während der Probe miteinander sprechen oder Blicke austauschen.

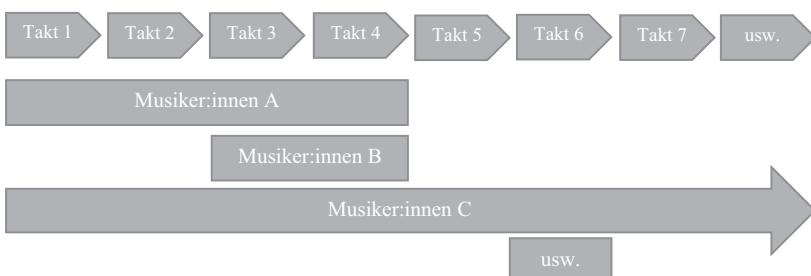
Die Zeitlichkeit in der Orchesterprobe kann folgendermaßen dargestellt werden:

### 1. Durchgang

a) Spielzeit



b) Partiturzeit

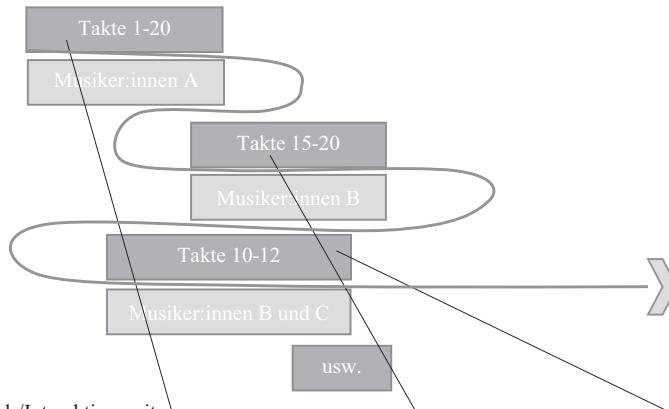


c) Real-/Interaktionszeit



### 2. Durchgang (und auch weitere Durchgänge)

a) Spielzeit und Partiturzeit



b) Real-/Interaktionszeit



**Abb. 1:** Zeitlichkeit in der Orchesterprobe.

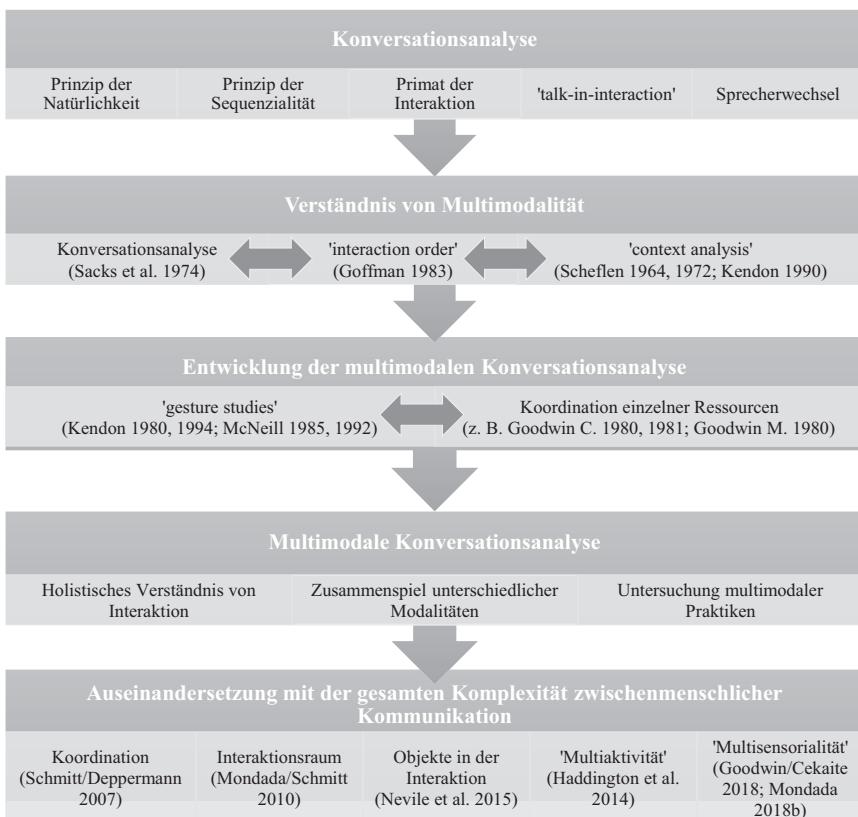
Aus dieser Darstellung lässt sich schließen, dass sich Segmente der Spiel- und auch der Partiturzeit in der Interaktions- oder Realzeit wiederfinden. Dabei ist die Interaktionszeit durch Sequenzialität geprägt, die Spielzeit und die Partiturzeit zeichnen sich durch Sukzessivität und Serialität aus, die aber ab dem zweiten Durchgang nur für jeweilige Segmente gilt, die meist, aber keinesfalls immer eine Sukzession ergeben, die der Partiturfolge entspricht. Während die Interaktionszeit immer weiter läuft, auch in den Spielphasen, in denen der/die Dirigent:in durch sein/ihr Dirigat mit den Musiker:innen kommuniziert, reguliert der/die Dirigent:in die Spiel- und die Partiturzeit durch Unterbrechungen der Musik und Wiedereinstiege in die Musik. Die Spielzeit ist immer partiturorientiert, in Takten kann vor- und zurückgesprungen werden und eine Stelle kann auch mehrmals geprobt werden, bis der/die Dirigent:in die Stelle als abgeschlossen erklärt – durch eine verbale (positive) Evaluation oder/und durch ein Weiterspielen-Lassen. Damit entstehen Interaktionssequenzen, die mit einer Aufforderung des/der Dirigent:in zum Spielen beginnen, worauf die Musiker:innen musikalisch antworten, der/die Dirigent:in eine problematische Stelle identifiziert und korrigiert, die Musiker:innen erneut musikalisch etwas anbieten und der/die Dirigent:in die jeweilige Sequenz durch eine Evaluierung beendet. In einer Orchesterprobe reihen sich mehrere solcher Sequenzen aneinander, in denen entweder die Partiturfolge berücksichtigt wird oder sich eine Art Zickzack-Bewegung ergibt, in der auch ein Vor- und Zurückspringen in der Partitur möglich ist: örtlich in der Interaktionszeit und zeitlich in der Spielzeit.

Außerdem kann an einer ganz anderen Stelle von neuem begonnen oder auch zu einem anderen musikalischen Werk gewechselt werden, falls für das Orchesterprojekt vorgesehen. Dabei erfolgt ein Wechsel zu einem anderen Stück nicht ohne Weiteres, sondern es kommt zu einer kurzen Pause zwischen Stück A und Stück B: Mit dem neuen Stück beginnt eine neue Probeneinheit.

### 2.1.3 Ethnomethodologische Konversationsanalyse

Die dieser Arbeit zugrundeliegende Perspektive auf Interaktion und die damit zusammenhängende Analysemethode gehen aus der ethnomethodologischen Konversationsanalyse hervor, deren Ansatz im Folgenden skizziert werden soll.

Dazu soll zuerst eine Grobskizze zur Orientierung angeführt werden (vgl. Abb. 2), in der ausgehend von den methodologischen Prinzipien der ‚klassischen‘ Konversationsanalyse ihre Entwicklung bis hin zur multimodalen Konversationsanalyse nachgezeichnet wird. Im Folgenden (Kap. 2.1.3 bis 2.1.5) werden die einzelnen Entwicklungsschritte und Ansätze genauer beschrieben.



**Abb. 2:** Entwicklungen und Wechselwirkungen innerhalb der Konversationsanalyse.

In diesem Kapitel geht es nicht um eine detaillierte, an Vollständigkeit orientierte Beschreibung der Forschungsrichtung der Konversationsanalyse, sondern vielmehr um die Veranschaulichung der methodisch-methodologischen Kerngedanken, die den Ausgangspunkt des analytischen Vorgehens bilden und für das spezifische Erkenntnisinteresse von Bedeutung sind. Dies soll – nach einer kurzen Darlegung der Entstehungsgeschichte und Weiterentwicklung bis heute – anhand eines Beispiels aus den in dieser Studie verwendeten Daten geschehen.

Die ethnomethodologische Konversationsanalyse wird in zahlreichen Beiträgen umfassend dargestellt, von denen hier einige angeführt sind: Bergmann (1981, 1994), Heritage (1995), Hutchby & Wooffitt (1998, <sup>2</sup>2008), ten Have (1999, <sup>2</sup>2007), Schegloff (2007), Gürlich & Mondada (2008), Sidnell (2010), Mondada (2013a, 2019a), Stivers & Sidnell (2013), Deppermann (2008, 2014a) sowie Clift (2016). Diese Arbeiten führen in die Konversationsanalyse ein, sowohl in Form von Artikeln (Bergmann 1981, 1994;

Heritage 1995; Mondada 2013a; Deppermann 2014a) als auch anhand monographischer Darstellungen (Hutchby/Wooffitt 1998, <sup>2</sup>2008; ten Have 1999; Schegloff 2007; Deppermann 2008; GÜlich/Mondada 2008; Sidnell 2010; Clift 2016). Daneben stellt das *Handbook of Conversation Analysis* von Stivers & Sidnell (2013) ein umfassendes Nachschlagewerk zu theoretischen Überlegungen und praktischen Arbeitsfeldern der Konversationsanalyse dar. Nicht zuletzt gibt es Methodendarstellungen (ten Have <sup>2</sup>2007; Deppermann 2008) sowie eine jüngst erschienene Arbeit von Mondada (2019a), die sich mit aktuellen Entwicklungen der und Herausforderungen für die Konversationsanalyse (u. a. auch hinsichtlich Multimodalität und die Erweiterung zur Multisensorialität) auseinandersetzt.

### **2.1.3.1 Entstehung und Entwicklung**

Der Begriff der Ethnomethodologie wurde in den 60er Jahren von Harold Garfinkel geprägt (vgl. Garfinkel 1967). Garfinkel interessierte sich für die Methoden, durch die Handelnde eine sinnhafte soziale Ordnung herstellen und sich gegenseitig anzeigen, was um sie vorgeht und was sie selbst tun (Bergmann 1981, 1991b, 1994). Dem ethnomethodologischen Ansatz zufolge ist soziale Wirklichkeit nicht interaktionsvorgängig vorhanden, sondern wird erst im Vollzug alltäglicher, sozialer Handlungen durch die Beteiligten lokal ausgehandelt und konstituiert (vgl. Garfinkel 1967). Bergmann (1994: 6) spricht hier auch von einer „Vollzugswirklichkeit“: Beteiligte an einer Interaktion machen ihre Handlungen noch während ihres Vollzugs verstehbar und interpretierbar bzw. „accountable“ (Garfinkel 1967: vii). Die sinnhafte Hervorbringung gesellschaftlicher Wirklichkeit ist somit als interaktive Leistung der Beteiligten zu verstehen. Diese beziehen ihre Handlungen wechselseitig aufeinander, stimmen sie ab und machen sie dadurch auch interpretierbar.

Aus diesem soziologischen Ansatz heraus hat sich die Konversationsanalyse – ebenfalls in den 60er Jahren – durch die Arbeit von Harvey Sacks und Emanuel Schegloff weiterentwickelt und als eigenständiges Untersuchungsfeld etabliert. Die Konversationsanalyse ergründet sprachliche Interaktion und fragt dabei nach Methoden, Mechanismen und Verfahren, mittels derer Beteiligte im Gespräch gemeinsam Bedeutung herstellen. Sie war ursprünglich als Instrument zur Erforschung von Alltagsgesprächen (*ordinary conversation*) gedacht, d. h., sie sollte zur Untersuchung von alltäglicher, nicht-zweckgerichteter Interaktion (v. a. Telefongespräche) zwischen Freunden und Bekannten eingesetzt werden. Allerdings wurde die Konversationsanalyse programmatisch weiterentwickelt und in vielen Untersuchungen zur institutionellen Kommunikation eingesetzt (vgl. z. B. Drew/Heritage 1992; Heritage/Clayman 2010). Dabei steht in der ethnomethodologischen Konversationsanalyse seit ihren Anfängen methodisch, theoretisch und konzeptionell die ‚conversation‘ im Mittelpunkt, d. h., es geht in erster Linie um

die verbalen Anteile menschlicher Interaktion. Später wurde der generelle Gegenstand der Konversationsanalyse von Schegloff (1991) durch ‚talk-in-interaction‘ ersetzt.

Vor allem in den letzten Jahren hat sich die klassische Konversationsanalyse allerdings dank des technologischen Fortschritts und durch den Einsatz video-gestützter Verfahren weiterentwickelt. Neben der Hörbarkeit ist nun auch die Sichtbarkeit interaktiver Bedeutungskonstitution analytisch zugänglich. Damit stehen nicht mehr nur die verbalen Anteile der Interaktion im Fokus, sondern es können auch Gestik, Mimik, Blick, Körperpositur und -bewegungen sowie die Benutzung und Manipulation von Objekten im Kontext der Interaktion mit in die Analyse einbezogen werden. Die heutige ethnomethodologische Konversationsanalyse ist demnach nicht mehr einzig auf Sprache fokussiert, sondern auf „human action“ (Mondada 2019a: 48), für die Sprache eine Ressource unter anderen ist: „[...] this approach considers situated action as the primordial locus for understanding the organization of language, sociability, and culture in their material and social ecologies.“ Sprache wird nicht mehr als Primat gehandelt, sondern es wird davon ausgegangen, dass verschiedene verbale und nonverbale Ressourcen gemeinsam zur Organisation sozialer Aktivitäten und deren situativer, emergenter und normativer Hervorbringung beitragen. „[T]o account for the organization of action“ (Mondada 2019a: 48), darin liegt das oberste Ziel der Konversationsanalyse.

### **2.1.3.2 Methodisches Vorgehen**

Ziel konversationsanalytischer Untersuchungen ist es, zu rekonstruieren, wie in methodischer Weise soziale Wirklichkeit durch soziales Handeln erzeugt wird (vgl. Garfinkel 1967). Dabei widersetzen sich sowohl Ethnomethodologie als auch Konversationsanalyse einer methodischen Anleitung zur schrittweisen Erschließung empirischer Daten, es gibt keinen festgelegten Methodenkanon (Bergmann 1988: 5). Vielmehr geht es darum, nicht nur ein interaktives Phänomen zu untersuchen, sondern gleichzeitig analytische Methoden und Verfahren zu entwickeln, die zur Erforschung dieses Phänomens angewendet werden können. Garfinkel (1967, 2002) führt dafür die Maxime ‚unique adequacy requirement‘ ein, d.h., die Methoden müssen dem jeweiligen Untersuchungsgegenstand angemessen sein (vgl. auch Bergmann 2004: 57). Trotz der Absage an jegliche Kanonisierung methodischer Regeln zeichnet sich die Konversationsanalyse doch durch eine besondere Analysehaltung aus, die sog. ‚analytische Mentalität‘ (*analytic mentality*) (Bergmann 1981: 18–24; Schenkein 1978). Diese besagt, dass keine vorformulierten Fragestellungen an das Material herangetragen werden, sondern dass sich die Fragestellungen aus dem Untersuchungsmaterial selbst ergeben müssen.

Diese analytische Mentalität lässt sich durch eine Reihe von Prinzipien und Maximen charakterisieren:<sup>10</sup>

1) *Prinzip der Natürlichkeit*: Gegenstand der Analyse sind ‚natürliche Daten‘, d. h. nicht experimentell erzeugte, authentische Gespräche, die auf Tonband oder Video aufgezeichnet und in Transkripten festgehalten werden. Dabei gilt bei der Aufnahme der Daten das Prinzip der ‚konstitutionslogischen Vollständigkeit des Datums‘:

Die aufgenommenen Daten sollen soweit wie möglich alle multimodalen Ausdrucksressourcen (Stimme, Gestik, Mimik, Körperbewegung) und alle räumlich-visuellen Sachverhalte (räumliche Strukturen, Gegenstände, die in der Interaktion benutzt werden) repräsentieren, mit denen [...] die InteraktionsteilnehmerInnen ihr Handeln organisieren. (Deppermann 2014a: 20)

Solche Aufzeichnungen machen Interaktionen beliebig oft rekonstruierbar und ermöglichen daher eine detaillierte Rekonstruktion der von den Beteiligten eingesetzten Methoden.

2) *Prinzip der Sequenzialität*: Auf der Grundlage von Transkripten wird der Gesprächsverlauf abgebildet, dessen Sequenzialität und Temporalität in der Analyse berücksichtigt werden müssen. Äußerungen werden nicht isoliert, sondern immer in ihrem sequenziellen Kontext analysiert. Man geht von einem zeitlichen Ablauf aus, von einem Nach- bzw. Aufeinanderfolgen. Damit einher geht auch die Serialität der Interaktion: Eine sprachliche Handlung (als Antwort, Erwiderung) bezieht sich auf eine andere, vorhergehende. Das heißt, es erfolgen eine Handlung und die Reaktion darauf.

3) *Prinzip der Ordnung*: In einem ersten Analyseschritt geht es darum, Gleichförmigkeiten und Regelmäßigkeiten im Gespräch aufzufinden. Sprachliche und nicht-sprachliche Interaktion hinterlässt bei oberflächlicher Betrachtung zwar einen chaotischen Eindruck, allerdings wird bei genauerer Analyse deutlich, dass sie sehr wohl geordnet und einer Struktur folgend abläuft.

4) *Order at all points'-Maxime* (Sacks 1984): Die beobachtbaren Geordnetheiten werden nicht als Zufallsprodukte, sondern als systematisch produzierte Ergebnisse der methodischen Lösung struktureller ‚Probleme‘ der Interaktionsorganisation aufgefasst. Diese werden in einem zweiten Schritt der Analyse identifiziert.

---

<sup>10</sup> Für eine ausführliche Darstellung der Arbeitsprinzipien vgl. die Ausführungen von Bergmann (1981: 18–24).

5) *Primat der Interaktion*: Die Beiträge der Beteiligten an einer Interaktion werden immer zu denen der Gesprächspartner:innen in Bezug gesetzt und als Ergebnis gemeinsamer Aktivitäten interpretiert. Es geht um die Frage nach der Verbindung von Handlungen bzw. Äußerungen von verschiedenen Personen, also Anzeichen, die sich aufeinander beziehen. Die Geordnetheit in einem Gespräch wird demnach immer interaktionsorientiert betrachtet.

6) *Perspektive der Interaktionsteilnehmer:innen*: Es gilt am Datenmaterial nachzuweisen, dass die beschriebenen formalen Mechanismen tatsächliche, reale Orientierungsgrößen für die Beteiligten darstellen. Es muss gezeigt werden, „an welchen interaktiven Aufgaben und Problemen und an welchem Verständnis von Kontext und Partnerhandeln sich die InteraktionsteilnehmerInnen in ihrem Handeln erkennbar orientieren“ (Deppermann 2014a: 22).

7) *Beschreibung der ‚Maschinerie‘*: In einem dritten Analyseschritt wird die ‚Maschinerie‘ der Gesprächsentwicklung beschrieben. Dafür werden die kleinsten sinnstiftenden Äußerungseinheiten (oder Turns) rekonstruiert sowie die Verteilung des Rederechts aufgeschlüsselt. Sacks, Schegloff & Jefferson (1974) definieren für die Rederechtsverteilung drei aufeinanderfolgende Optionen, die für alle übergaberelevanten Stellen (*transition relevance place, TRP*) gegeben sind:

- a) An einer übergaberelevanten Stelle kann der/die aktuelle Sprecher:in den/ die nächste/n Sprecher:in auswählen, z. B. indem er:sie diese/n anspricht oder ihm/ihr eine Frage stellt (Fremdwahl).
- b) Daneben kann sich ein/e nächste/r Sprecher:in auch selbst auswählen (Selbstwahl).
- c) Wenn sich niemand selbst auswählt, kann der/die aktuelle Sprecher:in fortfahren.

Die Konversationsanalyse setzt sich demnach mit der detailgenauen sequenzanalytischen Untersuchung von alltäglichen und institutionellen Interaktionssituationen auseinander (Bergmann 1985; Deppermann 2014a: 21). Dafür müssen Transkripte angefertigt werden, die die Daten schriftlich fixieren und das Material jederzeit verfügbar und rezipierbar machen. In solchen Transkripten wird nach Konventionen gearbeitet, die die Besonderheiten gesprochener Sprache (etwa Pausen, Abbrüche, Versprecher, Verzögerungen usw.) und nonverbaler Anteile der Interaktion (etwa Gestik, Mimik, Körperbewegungen, Blickverhalten usw.) sowie den Interaktionsablauf inklusive Sprecherwechsel und Überlappungen berücksichtigen (Deppermann 2014a: 21; vgl. auch Deppermann/Schütte 2008). Dafür werden im deutschen Sprachraum die Konventionen nach GAT 2 (Gesprächsanalytisches Transkriptionssystem 2,

vgl. Selting et al. 2009) angewandt, an die sich auch die Transkription des folgenden Ausschnitts aus unserem Korpus anlehnt:<sup>11</sup>

Haydn Orchester Bozen, Aufnahme\_3\_21072016, Kamera Dirigent, 0000  
 Beteiligte: Dirigent Arvo Volmer (D), Musiker:innen (MM), Konzertmeister (KM)  
 Werk: Prokofjew, Symphonie Nr. 1 D-Dur

19:36-20:26

```

01 MM  ♪((spielen))                ♪]
02 D   [((winkt ab)) (2.5) <<f> questa uno forte:>]
03     <<dim> con riserva pe:r prima di DUE forte>
04     ((räuspert sich, 1.4sec)) (0.4) ((blättert in Partitur, 0.8sec))
05     e::h ↑e::::h
06     (.) <<f> quindici> (3.0)
07 KM  maestro facciamo con questo (.) facciamo che: e:h qui noi
08     (.) diciotto (0.4) IL LEVARE
09     (.) ANCORA in giù allora xxx
10     ((zeigt neben ihm sitzenden Musiker etwas in den Noten und summt, 3.5sec))
11 D   ((geht zu KM und schaut in seine Noten, 4.5sec))
12     xxx
13 KM  <<p> no no no xxx
14 D   okay
15 KM  (i think it's) xxx>
16 D   =yeah (but one more play fifteen)
17 (1.0) (i don't) need i don't need
18 (.) you- you tell me please when you when you:: (0.6)
19 KM  xxx
20 D   =when you need something
21     ((blättert in Partitur, 1.6sec))
22     <<f> quindici> <<dim> per favore>
23     (4.5) ((streckt Arme nach außen)) °H:
24     ((beginnt zu dirigieren))
25 MM  ♪((beginnen zu spielen))♪

```

Dieses Transkript ist in Segmente bzw. Intonationsphrasen<sup>12</sup> gegliedert, die linksbündig durchnummerniert sind.<sup>13</sup> Der Transkripttext enthält die Abfolge der Spre-

<sup>11</sup> Vgl. im Anhang der vorliegenden Arbeit eine genaue Auflistung der verwendeten Transkriptionskonventionen.

<sup>12</sup> Für die Gliederung eines Transkripts in Segmente oder Intonationsphrasen ist die Prosodie ausschlaggebend, d. h. der Tonhöhenverlauf gesprochener Sprache, Lautstärke, Länge, Pausen sowie Sprechgeschwindigkeit und Rhythmus (vgl. Selting et al. 2009: 370). Ausführlicher zu Prosodie vgl. Heinz (2006: 11–14, 36–54), zu Prosodie und Konversationsanalyse vgl. Rabanus (2001: 5–62).

<sup>13</sup> Diese Beschreibung des Transkripts ist angelehnt an Deppermann (2014a: 21).

cher („D“ steht für Dirigent, „KM“ steht für Konzertmeister) bzw. auch der Spieler:innen (= die Musiker:innen, „MM“ im Transkript), einschließlich der Markierung, wann sie einander überlappen (in Z01<sup>14</sup> und Z02 spielen die Musiker:innen [noch], wenn der Dirigent bereits verbal eine Anweisung gibt). Außerdem wird der Wortlaut so präzise wie möglich wiedergegeben. Dazu zählen nicht nur Versprecher, Abbrüche und Selbstkorrekturen (Z17 und Z18), sondern auch nicht-lexikalische Laute („eh“ in Z05 und Z07). Auf prosodischer Ebene werden Pausen (gemessen, etwa in Z06, wobei der Punkt für 0.1sec steht, die Angabe „3.0“ für 3sec), Dehnungen<sup>15</sup> (durch einen oder mehrere Doppelpunkte, z. B. „forte::“ in Z02), Akzente (durch Großschreibung, z. B. „DUE“ in Z03) und Tonhöhenbewegungen (der Pfeil in Z05 gibt eine steigende Intonationsbewegung an) mit in das Transkript aufgenommen. Ebenso werden Veränderungen in der Lautstärke gegenüber zuvor gewählten oder etablierten Vergleichsnormen festgehalten (z. B. ein *forte* [*<>f>* text>] in Z02 oder ein *diminuendo* [leiser werden] in Z03). Zu den prosodischen Merkmalen zählt nicht zuletzt das Einatmen des Dirigenten in Z23 („H:“). Nonverbale Ereignisse (in doppelten Klammern, etwa das Abwinken in Z02, das Räuspern und das Blättern in der Partitur in Z04) und nicht verständliche („xxx“ in Z09) oder schlecht verständliche Passagen (in einfachen Klammern, etwa „i think it’s“ in Z15) werden ebenfalls notiert. Kommt es zu einem schnellen Anschluss von Intonationsphrasen, so wird dieser durch ein Gleichheitszeichen wiedergegeben (etwa in Z16 oder in Z20, wo der Dirigent seine jeweiligen Turns unmittelbar an jene des Konzertmeisters anschließt, ohne die übliche Mikropause). Im Transkript ist es zudem möglich, Aspekte der visuellen Interaktion zu notieren. Im Abschnitt zur Multimodalität in der Interaktion (Kap. 2.1.5) soll ein Teil desselben Transkripts um eine visuelle Transkription erweitert und hinsichtlich nonverbaler Phänomene erläutert werden.

Um ein solches Transkript zu analysieren, arbeitet die Konversationsanalyse mit einer detaillierten Sequenzanalyse:<sup>16</sup>

---

<sup>14</sup> Die Abkürzung „Z“ steht für „Zeile“.

<sup>15</sup> Dehnungen werden im Unterschied zu Pausen auditiv-impressionistisch und nicht durch Messung erfasst.

<sup>16</sup> Daneben kann auch mit Kollektionen gearbeitet werden. Die Kollektionsanalyse zeichnet sich durch komparative Verfahren aus, die beständig Einzelfall und Kollektionen gegenüberstellen (vgl. Deppermann 2008, Kap. 6). Durch diesen Prozess kann die Kollektion mittels Tilgungen, Erweiterungen und Untergliederungen bereinigt und systematisiert werden, bis sich eine Struktur herausbildet, die durch die Detailanalyse weiterer Fälle nicht mehr modifiziert wird (Deppermann 2010: 649). Diese Vorgehensweise bietet sich für die Untersuchung der Orchesterproben-Interaktion allerdings nicht an.

Die Sequenzanalyse hat zum Ziel, zu rekonstruieren, wie die GesprächsteilnehmerInnen gemeinsam ihre Interaktion als schrittweise aufeinander bezogenen Prozess der Sinnbildung und Herstellung von Interaktionsstrukturen organisieren. (Deppermann 2014a: 22)

Das heißt, Gesprächsbeiträge/Turns definieren eine Situation im Hier und Jetzt (vgl. Goodwin/Heritage 1990). Folgende Beiträge orientieren sich immer an vorhergehenden Handlungen und sind damit in einen sequenziellen Kontext eingebettet. Diese Sequenzialität von Handlungen in der Interaktion zeigt sich beispielsweise an Paarsequenzen (*adjacency pairs*). Damit sind Äußerungspaares gemeint, die aus zwei Teilen bestehen (*first pair part/FPP* und *second pair part/SPP*), direkt aufeinanderfolgen und von unterschiedlichen Sprecher:innen produziert werden, z. B. Gruß-Gegengruß, Frage-Antwort, Angebot-Annahme/Verweigerung. Dabei etabliert der erste Teil des Paares eine so starke konditionelle Relevanz, dass nur ein spezifischer Äußerungstyp eine „erwartbare, relevante und adäquate Folgehandlung“ darstellt (Gülich/Mondada 2008: 51; vgl. auch Schegloff/Sacks 1973). Allerdings steht es den Beteiligten frei, ob sie in präferierter (etwa wenn ein Angebot angenommen wird) oder in dispräferierter Form (etwa wenn ein Angebot verweigert wird) auf die erste Äußerungseinheit einer Paarsequenz reagieren. Auch in dem Beispiel oben taucht eine solche Sequenz auf: In Z15 macht der Konzertmeister einen Vorschlag („i think it's xxx“), auf den der Dirigent in Z16 in präferierter Weise durch eine Zustimmung reagiert („yeah“). Die normativ erwartete Zustimmung wird jedoch durch die Weiterführung des Turns durch die adversative Konjunktion „but“ (Z16) abgeschwächt. Durch das „but“ projiziert der Dirigent eine Gegenposition zum Vorschlag des Konzertmeisters. Damit zeigt er nicht nur an, wie er die vorausgehende Handlung interpretiert hat, sondern projiziert gleichzeitig eine künftige konversationelle Handlung. Der Turn ist demnach zum einen retrospektiv ausgerichtet („Why that now?“, Schegloff/Sacks 1973) und drückt aus, auf welche Kontextaspekte er sich bezieht, wie diese interpretiert werden und wie er durch den vorausgegangenen Kontext erwartbar (*accountable*) gemacht wurde. Zum anderen ist der Turn auch vorausschauend („What comes next?“, Schegloff/Sacks 1973), indem er demonstriert, welche Funktion er für den weiteren Gesprächsverlauf hat. Schegloff & Sacks (1973) sprechen hier auch von „sequential implicativeness“:

An utterance projects for the sequentially following turn(s) the relevance of a determinate range of occurrences (by their utterance types, activities, speaker selection, etc.). It thus has sequentially organized implications. (Schegloff/Sacks 1973: 296)

Dieser doppelte zeitliche Horizont impliziert ein Monitoring der Interaktionsbeteiligten, d. h., sie müssen ständig die Beiträge der anderen analysieren, um angemessen darauf reagieren zu können. Dieses Monitoring spielt auch beim Spre-

cherwechsel eine Rolle:<sup>17</sup> Interaktionsbeteiligte rezipieren Redebeiträge anderer *online*, also im laufenden Prozess, und platzieren ihre eigenen Beiträge in Momenten, in denen Turns potenziell abgeschlossen sind (Gülich/Mondada 2008: 39). So platziert der Konzertmeister in dem Ausschnitt oben seinen Beitrag in Z07 in einer Sprechpause des Dirigenten („maestro facciamo con questo ...“). Er identifiziert die Stelle als TRP und selektiert sich selbst als nächsten Sprecher. Ähnlich verhält es sich in Z19, allerdings vervollständigt hier der Konzertmeister die begonnene Äußerung des Dirigenten aus Z18 („you tell me please when you when you:“). Hier kann außerdem von einer (impliziten) selbstinitiierten Fremdkorrektur gesprochen werden: Ein problematisches Element (hier: eine Verzögerung, eine Wortsuche) wird vom Dirigenten in seinem Turn identifiziert und im darauffolgenden Turn vom Konzertmeister repariert. In Z20 kommt es zudem zu einer Selbstkorrektur des Dirigenten („when you need something“). Auch solche Verfahren, mit denen im Gesprächsverlauf auftretende Störungen bearbeitet bzw. „repariert“ werden, gehören zu den Grundformen der sequenziellen Organisation. Daneben beschäftigt sich die Konversationsanalyse mit der Eröffnung und Beendigung von Gesprächen, mit der thematischen Organisation und mit Erzählstrukturen. Diese Phänomene werden hier allerdings nicht weiter ausgeführt, da sie für das Erkenntnisinteresse der vorliegenden Studie nicht – oder nur wenig – von Belang sind.

Insgesamt kann die Vorgehensweise der Konversationsanalyse – Deppermann (2014a) folgend – in dieser Art zusammengefasst werden:

Die KA [Konversationsanalyse] wendet also zentrale Konstitutionseigenschaften der Interaktion, nämlich Sequenzialität (das zeitliche Nacheinander von Aktivitäten), Interaktivität (das Sich-aufeinander-Beziehen), den Handlungsbezug von Interaktionen, die Methodizität des Handelns (die systematische Verwendung von Interaktionspraktiken) und ihre multimodale Konstitution methodologisch an. (Deppermann 2014a: 24)

## 2.1.4 Sprecherwechsel-Systeme

Ein zentrales Thema in der Konversationsanalyse ist die Sprecherwechselsystematik („machinery“ nach Sacks/Schegloff/Jefferson 1974) bzw. das Turn-taking in Gesprächen. Diese Systematik geht auf den wohl bekanntesten Artikel der Konversationsanalyse von Sacks, Schegloff und Jefferson zurück (1974, „A Simplest Systematics for the Organization of Turn-taking for Conversation“) und stellt Regeln

---

<sup>17</sup> Für eine ausführliche Beschreibung von Sprecherwechsel-Systemen vgl. auch den eigenen Abschnitt dazu weiter unten (Kap. 2.1.4).

und Prämissen auf, die die Organisation von Redebeiträgen in Alltagsgesprächen regulieren. Im Mittelpunkt stehen dabei folgende Fragen:

- a) Wenn zwei oder mehr Personen miteinander sprechen, wie managen sie dann den Ablauf von Gesprächsbeiträgen?
- b) Wie werden Redebeiträge oder Turns ausgetauscht und abgewechselt, sodass weder längere Pausen noch Überlappungen oder gleichzeitiges Sprechen entstehen?

Sacks, Schegloff & Jefferson (1974) beschreiben anhand aufgezeichneter und transkribierter Telefongespräche, wie das wechselseitige Abstimmen und Koordinieren von Gesprächsbeiträgen funktioniert. Wichtig dabei sind zwei Komponenten (vgl. auch Göllich/Mondada 2008): das Konstruieren von Turns (*turn constructional units*) sowie das Verteilen des Rederechts (*turn-allocational component*).

Gesprächsbeteiligte orientieren sich an übergangsrelevanten Stellen, sog. *transition relevance places* (TRPs, Sacks/Schegloff/Jefferson 1974), die am Ende von Beitragskonstruktionseinheiten, sog. *turn constructional units* (TCUs, Sacks/Schegloff/Jefferson 1974), auftauchen können. Mit dem Abschluss solcher *units* wird signalisiert, ob ein Sprecherwechsel erfolgen soll: „transition-relevance places [...] occur at ‚possible completion points‘.“ (Sacks/Schegloff/Jefferson 1974: 721; vgl. auch Auer 2010) Sacks, Schegloff & Jefferson (1974) gehen davon aus, dass TCUs voraussagbar (*projectable*) sind, d. h., Gesprächsbeteiligte beobachten und rezipieren Redebeiträge online, nehmen den weiteren Verlauf oder die mögliche Vollständigkeit von Turns vorweg und antizipieren auf diese Weise mögliche übergaberelevante Stellen (vgl. auch Göllich/Mondada 2008; Couper-Kuhlen/Selting 2018a). Das System der TCUs erklärt nicht nur, warum die Wahl eines/einer Sprecher:in in Gesprächen überwiegend reibungslos vonstatten geht, sondern auch, warum und wie es zu Überlappungen und Pausen oder zur Produktion von Turnergänzungen sowie in bestimmten Interaktionskontexten auch zu Abbrüchen kommen kann (vgl. Deppermann 2014a).

Außerdem stellen Sacks, Schegloff & Jefferson (1974) fest, dass Gespräche durch das *lokale Management* des Sprecherwechsels gekennzeichnet sind (vgl. auch Levinson 1983): Am Ende einer jeden TCU wird ausgehandelt, wer als nächstes spricht (vgl. Deppermann 2014a). Dabei gelten folgende Regeln (Sacks/Schegloff/Jefferson 1974: 704):

- 1) Regel 1 – gilt für den ersten TRP jedes Beitrags:
  - a) Sprecher:in A wählt in seinem/ihrem Beitrag Teilnehmer:in B aus (Fremdwahl). Nach dem nächsten TRP darf B sprechen.
  - b) Wählt Sprecher:in A in seinem/ihrem Beitrag niemanden aus, so darf sich jede:r selbst auswählen und das Wort ergreifen (Selbstwahl).

- c) Wählt Sprecher:in A niemanden in seinem/ihrem Beitrag aus und es nimmt sich kein/e andere/r Sprecher:in durch Selbstwahl das Rederecht, so kann sich Sprecher:in A erneut auswählen und weitersprechen.
- 2) Regel 2 – gilt für alle weiteren TRPs:  
Ist Regel 1c) von Sprecher:in A angewandt worden, so gelten beim nächsten TRP die Regeln 1a) bis 1c) und fortlaufend bei jedem weiteren TRP, bis der Sprecherwechsel vollzogen wurde.

Dieses von Sacks, Schegloff & Jefferson (1974) aufgestellte Regelwerk sagt nicht voraus, wie in einer bestimmten alltäglichen Gesprächssituation einzelne Redebeiträge verteilt sein werden. Vielmehr handelt es sich um allgemeingültige Regeln, deren Einhaltung dazu führt, dass ein Gespräch strukturiert und geordnet verläuft (vgl. auch Glich/Mondada 2008). Diese Regeln und Prämissen können in unterschiedlichen Kontexten und unabhängig von Gesprächsthemen zum Einsatz kommen: „The system is described as [...] context-free, [...] locally managed, and party-administered.“ (Couper-Kuhlen/Selting 2018a: 31)

#### **2.1.4.1 Sprecherwechsel-Systeme in institutionellen Settings**

Die Systematik des Sprecherwechsels, so wie sie von Sacks, Schegloff & Jefferson (1974) für alltägliche Interaktionen beschrieben wird, kann nicht unverändert auf institutionelle Interaktionssituationen übertragen werden. Bereits Sacks, Schegloff & Jefferson (1974) halten fest, dass ihre Regeln – vor allem im Hinblick auf die Rederechtsverteilung – nicht für jegliche Arten von Interaktionen gültig sind. Institutionelle Interaktionen zeichnen sich durch spezialisierte Sprecherwechselsysteme aus (Heritage/Clayman 2010; vgl. auch Deppermann 2014a), die an den jeweiligen sozialen Zweck und die für institutionelle Kommunikation konstitutiven Diskurs-Rollen angepasst sind (vgl. Zimmermann 1998). In Interaktionen im Gericht (Atkinson/Drew 1979), im Klassenzimmer (McHoul 1978; Mehan 1979), im Rundfunkinterview (Clayman/Heritage 2002), im Arzt-Patienten-Gespräch (Heritage/Maynard 2006; Heritage/Robinson 2006), in politischen Debatten (Mondada 2013b) oder in der Pressekonferenz (Deppermann 2014a) haben wir es mit solchen spezifischen Sprecherwechselsystemen zu tun. Als Beispiel sei hier genauer auf die Situation des Arzt-Patienten-Gesprächs eingegangen (vgl. für die Interaktion zwischen Ärzt:in und Patient:in auch Brünner 2005; Neises/Ditz/Spranz-Fogasy 2005; Nowak/Spranz-Fogasy 2009; Spranz-Fogasy 2010). Nowak & Spranz-Fogasy (2009) erarbeiten in ihrer Studie Unterschiede zwischen der Arzt-Patienten-Kommunikation und der Kommunikation im Alltag:

- a) Die institutionelle Einbindung;
- b) die asymmetrische Beteiligungskonstellationen (ärztliche Gespräche sind Experten-Laien-Gespräche, vgl. auch Brünner 2005): Der/die Ärzt:in setzt das

- Gespräch und die einzelnen Schritte darin in Gang und beansprucht Initiativrechte mit typischen Aktivitäten wie Frage, Erklärung oder Instruktion;
- c) der Zweckbezug: ärztliche Gespräche sind aufgabenorientiert und dienen gesprächsübergreifenden Zwecken der Diagnosefindung und Therapieentwicklung.

Diese Beobachtungen lassen darauf schließen, dass es in institutioneller Interaktion zu Restriktionen kommen kann, die das Rederecht betreffen (etwa Schüler:innen dürfen im Klassenzimmer nur etwas sagen, wenn sie sich per Handzeichen melden und von der Lehrperson ausgewählt werden), wie auch die Konstruktion und Produktion von Redebeiträgen (etwa der/die Interviewende stellt Fragen, der/die Interviewte gibt Antworten), oder die Verteilung von Wissen und Macht:

In many forms of institutional discourse, by contrast, there is a direct relationship between status and role, on the one hand, and discursive rights and obligations, on the other. As we have detailed, institutional interactions may be characterized by role-structured, institutionalized, and omnirelevant asymmetries between participants in terms of such matters as differential distribution of knowledge, rights to knowledge, access to conversational resources, and participation to interaction. (Drew/Heritage 1992: 49)

Das gesamte Sprecherwechselsystem ist damit in hohem Maß prädeterriniert (*pre-allocated* nach Sacks/Schegloff/Jefferson 1974: 729), restriktiv und spezialisiert (vgl. Drew/Heritage 1992). Das bedeutet, dass die Form sowie die Ordnung von Gesprächsbeiträgen der Beteiligten bereits vorab relativ fix festgelegt sind und dass das Rederecht sowie die Verteilung von Turns (zumeist) auf eine Person fällt (etwa der/die Lehrer:in im Klassenzimmer, der/die Ärzt:in im Arzt-Patienten-Gespräch usw.).

- Vor allem in institutionellen Interaktionssituationen, an denen eine größere Gruppe von Interaktant:innen beteiligt ist – etwa im Gericht, im Klassenzimmer oder auch in der Orchesterprobe –, können spezielle Turn-taking-Formate bestimmte Personen/Rollen definieren, die das Recht haben zu sprechen oder jemanden auffordern können zu sprechen (vgl. Mondada 2013b). Damit wird die Aktivität des Sprechens auf eine Gruppe von Beteiligten oder auch auf nur eine/n Beteiligte:n beschränkt; diejenigen, die zuhören, werden dabei mehr als Kollektiv (*party*) und weniger als einzelne mögliche Sprecher:innen (*possible persons*) aufgefasst (vgl. Schegloff 1995; Mondada 2013b). Nach Heritage & Clayman (2010) gibt es zwei unterschiedliche Arten, um größere *Multi-party*-Interaktionen zu managen:
- a) Das *turn-type pre-allocation system* (z. B. in Interviews oder vor Gericht), wo es bestimmte Funktionsrollen gibt (Interviewende:r/Interviewte:r), die mit einem speziellen Handlungstyp verknüpft sind (vom/von der Interviewten wird erwartet)

- tet, dass er:sie auf Fragen antwortet und keine anderen sprachlichen Handlungen ausführt);
- b) das *mediated turn-taking system* (z. B. in Meetings), wo ein/e Vorsitzende:r die Verteilung der Turns, die jeweiligen Sprecher:innen sowie die Themen überwacht.

Außerdem konstatiert Mondada (2013b), dass in *Multi-party*-Interaktionssituationen vor allem visuellen Ressourcen hinsichtlich der Organisation des Turn-taking eine bedeutende Rolle zukommt (vgl. auch Sahlström 2002; Schmitt 2005; Ford 2008; Markaki/Mondada 2012; Mondada 2012c). Und zwar, wenn es darum geht, künftige Sprecher:innen zu identifizieren und (vor)auszuwählen. Beispielsweise zeigt das Heben einer Hand eines/einer Beteiligten in Meetings an, dass derjenige/diejenige sich selbst als (mögliche/n) nächste/n Sprecher:in auswählt und vom Vorsitz auch als solche/r wahrgenommen werden möchte: „the visible and public accountability of turn-taking is a continuous practical accomplishment.“ (Mondada 2013b: 41) Wir werden sehen, dass solche visuell-gestischen Ressourcen auch in der Orchesterprobe von Bedeutung sind, vor allem dann, wenn der/die Dirigent:in eine Unterbrechung der Musik initiiert und das Rederecht für sich beansprucht.

Institutionelle Settings zeichnen sich außerdem hinsichtlich des Sprecherwechselsystems dadurch aus, dass eine Person einer Gruppe gegenübersteht, die sich wiederum in Untergruppen oder einzelne Individuen aufspalten kann. So wie im Klassenzimmer der/die Lehrer:in das Rederecht hat und die Schüler:innen als Gruppe funktionieren, der etwas beigebracht wird und die instruiert wird, hat auch in der Orchesterprobe der/die Dirigent:in das Rederecht auf seiner/ihrer Seite und instruiert die Musiker:innen als Gruppe. Anders als im Klassenzimmer sind Turns in der Orchesterprobe allerdings nicht nur durch Sprechen, sondern auch durch Musik konstituiert. Die Turns der Musiker:innen bestehen aus dem Spielen auf ihren Instrumenten, ihr Rederecht ist stark eingegrenzt. Als Sprecher:in kommt in seltenen Fällen neben dem/der Dirigent:in der/die Konzertmeister:in ins Spiel, dem/der als Stellvertreter:in des Orchesters besondere Rechte hinsichtlich des Sich-Zu-Wort-Meldens in der Probe zukommen. Ein weiterer Unterschied zur Interaktion in schulischen Settings ist das Vorhandensein einer Partitur in der Orchesterprobe. Die Partitur gibt nicht nur vor, was gespielt werden soll, sondern sie determiniert in einer bestimmten Art und Weise auch das, was in den Unterbrechungsteilen besprochen wird. Das heißt, Handlungen und Äußerungen können nicht frei gewählt werden, sondern orientieren sich an einer fixierten Vorlage. Dieses spezifische „Sprecher“-Wechselsystem in der Orchesterprobe soll im Folgenden skizziert werden.

### 2.1.4.2 Das „Sprecher“-Wechselsystem in der Orchesterprobe

Für das „Sprecher“-Wechselsystem in der Orchesterprobe gilt:

- a) Es gibt eigentlich nur eine/n Sprecher:in, den/die Dirigent:in.
- b) Diese/r hat das Recht, die Musik an jedem erdenklichen Punkt zu unterbrechen.<sup>18</sup>
- c) Ist die Musik unterbrochen, spricht der/die Dirigent:in, die Musiker:innen hören zu (diachrone Instruktion).
- d) Kommt es innerhalb von Besprechungsparts zu einem Austausch zwischen Dirigent:in und einem/einer Musiker:in (in den allermeisten Fällen mit dem/der Konzertmeister:in), gelten dieselben Regeln für das Turn-taking wie von Sacks, Schegloff & Jefferson (1974) aufgestellt.
- e) Besprechungssteile dauern so lange, bis der/die Dirigent:in einen neuen musikalischen Teil einleitet und den *Floor* an die Musiker:innen übergibt.
- f) Die Musiker:innen haben in ihren Turns während Performance-Parts eine Spielpflicht und ein Spielrecht.
- g) Diese musikalischen Turns werden durch die Partitur (und auch durch die Korrekturen des/der Dirigent:in) bestimmt: Diese gibt vor, was gespielt wird.
- h) Während die Musiker:innen spielen, dirigiert der/die Dirigent:in (synchrone Instruktion); dabei nimmt er:sie sein/ihr Rederecht nicht wahr/zurück, allerdings hat er:sie durch das Dirigat ein gestisches Recht, das – so wie die Anweisungen und Korrekturen in Unterbrechungsphasen – instruierend ist. Trotzdem kann es durchaus der Fall sein, dass der/die Dirigent:in Instruktionen simultan zur Musik anbringt, indem er:sie diese in den Raum ruft.
- i) In solchen musikalischen Parts überlappen sich die Spielaktivität der Musiker:innen und die Dirigieraktivität des/der Dirigent:in, es kommt zu einem gemeinsamen Turn.

Wir haben es demnach mit einem Wechsel zwischen dem/der Dirigent:in als Sprecher:in (oder auch als Gestikulierer:in, Sänger:in) sowie den Musiker:innen als Zuhörer:innen in Besprechungssteilen und den Musiker:innen als Musizierende sowie dem/der Dirigent:in als Dirigierende:r in Performance-Teilen zu tun. Das heißt, das Rederecht des/der Dirigent:in alterniert mit dem Spielrecht/der Spielpflicht der Musiker:innen. Dabei gelten bestimmte Regeln für das Spielen der Musiker:innen: Diese werden durch die Partitur und die Korrekturen/Instruktionen des/der Dirigent:in festgelegt. Zugleich ist das Sprechen des/der Dirigent:in normgeleitet: Dieses findet in den Besprechungsphasen statt, d. h., wenn die Musiker:innen nicht spielen, und orientiert sich dabei an der Partitur und am soeben Gespielten. Darüber hinaus kann der/die Dirigent:in allerdings noch zusätzlich

---

<sup>18</sup> Vgl. die diskursiven *rights and obligations* im Zitat von Drew & Heritage (1992) weiter oben.

andere semiotische Ressourcen – etwa Gestik oder Singen – verwenden, mit denen er:sie instruiert und korrigiert, oder auch simultan zur Musik sprechen.

Die diskursiven Rechte und Pflichten des/der Dirigent:in sind also im Vergleich zu denen der Musiker:innen weniger eingeschränkt und determiniert. Wie diese *rights and obligations* in der Interaktion behandelt und relevant gesetzt werden, soll im empirischen Teil zu Instruktionen untersucht werden.

## 2.1.5 Multimodale Konversationsanalyse

Face-to-face interaction is, by definition, multimodal interaction in which participants encounter a steady stream of meaningful facial expressions, gestures, body postures, head movements, words, grammatical constructions, and prosodic contours.

Tanya Stivers & Jack Sidnell<sup>19</sup>

### 2.1.5.1 Multimodalität in der Interaktion

Der von Stivers & Sidnell (2005) in ihrer Einleitung zu einer Reihe von Studien zu multimodaler Interaktion postulierte grundsätzlich multimodale Charakter menschlicher Kommunikation blieb in der ‚klassischen‘ Konversationsanalyse lange Zeit außen vor. Die ethnomethodologische Konversationsanalyse konzentrierte sich auf das Verbale, auf ‚talk-in-interaction‘<sup>20</sup> und reduzierte damit die – zumeist per Tonaufnahme festgehaltenen – Ausgangsdaten auf das Hörbare in Kommunikationssituationen. Durch den technologischen Fortschritt, der die Arbeit mit digitalen Videodaten ermöglicht,<sup>21</sup> kann neben der Hörbarkeit auch die Sichtbarkeit interaktiver Bedeutungskonstitution analytisch zugänglich gemacht werden. Videoaufnahmen ermöglichen es für „another first time“ (Garfinkel 1967: 9) die detailgenaue Organisation menschlicher Interaktion zu beobachten. Das heißt, das gesamte körperliche Verhalten, das Teilnehmer:innen zur Herstellung und Aufrechterhaltung von Interaktion einsetzen, kann untersucht werden. Durch eine entsprechende multimodale Transkription (vgl. Mondada 2016b) können körperliche Aktivitäten in ein genaues Zeitverhältnis zu sprachlichen Äußerungen gesetzt werden. Dafür werden im Transkript Kürzel verwendet (z. B. „&“, „\*“ usw., siehe das Transkript unten), die den Start- und Endpunkt einer Aktivität festlegen. Außerdem kann das Verhältnis der

---

<sup>19</sup> Vgl. Stivers & Sidnell (2005: 2).

<sup>20</sup> ‚Talk-in-interaction‘ ersetzt als zentrale Beschreibung des Gegenstandes der Konversationsanalyse den ursprünglichen, namensgebenden Begriff dieses Ansatzes ‚conversation‘.

<sup>21</sup> Vgl. auch die Bezeichnung *video turn* bei Mondada (2006a) und Heath/Hindmarsh/Luff (2010).

Aktivitäten mehrerer Beteiligter präzise bestimmt werden. Wie ein solches Transkript aussieht, zeigt folgendes Beispiel aus einer Probe des *Haydn Orchester Bozen*:

---

19:36-19:43

```

01 D %&*(0.5)&* ~¤(1.3)¤ ¤(0.8) <<f> questa uno forte:>¤%
d &winkt ab&
    *blickt auf Partitur*
    ~Blick-->MM ----->
    ¤zieht Mundwinkel nach außen¤
        ¤zeigt mit linkem ZF nach vorne¤
mm %♪((spielen))♪ ----->%
```

02 D <<dim> &con& \*~riserva\*~ ¤¤#pe:r prima di DUE¤¤ forte>
 &bewegt linke Hand nach unten&
 \*hält LH auf Augenhöhe\*
 ~bewegt rechte Hand zur linken Hand
 ¤bewegt linke Hand nach oben¤
 ¤hält rechte Hand auf Augenhöhe¤

abb #Abb.3

---

Mithilfe der Erweiterung des ursprünglichen, auf die Verbalität fokussierten Transkripts (vgl. Kap. 2.1.3.2) durch kinesisch-visuelle Aspekte der Interaktion ist es möglich, die Verwendung einzelner Ressourcen genau nachzuvollziehen. Die Sprechpausen des Dirigenten in Z01 sind keine ‚leeren‘ Pausen, sondern ‚verbale Auszeiten‘, die durch Gesten, Blickrichtungen sowie Mimik gefüllt sind. In den ersten 0.5sec winkt der Dirigent die Musik ab und richtet den Blick auf die Partitur aus. In den nächsten 1.3sec orientiert er seinen Blick in die Richtung der Musiker:innen und zieht gleichzeitig die Mundwinkel nach außen. In den anschließenden 0.8sec zeigt er mit dem linken Zeigefinger nach vorne. Diese Gestik führt der Dirigent auch während seiner sprachlichen Äußerung am Ende von Z01 weiter (*questo uno forte*). Während dieser gesamten Sequenz spielen die Musiker:innen (noch). In Z02, während der Dirigent weiterhin spricht, kommen ebenfalls eine Reihe an unterschiedlichen Modalitäten zum Einsatz. Der Dirigent bewegt bei *con* die linke Hand nach unten, hält bei *riserva* die linke Hand auf Augenhöhe und nähert gleichzeitig die rechte Hand an die linke Hand an. Bei der Aussage *per prima di due* bewegt er die linke Hand nach oben, während er die rechte Hand auf Augenhöhe hält. Diese letzte Geste wurde auch per Screenshot festgehalten (im Transkript mit „abb“ gekennzeichnet):



**Abb. 3:** Linke Hand nach oben, rechte Hand auf Augenhöhe.<sup>22</sup>

Durch solche Screenshots kann die Ambiguität, die zumal von der Beschreibung von Aktivitäten ausgeht, reduziert werden. Außerdem kann das Zusammenspiel mehrerer Modalitäten (hier: Gestik in Kombination mit dem Blick zu den Musiker:innen, den der Dirigent auch in Z02 aufrechterhält) in seiner Veräußerung als multimodale Gestalt<sup>23</sup> (Mondada 2014e) exemplifiziert werden.

Das Beispiel zeigt, dass Multimodalität keine neue Analysemethode oder Methodologie darstellt, sondern eine konstitutive Eigenschaft von Interaktion ist. Interaktion ist als multimodal konstituiertes soziales Ereignis begreifbar: Die Modalitäten oder Ausdrucksressourcen sind von der Leiblichkeit der an der Interaktion Beteiligten nicht zu trennen, da diese ihren Körper als zentrale Ressource nutzen, um unterschiedliche Modalitäten zu realisieren (Putzier 2016: 44). Im Beispiel oben setzt der Dirigent den Körper ein, um zu gestikulieren, um in eine bestimmte Richtung zu blicken oder um einen bestimmten Gesichtsausdruck einzunehmen; die Musiker:innen nutzen ihre Körper, um auf ihren Instrumenten zu spielen. Die zentralen Ausdrucksressourcen in der Interaktion sind nach Mondada & Schmitt (2010: 24–25) folgende:

- Verbalität (Prosodie inbegriffen) und ihre Körperlichkeit;
- Vokalität;
- Blick;
- Kopfbewegung;
- Mimik;
- Gestikulation;
- Körperpositur;

---

<sup>22</sup> Bildrechte für alle Abbildungen in dieser Arbeit: Monika Messner; publiziert mit Einverständnis der abgebildeten Personen.

<sup>23</sup> Für eine Präzisierung des Terminus *multimodale Gestalt* vgl. den Abschnitt weiter unten (Kap. 2.1.5.3).

- Bewegungsmodus;
- Präsenzform;
- Proxemik, d. h. Verteilung im Raum, genaue Positionierung der Beteiligten (*face to face, side by side, back to face ...*), Nähe/Distanz;
- Räumlichkeit/Materialität: Manipulation von Objekten (Tools, Instrumente usw.).

Interaktion wird damit als ganzheitlicher Prozess angesehen, in dem nicht eine Modalität im Mittelpunkt steht, sondern ein „gleichzeitiges Zusammenspiel“ (Schmitt 2007b: 29) stattfindet. In diesem Zusammenspiel tragen alle Modalitäten gleichwertig zur Konstitution sozialer Bedeutung bei. Das heißt, aus einer multimodalen Analyseperspektive heraus sind sowohl die sequenzielle Abfolge von Handlungen als auch die simultane Abstimmung mehrerer Modalitäten zur Ausführung einer Handlung konstitutiv für die Herstellung interaktiver Ordnung. Nach Mondada & Schmitt (2010: 23) ist „Multimodalität [...] ein theoretischer Bezugsrahmen, der [...] auf eine Form der Interaktionsanalyse verweist, die auf der Basis von Videodokumenten authentischer Interaktionssituationen auf der Grundlage der konstitutionsanalytischen Methodologie der Konversationsanalyse durchgeführt wird.“ Diese multimodale Konzeption von Interaktion ist mit der Modifikation bestehender sowie der Konstitution neuer Untersuchungsgegenstände für die ethnomethodologische Konversationsanalyse verbunden.<sup>24</sup> Aus dieser Ausweitung und Re-Konzeptualisierung der klassischen Konversationsanalyse entstand in den letzten Jahren die *multimodale Konversationsanalyse* (Deppermann 2013, 2018b; Mondada 2014e, 2016a, 2019a, 2019b).

### **2.1.5.2 Das Verständnis von Multimodalität**

Laut Deppermann & Schmitt (2007: 16–18) geht das Verständnis von Multimodalität auf die von Sacks und Schegloff entwickelte Konversationsanalyse, Goffmans (1983) Konzeption der ‚interaction order‘ sowie auf den Forschungsansatz der ‚context analysis‘ (Scheflen 1964, 1972; Kendon 1990) zurück. Goffman (1983: 2) versteht soziale Interaktion als Zusammenkunft mehrerer Individuen in einer „face-to-face domain“. Das Konzept der ‚face-to-face domain‘ hebt die Bedeutung der Sichtbarkeit in der Interaktion hervor und deutet darauf hin, dass alles, was in einer sozialen Situation von den Interaktionsbeteiligten wahrnehmbar ist, grundsätzlich als interaktiv relevant zu betrachten ist und auch unter dieser Perspektive analysiert werden muss (Deppermann/Schmitt 2007: 16). Diese Konzeption der

---

<sup>24</sup> Für eine Auflistung dieser Untersuchungsgegenstände vgl. den Abschnitt weiter unten (Kap. 2.1.5.4).

,interaction order‘ verdeutlicht die Komplexität interaktiver Ordnung, für die nicht nur verbale Anteile, sondern auch nonverbale Verhaltensweisen konstitutiv sind. Zwar hat Goffman seine Beobachtungen, Vorstellungen und Konzepte nicht anhand empirischer Analysen entwickelt, allerdings sind seine Arbeiten auf unterschiedliche Forschungskontexte, wie z. B. die Konversationsanalyse, übertragbar. Werden die methodische Strenge in der Analyse der verbalen Interaktionsteile sowie die ,face-to-face‘-Vorstellung von Goffman zusammengeführt, so steht die empirische Analyse des Vollzugs der ,interaktiven Ordnung‘ im Mittelpunkt (Deppermann/Schmitt 2007: 17). Es soll erforscht werden, wie interaktive Ordnung und Struktur durch verschiedene Ausdrucksmittel, derer sich Beteiligte an einer Interaktion bedienen, hergestellt wird. Unter multimodaler Perspektive kann das Konzept der ,interaction order‘ reformuliert werden als „Gesamtzusammenhang aller simultan realisierten, sequenziell strukturierten und aufeinander bezogenen interaktiven Beteiligungsweisen aller Teilnehmer zu jedem Zeitpunkt im Interaktionsverlauf“ (Schmitt 2007a: 31).

Ein weiterer zentraler Bezugspunkt für die Konzeption multimodaler Forschung bildet der Forschungsansatz der ,context analysis‘ (vgl. Scheflen 1964, 1972; Kendon 1990). Die ,context analysis‘ hat bereits früh begonnen – auf der Basis von Goffmans Konzepten – empirisch mit Videoanalysen zu arbeiten. Kendon (1990: 16) betont „the importance of an integrated approach to the study of interaction“ und „refuses to assume that any particular modality of communication is more salient than another“. Das heißt, Simultaneität spielt eine wichtige Rolle als generativer Mechanismus multimodaler Bedeutungskonstitution und das Prinzip der Verbalität wird zugunsten eines integrierten Ansatzes aufgegeben, da alle Modalitäten als grundsätzlich gleichwertig betrachtet werden. Allerdings geht die ,context analysis‘ nicht sequenzanalytisch vor – wie es für die Konversationsanalyse üblich ist – und interessiert sich zumeist nicht unter konstitutionsanalytischer Perspektive dafür, wie unterschiedliche Aktivitäten bei der Herstellung eines Handlungszusammenhangs koordiniert werden. Vielmehr stehen die spezifischen Funktionen einer bestimmten Modalität bzw. eines bestimmten multimodalen Verfahrens im Vordergrund (Deppermann/Schmitt 2007: 18).

### **2.1.5.3 Die Entwicklung der multimodalen Konversationsanalyse**

Neben den drei beschriebenen Bezügen und deren Zusammenhängen bilden für die Entwicklung der multimodalen Konversationsanalyse die sogenannten ,gesture studies‘ einen wichtigen Ausgangspunkt. Diese untersuchen sehr früh den Zusammenhang zwischen Sprache und Gestikulation (Kendon 1980, 1994, 2004; McNeill 1985, 1992, 2000, 2005). Den ,gesture studies‘ zufolge werden Gesten zeitlich sehr präzise mit Sprache koordiniert (Müller 1998; Kendon 2004) und weisen dabei die-

selbe „improvisational quality as do words in conversation“ (Bavelas et al. 1992: 470) auf. Sie entstehen im Äußerungsprozess und produzieren gemeinsam mit der verbalen Äußerung sogenannte „visible acts of meaning“ (Bavelas/Chovil 2000) oder „visible action[s] as utterance“ (Kendon 2004). Außerdem sind Gesten auf das Verhalten der Gesprächspartner:innen abgestimmt, an die sie adressiert sind. Solche Gesten, die sich beispielsweise in einem Zeigen in die Richtung anderer Teilnehmer:innen äußern, werden als „interactive gestures“ (Bavelas et al. 1992) bezeichnet.

Allerdings bleibt in den ‚gesture studies‘ ein holistisches Verständnis von Interaktion, bei dem nicht nur Gesten, sondern auch andere unterschiedliche Ausdrucksressourcen eine Rolle spielen, außen vor.<sup>25</sup> Laut C. Goodwin (1981, 2000, 2007b) muss Gestikulation immer im Rahmen des gesamten ‚participation framework‘ verstanden werden, in das sowohl die Aktivitäten des/der Produzent:in als auch die des/der Adressat:in mit einbezogen sind. So zeigt z. B. M. H. Goodwin (1980), dass Gesten eng mit einem Monitoring dessen abgestimmt sind, was andere Interaktionsbeteiligte machen, oder Schegloff (1984), dass Gesten, die den propositionalen Gehalt einer Äußerung vorwegnehmen, von dem/der Adressat:in beobachtet und so interpretiert werden, dass er:sie in emergenter und relevanter Weise darauf reagieren kann. C. Goodwin (1979, 1981) konstatiert, dass für den Abschluss und auch für die Erweiterung eines Turns sowohl sprachliche als auch blickliche Ressourcen eingesetzt werden. Außerdem beobachten Goodwin & Goodwin (1987), dass ein Turn, in dem es darum geht, eine gemeinsame Beurteilung hervorzu bringen, durch ein Zusammenspiel von Syntax, Prosodie, Blick und Mimik abgeschlossen wird. Die Koordination einzelner Ressourcen spielt auch in der Weiterführung eines Turns eine Rolle, wie Ford/Fox/Thompson (1996) feststellen: Blick, Körperhaltung und Körperpositionierung wirken zusammen, wenn ein/e Sprecher:in einen Turn für sich behalten und weiter ausführen möchte.

Diesen und auch anderen multimodal orientierten Studien (vgl. z. B. Goodwin 1986b; Heath 1986, 1989; Streeck 1988, 1993) ist gemeinsam, dass nicht nur der Einsatz spezifischer Ausdrucksressourcen, wie Gestik, Mimik, Blickverhalten, Körperposition und Körperorientierung usw., in Verbindung mit Sprache systematisch untersucht wird, sondern auch die Koordination einzelner Ressourcen in ihrer gesamten Komplexität.<sup>26</sup> Die Arbeiten von Charles Goodwin (1979, 1980, 1981), Mar-

---

<sup>25</sup> Allerdings weist Kendon (1990) in seiner Arbeit zu F-Formationen auf das Teilen eines gemeinsamen Aufmerksamkeitsfokus der Interaktionsbeteiligten hin, der sich durch die Anordnung innerhalb einer F-Formation ergibt. Eine F-Formation ermöglicht es den Beteiligten, sich gemeinsam einer Aktivität zu widmen.

<sup>26</sup> Diese Koordinierung sowie der zusätzliche Miteinbezug von interaktiven Faktoren (etwa Kör pergebundenheit, Wahrnehmung, Raumbezogenheit, Simultaneität von verschiedenen Modalitäten usw.) zeigt sich vor allem in den zahlreichen Arbeiten zur Deixis in der face-to-face-Interaktion

jorie H. Goodwin (1980), Heath (1986) und Streeck (1988) in den 70er und 80er Jahren bereiten dafür den Weg. Sie analysieren alle in der Kommunikation zum Tragen kommenden Modalitäten, sowohl in alltäglichen (Tischgespräche) als auch in institutionellen Interaktionssituationen (Arzt-Patient-Gespräche). Damit legen sie den Grundstein für ein holistisches Verständnis von Interaktion, wo multimodale Ressourcen von unterschiedlichen Beteiligten koordiniert und elaboriert werden. Das Untersuchungsinteresse entfernt sich demnach von einer Fokussierung auf einzelne Modalitäten, die in Kombination mit Sprache auftreten, und bewegt sich hin zu einer Annäherung an Interaktion als „embodied and environmentally situated organization of activities“ (Broth/Mondada 2013: 42), und damit als leibliches Handeln.<sup>27</sup>

Dieser ‚embodied interaction‘ widmen Streeck/Goodwin/LeBaron (2011) einen Sammelband, in dem sie eine videobasierte, naturalistische und mikroanalytische Herangehensweise an die Erforschung sozialer Interaktion vorschlagen. In dem Band werden Möglichkeiten dafür aufgezeigt, wie Videodaten analysiert werden können (Beiträger:innen arbeiten beispielsweise mit Fotos, Zeichnungen und Screenshots). Dabei steht immer die Frage im Mittelpunkt, wie Beteiligte in ko-präsenter Interaktion durch Zuhilfenahme von unterschiedlichen semiotischen Ressourcen interaktive Handlungen produzieren. Denn – so halten die Herausgeber fest – „the simultaneous use of diverse semiotic resources is pervasive in the organization of endogenous human action.“ (Streeck/Goodwin/LeBaron 2011: 4)

In den letzten Jahren entstanden zahlreiche Studien, die sich mit der gesamten Komplexität zwischenmenschlicher Kommunikation auseinandersetzen (vgl. z. B. Mondada 2014e, 2016a; Goodwin 2017; Keevallik 2018). Das heißt, das Zusammenspiel unterschiedlicher Modalitäten wird nicht nur auf intrapersonaler, sondern auch auf interpersonaler Ebene untersucht. Denn Multimodalität inkludiert

[...] an interest not only in talk, gesture, and gaze, but more radically in the entire body – body posture, orientation, body-torque, and body movements. This not only concerns the individual participants, and their simple coordination, but also concerns the interactional space they visibly, dynamically, and specifically design and configure within the ongoing course of action. (Mondada 2019b: 64)

Dabei ist Sprache eine von vielen möglichen Modalitäten, es gibt keine a priori festgelegte Hierarchie. Das heißt, alle Ressourcen können gleichermaßen relevant für die Organisation von Aktivitäten und Interaktion und damit auch für die Herausbildung von komplexen multimodalen Gestalten sein (Mondada 2014e):

---

(vgl. u. a. Heath 1986, 2002; Goodwin 1994, 2003; Hindmarsh/Heath 2000; Clark 2003; Mondada 2007c, 2012a; Stukenbrock 2008, 2009a, 2009b, 2014a, 2018).

<sup>27</sup> Vgl. auch die Bezeichnung *embodied turn* bei Neville (2015) und bei Deppermann & Streeck (2018).

[...] human action is organized by exploiting a large array of multimodal resources, mobilized and packaged in an emergent, incremental, dynamic way, in response to the contingencies of the setting and of the interaction. (Mondada 2014e: 140)

Ein Archäologe beispielsweise, der auf den Boden schaut, gleichzeitig in einer gewissen Körperposition seinem Kollegen etwas erklärt, Objekte handhabt, gestikuliert und sich mit seinem Gesprächspartner koordiniert, produziert eine komplexe multimodale Gestalt (vgl. Goodwin 2000, 2017). Diese neue Herangehensweise an Interaktion und Multimodalität zieht auch ein verstärktes Interesse an Objekten und deren Einsatz und Relevanzsetzung in der Interaktion mit sich (vgl. z. B. Goodwin 2010). Studien zeigen, dass Objekte für die Organisation sozialer Interaktion (Robinson/Stivers 2001; Mondada 2006b; Day/Wagner 2015) oder als Tools für spezifische Aktivitäten (Goodwin 1994: Archäologische Ausgrabungen; Mondada 2011a, Heath et al. 2018: Operationssaal) eingesetzt werden. Darüber hinaus werden mehrere parallel verlaufende Aktivitätsstränge in der Interaktion („Multiaktivität“, z. B. Autofahren und die gleichzeitige Koordination von Sprache, Blick und Gestik, vgl. Haddington et al. 2014), Formen der leiblich-sozialen Koordination („Interkorporealität“, vgl. Cekaite 2010; Streeck 2013; Meyer/Streeck/Jordan 2017; Goodwin/Cekaite 2018), das Zusammenspiel mehrerer Sinne („Multisensorialität“, vgl. Goodwin/Cekaite 2018; Mondada 2018b) sowie Bewegung im Raum und Mobilität (Haddington/Mondada/Nevile 2013; Mondada 2016a) untersucht.

#### **2.1.5.4 Re-Modifikation und Neu-Konstitution von Untersuchungsgegenständen**

Die Verschiebung des Erkenntnisinteresses der Konversationsanalyse hin zur Analyse leiblichen Handelns in der Interaktion ist auch mit der Modifikation bestehender sowie der Konstitution neuer Untersuchungsgegenstände verknüpft. In der multimodalen Konversationsanalyse werden multimodale Praktiken untersucht, die Teilnehmer:innen in der Interaktion benutzen (Deppermann 2018b: 58). Diese bestehen aus einer Trias aus dem formbestimmten Einsatz multimodaler Ressourcen (Wie, Form), der Funktion einer Handlung (Wozu, Funktion) und der Gebundenheit dieser Ressourcen an bestimmte Kontexte (Wann, Kontext) (Deppermann 2018b: 58). Die multimodale Perspektive erweitert auch den Blick auf die Zeitlichkeit in der Interaktion (Deppermann 2018b: 62). In der klassischen Konversationsanalyse bildete Sequenzialität die Norm und sog. „overlaps“, bei denen es zu einer Gleichzeitigkeit von Aktivitäten unterschiedlicher Interaktionsbeteiligter im gleichen Modus kommt (z. B. wenn zwei oder mehrere Beteiligte gleichzeitig sprechen), wurden als „systemkonforme“ Ausnahmen konzeptualisiert (Deppermann/Schmitt 2007: 29; vgl. auch Sacks/Schegloff/Jefferson 1974). Wenn der Fokus allerdings auf Multimodalität und damit auf die Gleichzeitigkeit multimodaler Aktivitäten verschoben wird, wird die Ausnahme zu einem für die Interaktion „konstitutive[n]

Faktum“ (Deppermann/Schmitt 2007: 29). Das heißt, in leiblicher Interaktion sind Sequenzialität und Simultaneität gleichermaßen konstitutiv:

[S]equentiality might not be organized turn-at-talk after turn-at-talk, strictly successively, but rather in parallel flows of action, as emergent embodied conduct responds to a previous action and unfolds simultaneously with it. (Mondada 2016a: 346)

Durch den Miteinbezug von simultanen Ereignissen in der Interaktion, der sich beispielsweise darin äußert, dass sich die Turn-Produktion an simultanen non-verbalen Rezipient:innen-Handlungen orientiert, ist auch die ursprüngliche Sprecher-Hörer-Dichotomie irreführend. Besser ist die Bezeichnung ‚participant‘ – die bereits Goffman (1979) vorgeschlagen hat<sup>28</sup> da der Terminus sowohl verbale als auch nonverbale Handlungen der Teilnehmer:innen mit einschließt. Damit einher geht auch die Einführung von neuen Untersuchungsgegenständen, etwa Koordination, (Interaktions-)Raum und Objekte. Koordination wird als ein auf unterschiedlichen Ausdrucksebenen hervorgebrachtes, von allen Beteiligten zu jedem Zeitpunkt gemeinsam konstituiertes und praxeologisch gerahmtes Ereignis begriffen, das sowohl intrapersonal (ein/e Teilnehmer:in koordiniert seine/ihre eigenen Aktivitäten) als auch interpersonal (mehrere Teilnehmer:innen stimmen ihre Aktivitäten aufeinander ab) geschieht (Deppermann 2018b: 60). Interagieren mehrere Beteiligte miteinander, so spielt auch Raum als Begrenzung (z. B. Begehbarkeit, Sichtbarkeit usw.), als Ressource (z. B. die Nutzung bestimmter Raumabschnitte für spezifische Handlungen) sowie als Produkt der Interaktion (z. B. Konfigurationen, Eigenschaften und Objekte, die in der Interaktion relevant gesetzt und hergestellt werden) eine Rolle. Raum wird von den Interaktionsbeteiligten nicht als etwas Gegebenes betrachtet, sondern muss als für die Interaktion wichtiger Relevanzrahmen vielmehr für die jeweiligen Zwecke in Form von „Interaktionsräumen“<sup>29</sup> situativ hergestellt werden (Mondada/Schmitt 2010: 27; vgl. auch Mondada 2007b; Hausendorf/Schmitt/Kesselheim 2016; Hausendorf/Schmitt 2018). Nicht zuletzt stellen auch Objekte Ressourcen für die Interaktion dar (vgl. Nevile et al. 2015). Objekte erzeugen Bedingungen und Ereignisse, auf die Interaktionsteilnehmer:innen reagieren, wobei die interaktive Nutzung häufig nicht der konventionellen entspricht (z. B. wenn ein Stift zum Zeigen oder zur Beanspruchung des Rederechts benutzt wird, vgl. Mondada 2007c).

---

<sup>28</sup> Goffman (1979) unterscheidet zwischen Personen, die an der Produktion einer verbalen Äußerung beteiligt sind (z. B. als *animator*, *author* oder *sponsor*), und Personen, die als Rezipient:innen an einem kommunikativen Ereignis teilnehmen können (z. B. als *ratified* oder *non-ratified participant*, als *overhearer* oder als *eavesdropper*).

<sup>29</sup> Vgl. auch „interactional space“ bei Kendon (1979, 1990).

Dem Forschungsansatz der multimodalen Konversationsanalyse folgend wird soziale Interaktion durch beobachtbare Handlungen und Praktiken der Beteiligten organisiert, durch die sie sich wechselseitig anzeigen, wie sie die Aktivitäten der Interaktionspartner:innen interpretieren. Für solche Handlungen greifen Teilnehmer:innen auf körperliche Ressourcen zurück, die im Laufe der Interaktion auch neu konstituiert („(re)constituted“) und ausgehandelt werden können („(re)negotiated“, Mondada 2016a: 361). Diese holistische Konzeption von Multimodalität antwortet auf die ursprüngliche Aufforderung der Konversationsanalyse und der Ethnomethodologie, die *accountability* von Handlungen zu untersuchen (vgl. Garfinkel 1967; Sacks 1992): „how action is accountably produced and how it is inter-subjectively received and understood“ (Mondada 2016a: 360). Die ‚next-turn proof procedure‘ (Sacks/Schegloff/Jefferson 1974), d. h. die unmittelbar folgenden Reaktionen der Rezipient:innen in den nächsten Gesprächsbeiträgen, mit denen diese ihr Verständnis des vorhergegangenen Beitrages zu erkennen geben, wird damit zu einer ‚next-action proof procedure‘. Durch solche ‚next-action proof procedures‘ kann der/die Analytiker:in sehen, wie Beteiligte sequenziell und simultan auf multimodale Gestalten reagieren.

### 2.1.6 Dirigiergesten vs. sprachbegleitende Gesten

---

```

01 D &%FORTE! (1.2) *(1.0)*&
d &dirigiert ----->&
*schüttelt Kopf*
mm %♪((spielen 6.5''))♪----->
02 D &(4.0)&%
d &winkt ab&
mm --->%
[...]
06 D (1.2) *(0.4)* ~¤let's go BACK again¤~
*hebt Arme*
~geht in die Knie~
¤bewegt Arme nach unten¤
Dirigentin und Musiker:innen30

```

---

<sup>30</sup> Das Transkript dokumentiert einen Ausschnitt aus einer Probe des *Haydn Orchester Bozen* mit der Dirigentin Sybille Werner. Die ausführliche Analyse des zugrundeliegenden Videoausschnitts erfolgt in Abschnitt 6.4.1.

Dieser kurze Ausschnitt aus einer Probe des *Haydn Orchester Bozen* mit der Dirigentin Sybille Werner zeigt, dass sich in der Orchesterprobe Dirigiergesten (vgl. Z01: *dirigiert*) mit sprachbegleitenden Gesten (vgl. Z06: *hebt Arme, geht in die Knie, bewegt Arme nach unten*) abwechseln. Dirigiergesten werden von der Dirigentin während einer Spielphase verwendet (die Musiker:innen spielen in Z01, während die Dirigentin dirigiert), sprachbegleitende Gesten kommen in Besprechungsphasen zum Einsatz. Gesten<sup>31</sup> spielen auch eine wichtige Rolle, wenn es darum geht, eine Unterbrechung der Musik einzuleiten, wie es in Z02 der Fall ist (*winkt ab*). Die Dirigentin führt mit ihren Händen eine abwinkende Bewegung aus und gibt den Musiker:innen dadurch zu verstehen, dass sie mit dem Spielen aufhören sollen. Die Gestik fungiert hier als körperliches Zeichen (*body cue*), um die Transition von einer Spiel- zu einer Besprechungsphase zu koordinieren. Für die Musiker:innen reicht diese Geste aus, damit sie die Musik auf ihren Instrumenten abklingen lassen: Sie hören in Z02 mit dem Spielen auf. Außerdem ist es möglich, dass Dirigiergesten mit Sprache („*forte*“ in Z01) oder auch mit sprachbegleitenden Gesten während der Musik koordiniert werden (*schüttelt Kopf* in Z01).

Die Funktionen von Gesten in der Orchesterprobe können also vielseitig sein:

- a) Orientierungs-, Koordinierungs- und Gestaltungshilfe für die Musiker:innen während der Musik (Z01, Dirigiergeste);
- b) negative Evaluierung der Umsetzung des instruierten *forte* (Z01, Kopfgeste);
- c) visueller Hinweisreiz für die Einleitung von Unterbrechungen der Musik und damit von Transitionen zwischen Spiel- und Besprechungsphasen (Z02, Handgeste);
- d) visuelle Illustrierung von korrekten Instruktionen (Z06, Arm- und Beingeste).

Dabei sind Dirigiergesten weitestgehend kodifiziert<sup>32</sup> und werden so auch in Dirigierhandbüchern dargestellt: Mit der rechten Hand/dem rechten Arm werden das Metrum (Takt) sowie Anfang und Ende (Organisation) angegeben, mit der linken Hand/dem linken Arm werden expressive Qualitäten und andere Aspekte wie Dynamik, Stil, Klangfarbe, Artikulation, Warnungen sowie Einsätze angezeigt (Bräm/Boyes Bräm 1998: 221). Dirigiergesten, vor allem jene der linken Hand, gleichen – wie Bräm & Boyes Bräm in ihrer Klassifizierung von Ausdrucksgesten

---

<sup>31</sup> Im allgemeinen Verständnis wird der Ausdruck *Geste* mit kommunikativen Bewegungen der Arme und Hände, also der Vordergliedmaßen, in Verbindung gebracht. Diese Auffassung von Gestik wird teils auch in der Gestenforschung vertreten. Nach Müller (1998: 30) etwa wird der Terminus *Geste* „als Bezeichnung der kommunikativen Bewegungen der Arme und Hände“ verwendet. Auf ähnliche Weise definiert Goldin-Meadow (2003: 4) *gesture* als „the hand movements that co-occur with speech“.

<sup>32</sup> Vgl. auch die Ausführungen zu semiotischen Werten von Gesten weiter unten.

bei Dirigent:innen feststellen – häufig alltäglichen Gesten, da sie „eine metaphorische Verbindung zwischen musikalischen Aspekten und physischen Erfahrungen her[stellen], die Menschen mit Objekten im täglichen Leben machen“ (Bräm/Boyes Bräm 1998: 238). So drückt beispielsweise eine Dirigiergeste, bei der der/die Dirigent:in seine/ihre linke Hand hin zum Gesicht bewegt und dabei den Daumen und den Zeigefinger berührt („Pinzetten-Stellung“ bei Bräm/Boyes Bräm 1998: 229) aus, dass ein „schlanker, dünner, kontinuierlicher Ton“ gespielt werden soll. Der/die Dirigent:in greift hier metaphorisch nach dem Ton des spielenden Instruments und zieht diesen aus dem Instrument heraus.

Aus solchen spezifischen Linke-Hand-Gesten des/der Dirigent:in versuchen Bräm & Boyes Bräm ‚lexikalische‘ Bedeutungen abzuleiten. Dabei stützen sie sich auf die Konzeptualisierungstheorie für Metaphern<sup>33</sup> (oder auch *conceptual metaphors*) von Lakoff & Johnson (1980), die ebenso in der Musikwissenschaft für die Analyse der Beziehung zwischen Sprache und Musik angewandt wird. Dirigent:innen setzen Gesten ein, die mit der Manipulation von Objekten zu tun haben (z. B. greifen, berühren), mit biologischen Faktoren verknüpft sind (z. B. riechen, hören), visuell orientiert sind (z. B. zeichnen, malen) oder Sprechakt-Charakter haben (z. B. erhobener Zeigefinger für „Achtung!“) (Bräm/Boyes Bräm 1998: 238). Ausgehend von diesen Beobachtungen erstellen die Autor:innen ein Inventar an Ausdrucksgesten der linken Hand von Dirigent:innen:

- a) Klänge wie Objekte handhaben (z. B. Objekte greifen, etwas schlagen, etwas in Bewegung setzen, das Instrument spielen, zeichnen/malen, etwas berühren);
- b) Hinweise auf Körperteile als Metaphern (z. B. Herz für Emotion, Ohr für Hören/Intonation, Lippen für leiser oder sinnlich, Nase für Klangqualität);
- c) räumliche Form und Wege als Metaphern für musikalische Strukturen (z. B. Aufzeigen der Form, Aufzeigen des Weges);
- d) vertikale Orientierung im Raum als Metapher für Dynamik (z. B. nach oben = mehr, nach unten = weniger);
- e) Aufwärtsbewegung und Öffnung der Hand als Metapher für strahlende Klangqualität;
- f) Sprechaktgesten (z. B. Kreisen für „Vorwärts!“, Zeigefinger für „Achtung!“; flache Hand, Blickkontakt und positiver Gesichtsausdruck für Ermutigung).

---

<sup>33</sup> Lakoff & Johnson (1980: 5) definieren (konzeptuelle) Metaphern als „understanding and experiencing one kind of thing in terms of another“. Das heißt, zwei verschiedene konzeptuelle Domänen verknüpfen sich auf systematische Art und Weise: Eine Quelldomäne wird auf eine Zieldomäne projiziert, es findet eine metaphorische Übertragung statt. Dies ist beispielsweise zwischen den Domänen Krieg und Argumentation der Fall: Metaphern wie *verteidigen*, (*Argumenten*) *ausweichen*, *angreifen* usw. konzeptualisieren argumentative (Sprach-)Handlungen (Zieldomäne) als kriegerische Handlungen (Quelldomäne) (vgl. Ziem 2015: 53).

Auch Parton & Edwards (2009) untersuchen Gesten von Dirigent:innen und stellen dabei Unterschiede in den musikalischen Antworten von Ensemblemusiker:innen fest, ausgehend von Änderungen in der Gestik der Dirigent:innen. Die Autor:innen entschlüsseln beispielsweise, dass die Musiker:innen mehr Zeit benötigen, um auf die Gestik der rechten Hand des/der Dirigent:in zu reagieren (diese ist während des Dirigierens für zeitliche Koordination und Rhythmus zuständig), wenn diese/r die Hand vor (und nicht neben) dem Oberkörper hält. Parton & Edwards (2009: 73) argumentieren, dass die Gesten der Dirigent:innen eine interktionale Funktion haben: „[...] conductor gesture is associated with responses from musicians, and spatial orientation of conductor gesture affects how musicians respond to that gesture.“

Ebenso gehen Poggi & Ansani (2017) von der Annahme aus, dass Dirigiergesten systematisch eingesetzt werden und zum geteilten Wissen von Dirigent:innen gehören. Sie untersuchen in Proben und Konzerten, wie unterschiedliche Dirigent:innen die dynamischen Parameter *piano* (leise), *forte* (laut), *crescendo* (lauter werden) und *diminuendo* (leiser werden) durch ihre Gestik ausdrücken. Die Autor:innen stellen fest, dass Dirigent:innen zum Anzeigen einer Veränderung in der Dynamik unterschiedliche Gesten einsetzen:

- symbolische Gesten (z. B. Zeigefinger auf der Lippe für *piano*);
- direkte ikonische Gesten (z. B. beide Arme nach vorne gestreckt, Handflächen zeigen nach oben, Finger bewegen sich leicht zum Dirigenten für *crescendo*);
- indirekte ikonische Gesten (z. B. Ausdruck der Emotion Wut durch Gestik, um ein *forte* anzuzeigen).

Solche Dirigiergesten weisen metaphorisch oder metonymisch auf dynamische Aspekte in der Musik hin. Wenn ein/e Dirigent:in etwa seine/ihre Fäuste ballt, so vermittelt er:sie implizit auf metonymischer Ebene, dass die Musiker:innen mit derselben Kraft spielen sollen (*forte*), mit der er:sie seine/ihre Muskeln in den Händen anspannt. Wie Bräm & Boyes Bräm (1998) vertreten auch Poggi & Ansani (2017: 118) die Ansicht, dass Dirigiergesten starke Parallelen zu Gesten aufweisen, die im Alltag eingesetzt werden: „These gestures do not make up a specialist jargon, rather they are quite similar to everyday gestures and likely comprehensible by laypeople.“

Diese Verbindung zwischen Alltags- und Dirigiergesten kann sich auch in umgekehrter Richtung äußern, das heißt, Gesten, die zum Dirigieren verwendet werden, können auch sprachbegleitend eingesetzt werden. In dem einleitenden Beispiel benutzt die Dirigentin in der Besprechungsphase eine Geste, die aus der Gestenkategorie der vertikalen Orientierung als Metapher für die Dynamik von Bräm & Boyes Bräm (1998) abgeleitet werden kann. Die Dirigentin geht in die Knie und bewegt dabei ihre Arme nach unten (die Handflächen zeigen nach unten), um anzudeuten, dass die Musiker:innen „weniger“ bzw. „leiser“ spielen sollen (vgl. auch

die verbale Äußerung „let's go back again“ in Verbindung mit der Geste). Diese Geste wird hier sprachbegleitend in einer Besprechungsphase eingesetzt, mit dem gleichen Ziel, wie sie auch als Dirigiergeste in einer Spielphase gebraucht wird: zum Anzeigen eines dynamischen Aspekts.

Das bedeutet für sprachbegleitende Gesten in Orchesterproben, dass diese bis zu einem bestimmten Grad konventionalisiert und ähnlich kodiert sein können (z. B. kann die abwinkende Handbewegung zum Unterbrechen der Musik bei unterschiedlichen Dirigent:innen beobachtet werden), während andere Gesten weniger eindeutig zu bestimmen sind und von Körperbau, Persönlichkeitsprägungen und individueller Körpersprache abhängen. Gemeinsam ist den sprachbegleitenden Gesten von Dirigent:innen, dass die Hände – teils auch in Verbindung mit dem Dirigierstock –, Kopf- und Körperbewegungen zu ‚stillen Instrumenten‘ werden, die illustrativ, ikonisch und musikalisch performieren. Das heißt, unter Gestik sind nicht nur die kommunikativen Bewegungen der Finger, Hände und Arme zusammengefasst, sondern eine Geste umfasst die gesamte „visible bodily action“ (Kendon 2004: 7). Haviland (2004: 206) beispielsweise führt neben Fingern, Händen und Armen auch andere Körperteile an (Gesicht, Lippen, Kinn, Blick, Kopf, Körperpositur, Torso- und Schulterbewegungen, Beine), die als gestische Ausdrucksressourcen fungieren können. Diese Auffassung von Gestik deckt sich mit dem, was Poggi & Ansani (2017: 111) hinsichtlich der Gestik von Dirigenten feststellen: „[...] the conductor's body signals are all, by definition, communicative.“ Außerdem werden sprachbegleitende Gesten nicht willkürlich eingesetzt, sondern die Wahl einer bestimmten Geste erfolgt immer bewusst und orientiert sich am kommunikativen Zweck sowie am Ziel der gestischen Nachahmung (Bressem 2018: 39). Neben diesem Fokus auf Dinge und Handlungen in der Welt zeigen Gesten auch als „Fenster zum Geist“ (McNeill/Duncan 2000: 42) an, wie der/die Sprecher:in darzustellende Handlungen und Dinge konzeptualisiert.<sup>34</sup>

Darüber hinaus kann die sprachbegleitende Geste der Dirigentin in dem Beispiel oben – angelehnt an die Gestentypologien von McNeill (1992, 2000, 2005) und Kendon (2004) – als ikonische/darstellende Geste bezeichnet werden. Ikonische Gesten bilden neben deiktischen Zeigegesten, metaphorischen Gesten und Taktstock-Gesten eine der vier Subtypen rede- oder sprachbegleitender Gesten. Unter ikonische Gesten werden Körperbewegungen gefasst, die konkrete Entitäten oder Handlungen repräsentieren und eine enge Verbindung zu einem konkreten Element der verbalen Äußerung aufweisen (z. B. gestische Darstellung

---

<sup>34</sup> Allerdings halten andere Studien daran fest (Kendon 1990, 2004; Goodwin 2003; Heath/Luff 2007; Streeck 2009), dass Gesten damit erklärt werden müssen, wie sie zum interaktionalen Prozess und zum intersubjektiven Verständnis beitragen, und nicht damit, was sie über mentale Prozesse aussagen.

von Klavierspielen). Zeigegesten zeichnen sich dadurch aus, dass – prototypisch mit dem ausgestreckten Zeigefinger (vgl. McNeill 1992, 2000; Kendon 2004) – auf eine Person oder ein Objekt hingewiesen wird, um diese oder dieses als gemeinsamen Referenten des Gesagten zu identifizieren oder um die Aufmerksamkeit der Gesprächsbeteiligten darauf zu lenken (vgl. z. B. Mondada 2014c). Metaphorische Gesten übertragen konkrete Objekte bildlich in eine abstrakte Domäne, etwa wenn beide Hände (mit den Handflächen nach oben) eine Aufteilung hinsichtlich positiver und negativer Aspekte einer Angelegenheit andeuten. „Beats“ (McNeill 1992) oder Taktstockgesten (Müller 1998: 103) sind Gesten, bei denen ein Körperteil eine Hin-und-Her-Bewegung oder eine Auf-und-Ab-Bewegung beschreibt, die mehr oder weniger parallel zum Sprechrhythmus verläuft. Solche Gesten werden eingesetzt, um – wie ein Taktstock – das Gesagte zu strukturieren und bestimmte Aspekte davon hervorzuheben. Diese Gestentypen dürfen allerdings nicht als streng abgegrenzte und sich wechselseitig ausschließende Kategorien verstanden werden, sondern sie können auch ineinander übergehen oder sich überlagern (Kendon 2004: 104; Stukenbrock 2014a: 22).

Streeck (2009) geht in seiner Gestentheorie über jene von McNeill und Kendon hinaus, indem er einzelne Gesten nicht isoliert als semiotische Bedeutungsträger oder pragmatische Instrumente betrachtet, sondern sich mit der Organisation von Gesten innerhalb körperlicher Interaktion beschäftigt (vgl. Meyer 2010). Er setzt Gesten in eine sequenzielle Beziehung mit Sprache und Turn-taking und zeigt, wie Gestik genutzt wird, um soziales und interaktionales Handeln zu organisieren. Streeck (2009) teilt Gesten ebenfalls anhand unterschiedlicher Kategorien ein. Die erste Kategorie bilden „environmentally coupled gestures“ (mit Bezug auf Goodwin 2007a), worunter Zeigegesten fallen, oder auch Gesten, die anzeigen, wie etwas getan/ausgeführt werden soll. Zur zweiten Kategorie gehören Gesten der „Darstellung“ (*depiction*), die identisch mit den ikonischen Gesten von McNeill sind. Streeck führt für diese Kategorie das Beispiel eines Architekten an, der die architektonische Struktur eines Hauses verbal und gestisch erläutert. Daneben gibt es noch die Kategorie „Ceiving“ oder „Thinking by hand“, d. h. Gesten, die einen Inhalt visuell konzeptualisieren. Dabei drückt Gestik das Denken und Fühlen nicht aus, sondern *ist* es. In die vierte Kategorie fallen pragmatische oder meta-pragmatische Gesten; diese fasst Streeck unter „Speech-handling“ zusammen. Gesten sind dann pragmatisch, wenn sie selbst kommunikative Funktion ausüben (z. B. die Stopp-Hand, die den/die Gesprächspartner:in vom Sprechen abhält). Streeck (2009) betont dabei, dass Gesten nur in ihrer Umgebung verstanden werden können, jedoch nicht, wenn sie aus dem Setting herausgetrennt und isoliert betrachtet werden.

Eine weitere Beobachtung, die sich aus dem Beispiel oben ableiten lässt, ist, dass (sprachbegleitende) Gestik und Verbales beim Vermitteln von Bedeutung zusammenwirken. Die Dirigentin gibt verbal eine Anweisung („let's go back again“) und

unterstützt diese durch eine Geste, die das Verbale visuell konzeptualisiert (*geht in die Knie, bewegt Arme nach unten*). Die Geste bildet hier mit Sprache gemeinsam eine Äußerung: „Speech and movement appear together as manifestations of the same process of utterance.“<sup>35</sup> (Kendon 1980: 208) Sprache und Gesten bilden dabei nicht nur strukturell, sondern auch auf funktionaler Ebene eine Einheit. Durch das Zusammenspiel der unterschiedlichen Modalitäten wird ein Rede-Geste-Ensemble kreiert (Kendon 2004: 127), in dem sowohl Verbales als auch Gestisches an der Herstellung einer multimodalen Äußerung beteiligt sind.<sup>36</sup> Daraus lässt sich eine enge Verknüpfung von Gestik und Sprache ableiten, innerhalb derer Gesten in der Lage sind, Sprachliches zu ergänzen, zu modifizieren oder auch zu ersetzen, und – ähnlich wie Sprache – dazu eingesetzt werden, „die Gedanken, Gefühle und Intentionen des Sprechers zum Ausdruck zu bringen und die soziale Ordnung des Gesprächs aktiv herzustellen“ (Müller 1998: 13).

Nicht zuletzt kann an dem kurzen Ausschnitt aus der Probe des *Haydn Orchester Bozen* mit Sybille Werner abgelesen werden, dass die Dirigentin bereits während der Sprechpause in Z06 (0.4sec) ihre Arme nach oben hebt und damit die anschließende Geste – bei der sie in die Knie geht und die Arme nach unten bewegt – vorbereitet. Diese Vorbereitung kann als *preparation* und damit als eine der Bewegungsphasen innerhalb einer sog. ‚Gestenphrase‘ – wie Kendon (2004: 111–113) kinetische Abläufe bezeichnet – beschrieben werden: Der Körper oder ein Körperteil wird in eine Position gebracht, in der der eigentliche, bedeutungsvolle Teil der Geste beginnt. Die Geste der Dirigentin erreicht ihren Höhepunkt (*apex*) in der Durchführung der Nach-unten-Bewegung der Beine und der Arme (*stroke*). An diese Durchführungsphase kann ein Nachhalt (*post-stroke hold*) bzw. ein Anhalten der Aktivität anschließen. Die Durchführungsphase (*stroke*) und der fakultative Nachhalt (*post-stroke hold*) bilden den ‚Nukleus‘ (*nucleus*) einer Gestenphrase: „Der Nukleus weist die größte gestalthafte Präzision auf und konstituiert funktional die Bedeutung der Gestenphrase.“ (Stukenbrock 2014a: 23) Dieser Kern einer Gestenphrase wird durch die vorausgehende, aus der Ruhe- oder Ausgangsposition des Körperteils (*position of rest or relaxation*) eingeleitete Vorbereitungsphase und die anschließende Rückzugsphase des gestikulierenden Körperteils in die Ruheposition gerahmt. An eine Gestenphrase muss nicht zwingend ein Rückzug des Körperteils in die Ruheposition anschließen, sondern es kann auch ein Übergang zu einer neuen Gestenphrase stattfinden. Mehrere solcher Gestenphasen bilden eine Gesteneinheit (*gesture unit*). Mit

---

35 Vgl. auch McNeill (1992: 2), der feststellt: „gesture and language are one system.“

36 Vgl. auch „environmentally coupled gestures“ bei Goodwin (2007a): Gesten müssen in einen Kontext eingebettet sein, um verständlich zu sein.

der Rückkehr in die Ruheposition kommt eine Gesteneinheit – bestehend aus einer oder mehrerer Gestenphrasen – zum Abschluss.

### 2.1.7 Höflichkeit und Face

The human personality is a sacred thing; one dare not violate it  
nor infringe its bounds, while at the same time the greatest good  
is in communion with others.  
Émile Durkheim<sup>37</sup>

Höflichkeit ist ein omnipräsentes Phänomen sowohl in alltäglichen als auch in institutionellen Interaktionskontexten. Goffman (1971: 62) versteht Höflichkeit als „Ritual“,<sup>38</sup> das eine wiederkehrende Verhaltensweise von sozialen, harmoniegeleiteten Individuen in sich fasst, die ständig versuchen, ihr Face oder öffentliches Selbstbild (das jede:r für sich in Anspruch nehmen will) zu wahren. Die *Klassiker* in der Höflichkeitsforschung Brown & Levinson (1978/1987: 61) schreiben, dass Höflichkeit auf Kooperation aufbaut: „In general, people cooperate (and assume each other's cooperation) in maintaining face in interaction, such cooperation being based on the mutual vulnerability of face.“

Dieses Verständnis von Höflichkeit oder *politeness* aus linguistischer Perspektive ist nicht zu verwechseln mit Höflichkeit im alltäglichen Sprachgebrauch. Höflichkeit in der Alltagssprache bedeutet, in seinem Verhalten anderen Menschen gegenüber aufmerksam und rücksichtsvoll zu sein, das heißtt, so wie es die Umgangsformen gebieten (vgl. Duden online 2019). Höflichkeit steht demnach in engem Zusammenhang mit Umgang und ist durch eine Ausrichtung am Gegenüber charakterisiert (vgl. Zarend 2015).

In der Linguistik ist Höflichkeit seit den 1970er Jahren in der linguistischen Pragmatik ein viel diskutiertes Thema. Linguistisch betrachtet wird Höflichkeit als zielorientierter, rationaler Prozess beschrieben, der auf Ausgewogenheit und soziale Harmonie abzielt (Fraser 2001: 1407). Unter Höflichkeit werden verbale, paraverbale und nonverbale Verhaltensformen gefasst, die – anscheinend – universelle Geltung haben (vgl. Held 1995). Das heißtt, „höfliche“ Kommunikation – wie auch Kommunikation generell – steht im Dienst des sozialen Harmoniestrebens (Held 2010: 6).

---

<sup>37</sup> Vgl. Durkheim (1974: 37).

<sup>38</sup> Goffmans (1971) Ritualkonzeption ist eng an Durkheims (2014 [1912]) Bezeichnung für Handlungsmuster angelehnt, die der symbolischen Repräsentation von kollektiven Glaubensvorstellungen in einer religiösen Welt dienen.

Dabei muss in Betracht gezogen werden, dass es sich bei Höflichkeit – wie sie in der Pragmatik aufgefasst wird –, nicht um tatsächliche Höflichkeit handelt, sondern um ein Anscheinwahren, um eine bestimmte Haltung, die Beteiligte an einer Kommunikation einnehmen. Höflichkeit umfasst oberflächliche Harmonisierungsstrategien, die vorgeben, dass der/die Gesprächspartner:in gewisse Vorteile habe. Allerdings wird beispielsweise eine Bitte immer noch als Bitte ausgedrückt, nur gewissermaßen *aufgehübscht* bzw. indirekter geäußert (vgl. *Gib mir das Salz!* vs. *Könntest du mir bitte das Salz reichen?*).

Zu den Klassikern in der Höflichkeitsforschung zählen Brown/Levinson (1978/1987) und Leech (1983). Beide sind sprechakttheoretisch orientiert (vgl. Austin 1962, Searle 1969). Höflichkeit wird als Strategie zur Konfliktvermeidung oder -minimierung verstanden, indem die Kommunikationsbeteiligten, etwa durch indirekte Formulierungen, jegliche Form des Konflikts zu meiden versuchen (vgl. auch Watts 2005: xiii; Zarend 2015).

Beiden Ansätzen liegen die Ausführungen zum Face-Begriff von Goffman (1967) sowie das Kooperationsprinzip von Grice (1975) zugrunde. Nach Goffman (1967) ist das Face ein positives Selbstbild, ein positiver sozialer Wert, das/der in sozialen Interaktionen vermittelt und beansprucht wird: „The term face may be defined as the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact.“ (Goffman 1967: 5) Mit ‚line‘ sind dabei verbale und nonverbale Verhaltensweisen in sozialer Interaktion gemeint, mit denen Gesprächsbeteiligte ihre Sicht auf die Situation sowie die Einstellung gegenüber den anderen Teilnehmer:innen vermitteln. Diese sind für die Darstellung des eigenen Selbstbildes wie auch für dessen Wahrnehmung durch andere Beteiligte entscheidend.

Eine weitere Begrifflichkeit nach Goffman (1967), die von Brown & Levinson (1978/1987) und Leech (1983) in ihre Höflichkeitsparadigmen mit aufgenommen wurde, ist die Imagepflege oder die Beziehungsarbeit (*framework*). Aufbau, Erhalt und Pflege des Selbstbildes spielen in jeder Kommunikation eine wichtige Rolle: Mitglieder einer Gesellschaft sind darauf bedacht, sowohl das eigene als auch das Face der anderen Beteiligten zu wahren (Goffman 1967: 11).

Daneben kommt auch das Kooperationsprinzip (*cooperative principle*) nach Grice (1975) in Hinblick auf Höflichkeit zum Tragen. Dieses Prinzip liegt der Kommunikation im Normalfall zugrunde und wird meist unbewusst befolgt, mit dem Ergebnis, dass die Kommunikation aufrechterhalten wird: „Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange.“ (Grice 1975: 45–46) Das Kooperationsprinzip konkretisiert sich in 4 Maximen:

- die Maxime der Quantität (*Mach deinen Beitrag so informativ, wie es der aktuelle Gesprächszweck erfordert*);

- b) die Maxime der Qualität (*Versuche, einen wahren Beitrag zu geben*);
- c) die Maxime der Relevanz (*Sei relevant*);
- d) die Maxime der Modalität (*Sei klar; vermeide unklare Ausdrucksweisen*).

Im Normalfall werden diese Maximen eingehalten; interessiert hat die Pragmatik allerdings vor allem, was passiert, wenn sie nicht eingehalten werden. Der Kern von Grices Theorie sind die konversationellen Implikaturen, bei denen einer oder auch mehreren Maximen ostentativ – d. h. für den/die Empfänger:in gut erkennbar – zuwidergehandelt wird, mit dem Zweck, einen nicht expliziten Zweitsinn durch die Äußerung zu transportieren. Bei einer ironischen Äußerung wird z. B. das genaue Gegenteil des Gesagten impliziert.

Die pragmatische Höflichkeitstheorie setzt nun dort an, wo Maximen nicht eingehalten werden, ohne dass damit eine Zweitbedeutung impliziert werden soll. Hier wurde Grices Theorie als ungenügend empfunden, weil sie Phänomene wie glatte Lügen bei Komplimenten oder auch Verschweigen von relevanten, aber unangenehmen Fakten nicht erklären kann. Es ist kein Zufall, dass eine der beiden – ungefähr gleichzeitig entstandenen – grundlegenden Höflichkeitstheorien, jene von Leech (1983), genau die Struktur von Grices Theorie nachahmt, mit einem das Kooperationsprinzip ergänzenden fundamentalen Prinzip, dem Höflichkeitsprinzip (*Be polite!*), und untergeordneten, konkretisierenden Maximen.<sup>39</sup> In einem ersten Schritt soll allerdings ein spezifischerer Blick auf die andere maßgebliche Höflichkeitstheorie, nämlich jene von Brown & Levinson (1978/1987), geworfen werden.

### **2.1.7.1 Das Höflichkeitsparadigma von Brown & Levinson**

Die Theorie zu Höflichkeit von Brown & Levinson (1978/1987) stützt sich ganz wesentlich auf Goffmans Face-Theorie, erweitert diesen Ansatz jedoch um eine zweite Art von Face: Dem positiven Face, dem nach außen projizierten Selbstbild (s. o.) steht das negative Face, die Handlungsfreiheit und Selbstbestimmung, gegenüber. Beide Faces können durch facebedrohende Akte (FTAs) angegriffen werden, was Face-kompensatorischen Handlungen den Anlass gibt.

Das Modell von Brown & Levinson (1978/1987) macht das Phänomen Höflichkeit (*politeness*) als universale Strategie kommunikativer Konfliktvermeidung an sprachlichen Mitteln (dreier verschiedener Kultursprachen) fest. Höflichkeit wird dabei auf strategisches, zweckrationales Handeln von Modellpersonen (MP) mit folgenden Eigenschaften zurückgeführt (Brown/Levinson 1978/1987: 59–60):

---

<sup>39</sup> Das Höflichkeitsprinzip und die Höflichkeitsmaximen werden weiter unten genauer ausgeführt.

- a) Alle MP handeln rational. Rationalität bedeutet die Möglichkeit, Mittel daraufhin abzuschätzen, ob sie dazu dienen, möglichst ökonomisch ein Ziel zu erreichen.
- b) Alle MP versuchen das (positive) Selbstbild, das sie von sich haben, vor anderen aufrechtzuerhalten. Diesem Selbstbild liegen zwei Grundbedürfnisse (*face wants*) zugrunde: das positive Face, d. h. der Wunsch, von anderen geschätzt und anerkannt zu werden, sowie das negative Face, d. h. das Bedürfnis nach Handlungsfreiheit.
- c) Es liegt im wesentlichen Interesse aller MP, gegenseitig ihr Face zu wahren.
- d) Manche Handlungen bedrohen intrinsisch das Face (*face threatening act, FTA*).
- e) Der/die Sprecher:in versucht, die Gesichtsbedrohung durch den FTA zu minimieren, außer der FTA soll offen und unverblümt geschehen.
- f) Je facebedrohender der FTA, desto indirekter und höflicher wird er durchgeführt.
- g) Es werden nicht mehr Höflichkeitsstrategien als nötig eingesetzt.

In der Kommunikation streben die Beteiligten also danach, die menschlichen Grundbedürfnisse im Bereich des Face zu erfüllen, d. h. das positive und negative Face zu wahren. Um eine konfliktfreie Kommunikation zu gewährleisten, gilt es, Bedrohungen für das Face abzuwenden oder möglichst gering zu halten. Es können allerdings verschiedene gesichtsbedrohende Handlungen auftreten, die das negative und das positive Face des/der Hörer:in und/oder des/der Sprecher:in inhärent bedrohen können, sogenannte FTAs.

Das positive Face des/der Sprechenden kann etwa durch das Bitten um Verzeihung bedroht werden (*Kannst du mir bitte verzeihen?*), wodurch das eigene Fehlverhalten eingestanden wird. Das positive Face des/der Hörenden kann dagegen z. B. durch Lächerlichmachen angegriffen werden (*Wie siehst du denn aus mit deiner neuen Frisur?*); in diesem Fall wird eine gewisse Respektlosigkeit gegenüber dem/der Hörer:in ausgedrückt. Daneben kann das negative Face des/der Sprecher:in z. B. durch Ausdruck von Dank (*Ich bin dir sehr dankbar, dass du mir in dieser schwierigen Situation geholfen hast.*) in Bedrängnis geraten, da er:sie sich dadurch bescheiden machen und sich selbst auferlegen kann, etwas an den/ die Hörer:in zurückgeben zu müssen. Nicht zuletzt kann das negative Face des/der Hörenden etwa durch einen Befehl bedroht werden (*Mach das Fenster zu!*): Der/ die Sprechende schränkt durch solche Äußerungen die Selbstbestimmung und Handlungsfreiheit des/der Hörer:in ein.

Mögliche weitere Bedrohungen für das positive Face des/der Sprecher:in sind z. B. auf ein Kompliment reagieren (er:sie muss dafür eventuell das eigene Face beschädigen), Verlust der physischen Kontrolle über den Körper, Selbsternidrigung, dummes Verhalten, Bekenntnisse, Zugeben von Schuld oder Verantwortung,

Durchsickern von Gefühlen, Nichtkontrolle von Weinen oder Lachen. Mögliche Bedrohungen des positiven Face des/der Hörer:in sind z. B. direkte Angriffe des positiven Face (Missbilligung, Kritik, Tadel, Verachtung, Anschuldigungen, Widerspruch, Nichtübereinstimmung). Diese implizieren, dass sich der/die Sprecher:in möglicherweise nicht um das Selbstbild des Gegenübers kümmert. Eine weitere Bedrohung kann durch die Gefährdung der Annahme entstehen, dass Hörer:in und Sprecher:in derselben Gruppe mit denselben Idealen und Zielen angehören (Ausdruck unkontrollierter Gefühle, Respektlosigkeit, Tabu-Themen, Unpassendes, schlechte Nachrichten über den/die Hörer:in oder prahlerisch gute Nachrichten über den/die Sprecher:in, Ansprechen heikler Themen, offensichtliches Nichtkooperieren, falscher Gebrauch von Anredeformen).

Die Bedrohungen für das negative Face des/der Sprechenden können ergänzt werden durch das Akzeptieren einer Bitte um Verzeihung, Entschuldigungen, das Akzeptieren von Angeboten (erzeugt Schuldgefühl, kann den/die Sprecher:in einschränken) sowie unwillige Versprechen und Angebote. Das negative Face des/der Hörenden kann außerdem bedroht werden durch a) FTAs, die zukünftige Handlungen des/der Hörer:in präzisieren (Befehle, Bitten, Vorschläge, Ratschläge, Erinnerungen, Drohungen, Warnungen); b) FTAs, die zukünftige Handlungen des/der Sprecher:in gegenüber des/der Hörer:in präzisieren, davon ist auch die Freiheit des/der Hörer:in betroffen (Angebote, Versprechen); c) FTAs, die Wünsche des/der Sprechenden gegenüber Gütern des/der Hörenden präzisieren (Komplimente; Ausdrücke der Bewunderung, der/die Sprecher:in wird vielleicht dem/der Hörer:in etwas wegnehmen; Ausdrücke von starken negativen Gefühlen gegenüber dem/der Hörer:in, der/die Sprecher:in wird vielleicht die Güter des/der Hörer:in beschädigen).

Außerdem beschreiben Brown & Levinson (1978/1987) in ihrem Höflichkeitsmodell fünf Strategien der Gesichtswahrung bzw. der Abmilderung eines FTA.<sup>40</sup> Beteiligte an einer sozialen Interaktion haben bei der Ausführung eines FTA (zumindest) drei einander teilweise widerstrebende Bedürfnisse:

- 1) den Inhalt des FTA zu vermitteln;
- 2) effizient oder rasch zu handeln;
- 3) die Face-Bedrohung zu minimieren.

Wenn 2) nicht größer ist als 3) (z. B. wenn keine Gefahr im Verzug ist), wird die Face-Bedrohung minimiert, an diesem Punkt kommt Höflichkeit ins Spiel.

---

**40** Der Begriff der *Strategie* wird von Brown & Levinson (1978/1987: 92) in drei verschiedenen Bedeutungen verwendet: 1) als übergeordneter (handlungsleitender) Wunsch des/der Sprecher:in, 2) als von diesem abgeleitete, spezifischere Wünsche, 3) als der Einsatz sprachlicher Mittel, die diese Wünsche erfüllen. Eine Strategie verbindet demnach die nicht-sprachliche Ebene (Wünsche, Zwecke) mit der sprachlichen Ebene (konkrete Äußerung) (vgl. auch Thaler 2012).

In Abbildung 4 wird gezeigt, welche Strategien je nach Schwere des FTA und damit je nach Größe der Gefahr, sein Gesicht zu verlieren, angewendet werden:

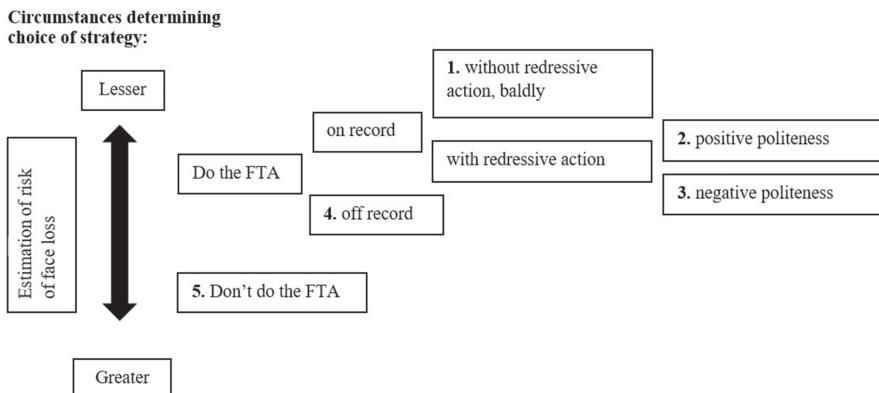


Abb. 4: Strategien der Gesichtswahrung (Brown/Levinson 1978/1987: 69).

Unter 1.) *Bald-on-record* fallen explizite FTAs, die direkt geäußert werden, ohne Versuch der Gesichtswahrung. Diese sind typisch, wenn Effizienz und Dringlichkeit von hoher Bedeutung sind bzw. wenn Gefahr im Verzug ist (*Hilfe!*); bei Einladungen, die den/die Adressierte:n auffordern, gegen mögliche (Gesichts-)Interessen des/der Sprechenden zu handeln (*Setzen Sie sich!*); bei aufgabenorientierten Diskursen (*Geben Sie 3 Löffel Zucker hinzu.*); oder auch wenn Dringlichkeit suggeriert wird (*Vergeben Sie mir!*). Brown & Levinson (1978/1987: 94–95) bezeichnen mit *bald-on-record* solche Äußerungen, für die die Gesichtswahrung sekundär ist und die für eine maximal effiziente Informationsvermittlung eingesetzt werden.

Unter 5.) Verzichte auf den FTA fallen solche Handlungen, die so heikel sind, dass sie ganz unterlassen werden. Von den dazwischenliegenden Strategien sind 2.) Positive Höflichkeit und 3.) Negative Höflichkeit als eigentliche Höflichkeit anzusehen, 4.) *Off record* ist eine Form der verhüllenden und nur mit Andeutungen arbeitenden Sprechweise, die nicht zu den Höflichkeitsstrategien im engeren Sinn zählt. Zu den impliziten, verdeckten (*off record*-)Strategien zählen etwa Andeutungen, Tautologien (z. B. *Wenn es kalt ist, ist es kalt.*), Metaphern, Ironie, rhetorische Fragen, mehrdeutige und vage Aussagen.<sup>41</sup>

---

41 Beispiel: Ein Gast findet es heiß und stickig und möchte, dass das Fenster trotz des stürmischen Wetters geöffnet wird. Er:sie vermutet allerdings, dass das seine/ihre Gastgeber:innen als Zumutung empfinden könnten, und sagt deshalb vorsichtig: *Sie mögen es gerne sehr warm, nicht wahr?*

Ausgleichende Handlungen (*redressive actions*) sollen zeigen, dass eine Face-Verletzung nicht intendiert ist. Je nachdem, ob sie eher dem positiven Face des/der Hörer:in *schmeicheln* oder dem negativen Face Rechnung tragen, unterscheiden Brown & Levinson (1978/1987) negative Höflichkeit (*negative politeness*) und positive Höflichkeit (*positive politeness*):

- a) Negative Höflichkeit, durch die der/die Sprecher:in den Effekt des FTA auf das negative Face des/der Hörer:in vermindert. Dazu gehören konventionelle Indirektheit (*Können Sie mir das Salz reichen?*), Hedges und Vagheit (*Das neue Buch von XY soll gut sein.*), Situationen, in denen der/die Adressierte zu keinen Handlungen gezwungen wird (*Könntest du mir vielleicht dein Auto leihen?*) sowie Entschuldigungen (*Ich möchte Sie nicht stören, aber könnten Sie hier weggehen?*).
- b) Positive Höflichkeit, durch die der/die Sprecher:in sichtbar macht, dass es ihm/ihr ein Anliegen ist, das positive Face des/der Hörer:in zu wahren: Verweis auf geteiltes Wissen (*common ground*, Verwendung von In-group-Markierungen, Anreden, Small Talk, Vermeidung von Differenzen, Präsupponieren von Gemeinsamkeiten). Der/die Sprecher:in suggeriert, dass er:sie und der/die Adressierte kooperieren (der/die Sprecher:in zeigt an, die Wünsche des/der Hörer:in zu kennen, der/die Sprecher:in macht Angebote, Versprechen, erfüllt Wünsche des/der Adressierten).

Es kann auch zu einer Kombination von positiver und negativer Höflichkeit kommen, etwa in Äußerungen wie: *Du bist so toll! Du kannst das so gut! Bitte mach das für mich!*. Im ersten Teil der Äußerung (*Du bist so toll! Du kannst das so gut!*) bringt der/die Sprecher:in seinem/ihrem Gegenüber Anerkennung und Wertschätzung entgegen und formuliert davon ausgehend im zweiten Teil (*Bitte mach das für mich!*) eine höfliche Bitte.

In Bezug auf Höflichkeit spielen außerdem drei Einflussfaktoren als Messwerte für die Gewichtigkeit (*amount of politeness*) eine Rolle: das jeweilige Machtverhältnis der Interagierenden ( $P = power slope$ ), die soziale Distanz zwischen ihnen ( $D = social distance$ ) sowie das soziale Gewicht der verbalisierten Handlung ( $R = rank$  oder auch *size of imposition*) (vgl. Brown/Levinson 1978/1987; Leech 1980; hier adaptiert nach Thomas 1995: 124 angeführt). Dabei gilt in Hinblick auf das Machtverhältnis: Je größer das Machtgefälle zwischen den Gesprächsbeteiligten ist, desto mehr Höflichkeit und Indirektheit werden eingesetzt. Mit der sozialen Distanz geht einher, dass, je vertrauter die Gesprächspartner:innen sind, desto direkter können ihre Äußerungen sein (Bitten können z.B. direkter geäußert

---

Velleicht verstehen die Gastgeber:innen die Andeutung sofort, vielleicht erkennt der Gast auch aus deren Reaktionen, ob er:sie zu Strategie 3 übergehen kann.

werden). Im Zusammenhang mit der *size of imposition* steht die Frage, wie stark gesichtsbedrohend z. B. eine Bitte oder Aufforderung ist bzw. wie sehr der Handlungsspielraum eingegrenzt wird.<sup>42</sup>

Das Gewicht eines FTA kann ausgehend von diesen drei Faktoren nach folgender Formel bestimmt werden:

$$W_x = D(S,H) + P(H,S) + R_x$$

$W_x$ , das Maß für das Gewicht eines FTA, ergibt sich aus der Summe von:

- a)  $D$  = Maß für die Distanz;
- b)  $P$  = Maß für die Power/Macht, die der/die Hörer:in über den/die Sprecher:in besitzt;<sup>43</sup>
- c)  $R$  = Maß für die Stärke der Bedrohung durch den FTA.

Das heißt, ein FTA ist umso gewichtiger, je weniger vertraut die Kommunikationsbeteiligten sind, je größer das Machtgefälle ist und je bedrohlicher eine bestimmte Handlung an sich ist. So stellt die Aussage *Mach mal das Fenster zu!* eher einen FTA dar, wenn ein/e Student:in diese gegenüber einem/einer Professor:in tätigt, als wenn eine Mutter diesen Befehl an ihr Kind richtet. Besteht hingegen ein kollegiales Verhältnis zwischen Professor:in und Student:in und befinden sich beide in einem situativ ungezwungenen Kontext außerhalb des Seminarraums, erhält auch diese Aussage trotz hierarchischer Strukturen eine andere Gewichtung.

Alle drei Faktoren sind in dem Höflichkeitsparadigma von Brown & Levinson (1978/1987: 74, 76) nicht als deren tatsächliche Ausprägungen zu interpretieren, sondern sie sind davon abhängig, wie sie von den Gesprächsbeteiligten empfunden und vorausgesetzt werden, und stehen damit in Verbindung mit deren Einstellungen bzw. Einschätzungen der Situation: „these [factors] are not intended as sociologists' ratings of actual power, distance, etc., but only as actors' assumptions of such ratings, assumed to be mutually assumed, at least within certain limits.“ (vgl. auch Thaler 2012 sowie die Ausführungen weiter oben)

---

<sup>42</sup> Vgl. auch den Unterschied zwischen *free* und *non-free goods* bei Goffman (1967). *Free* bedeutet minimale Indirektheit, z. B. jemanden fragen, wie spät es ist; *non-free* steht in Verbindung mit einem höheren Maß an Höflichkeit, z. B. jemanden darum bitten, eine größere Geldsumme zu leihen.

<sup>43</sup>  $P$  kann sowohl rollen- als auch situationsabhängig sein.

### 2.1.7.2 Das Höflichkeitsparadigma von Leech

Die zweite die Forschungsrichtung begründende Höflichkeitstheorie stammt von Leech (1983). Nach Leech (1983) impliziert Höflichkeit fundamental, dass es eine Unterscheidung zwischen *ego* und *alter* gibt. *Ego* und *alter* werden unterschiedlich behandelt, nicht wie bei Brown & Levinson (1978/1987), wo alle vier Faces gleichberechtigt sind und nichts darüber ausgesagt wird, welches wichtiger ist. Leech (1983) hält fest, dass in der Haltung, die in einer Kommunikationssituation zur Schau getragen wird, die beiden Faces des/der angesprochenen Hörer:in vorrangiger geschont werden müssen als die des/der Sprecher:in.

In der Höflichkeitstheorie von Leech (1983: 80) wird *politeness* als Erklärungsinstrument beschrieben, „why people are often so indirect in conveying what they mean“. Leech (1983) geht in seinem Paradigma von einem Höflichkeitsprinzip (*politeness principle*) aus:

Minimize (all the things being equal) the expression of impolite beliefs; maximize (all the things being equal) the expression of polite beliefs. (Leech 1983: 79)

Leech (1983) stellt dieses Höflichkeitsprinzip dem Kooperationsprinzip von Grice (1975) zur Seite. Dieses erklärt das Kommunikationsverhalten in Gesprächssituativen, in denen sich Beteiligte nicht an das den Maximen von Grice zugrunde gelegte Kooperationsprinzip halten – aber ohne eine Implikatur erzeugen zu wollen. Betritt z. B. ein Gast das Haus mit schmutzigen Schuhen und sagt zum/zur zusehenden Bewohner:in *Ich mache ganz schönen Dreck hier*, so verstößt er:sie gegen die Maxime der Relevanz, da beide Personen sehen, dass es dreckig ist. Die Äußerung wird allerdings entsprechend nicht als Information verstanden, sondern als höfliche Entschuldigung.

Das Höflichkeitsprinzip unterteilt Leech (1983: 132) in sechs verschiedene Höflichkeitsmaximen:<sup>44</sup>

a) Takt-Maxime:

- Minimiere die Kosten von anderen.
- Maximiere den Nutzen von anderen.
- Diese Maxime steht in Verbindung mit der Variabel *size of imposition*, es können etwa sog. *Minimizer* eingesetzt werden, um den Grad der Gesichtsbedrohung einer bestimmten Handlung abzuschwächen, z. B. *Warte mal eine Sekunde!*.

b) Anerkennungs-Maxime:

- Minimiere die Missbilligung anderer.
- Maximiere das Lob anderer (z. B. *Was für ein gutes Essen du gekocht hast!*).

---

<sup>44</sup> Diese Maximen können allerdings von Kultur zu Kultur unterschiedlich sein.

- c) Großzügigkeits-Maxime:
  - Minimiere den Nutzen für dich selbst.
  - Maximiere die Kosten für dich selbst.
  - Das heißt, den Bedürfnissen des/der Sprecher:in wird im Gegensatz zu jenen des Gegenübers ein geringerer Stellenwert beigemessen, z. B. *Ich kann dir mein Auto leihen*.
- d) Bescheidenheits-Maxime:
  - Minimiere Selbstlob.
  - Maximiere Selbstkritik (z. B. *Wie dumm von mir!*).
- e) Übereinstimmungs-Maxime:
  - Minimiere Nichtübereinstimmung zwischen dir und anderen.
  - Maximiere Übereinstimmung.
- f) Sympathie-Maxime:
  - Minimiere Antipathie zwischen dir und anderen.
  - Maximiere Sympathie.

Leechs Höflichkeitsverständnis basiert auf einem Kalkül der Kosten (*cost*) und des Nutzens (*benefit*) innerhalb einer Interaktion für die Beteiligten. Es gibt viele Situationen, in denen der/die Sprecher:in sein/ihr eigenes Face zurücknehmen muss, um die Kosten für das Gegenüber so gering wie möglich zu halten und gleichzeitig dessen Nutzen zu erhöhen. Es passiert demnach eine Zielrichtung auf das *alter*, zugleich wird das *ego „heruntergemacht“*. Demnach wäre eine Äußerung wie *Ich kann dir mein Auto leihen* im Sinne Leechs als höflich aufzufassen, denn sie erhöht den Nutzen des Gegenübers und schränkt gleichzeitig den/die Sprecher:in ein.

### 2.1.7.3 Die Weiterentwicklung von Modellen zu (Un-)Höflichkeit

Nach den grundlegenden Höflichkeitsmodellen von Brown & Levinson (1978/1987) und Leech (1983) ist die Höflichkeitsforschung nicht stehen geblieben, sondern es haben sich weitere Ansätze entwickelt, die hier kurz wiedergegeben werden. 2001 entstand eine neue Schule, die durch Eelen (2001) in Gang gesetzt wurde. Die sogenannte postmoderne oder diskursive Annäherung an die Höflichkeit, z. B. bei Mills (2003), Watts (2003, 2006), Culpeper (2005, 2008, 2011), Kerbrat-Orecchioni (2007), Locher & Watts (2008) sowie Held (2009, 2010), führte die Unterscheidung zwischen pol1 (*first-order politeness*) und pol2 (*second-order politeness*) ein. Der Höflichkeitsbegriff erster Ordnung umfasst die idealisierten Verhaltensformen, die in einer Sprachgemeinschaft als höflich bzw. unhöflich wahrgenommen und bezeichnet werden. Es handelt sich hierbei also um den alltagssprachlichen Gebrauch der Begriffe *Höflichkeit* und *höflich* (Thaler 2012: 15). Demgegenüber steht ein Begriffsverständnis zweiter Ordnung, in dem Höflichkeit als theoreti-

sches Konstrukt aufgefasst wird. Es geht darum, grundlegende Fragen danach zu klären, wie sich Höflichkeit in der Interaktion manifestiert und welche Faktoren die Strategiewahl des/der Sprecher:in beeinflussen, wenn er:sie etwas höflich vermitteln möchte (Fraser 2001: 1424). In *first-order politeness* stehen die Einschätzungen des/der Sprecher:in und/oder des/der Hörer:in im Zentrum, ob ein bestimmtes sprachliches Verhalten höflich oder unhöflich ist; *second-order politeness* beschreibt die Forschungsperspektive, aus der heraus dieses Verhalten erklärt wird (Thaler 2012: 16).

Watts (2003: 144) entwickelte ein Höflichkeitsmodell, in dem er die Auffassung vertritt, dass unmarkiertes (höfliches) Verhalten als „political behaviour“ beschrieben werden sollte, während der Begriff Höflichkeit („politeness“) für Fälle reserviert werden sollte, in denen Sprecher:innen mehr Höflichkeit einsetzen, als es die Situation erfordert: „Political behaviour is that behaviour, linguistic and non-linguistic which the participants construct as being appropriate to the ongoing social interaction.“ (Watts 2003: 257). Höflichkeit wird dahingegen als „excess of political behaviour“ (Watts 2003: 259) definiert; dieser Überschuss kann als höflich interpretiert werden, er kann aber auch ins Gegenteil umschlagen.

Culpeper (1996, 2005, 2008) weitete das Forschungsinteresse auf Unhöflichkeitsphänomene (*impoliteness, rudeness*) aus.<sup>45</sup> Mit seiner klassischen Arbeit zur Unhöflichkeit („Towards an anatomy of impoliteness“, 1996), setzte sich der Autor zum Ziel, unhöfliches sprachliches Verhaltens in Anlehnung an den Face-Begriff von Brown & Levinson zu klassifizieren. Darüber hinaus untersuchte Culpeper die soziale Dimension von Unhöflichkeit sowie den Zusammenhang zwischen Unhöflichkeit und Macht. In seinen späteren Arbeiten (2005, 2008, 2011) ergänzte er seine Konzeption durch die Erkenntnis, dass sich (Un-)Höflichkeit nicht allein an sozialen und linguistischen Normen ablesen lässt, die für alle Sprecher:innen gleichermaßen und für jeglichen Kontext gleich gelten: „what might be evaluated negatively for some might not be by others, or even might be considered positive.“ (Culpeper 2011: 25) Diese Auffassung vertritt auch Bonacchi (2013) in ihrer Veröffentlichung zu (Un-)Höflichkeit im Deutschen, Italienischen und Polnischen. Nach Bonacchi (2013) lassen sich Höflichkeit und Unhöflichkeit nicht per se feststellen oder analysieren, sondern beide Phänomene sind interaktionsabhängig. Dabei spielen die in der jeweiligen Gesellschaft etablierten sozialen Normen, die Realisierung sozialer Zwecke sowie geltende Werteschemata eine Rolle (Bonacchi 2013: 83–84). Die Autorin (2013: 84) unterscheidet drei Typen von Verhalten:

---

<sup>45</sup> Daneben beschäftigten sich auch Watts (2003, 2006) und Kerbrat-Orecchioni (2010) mit Unhöflichkeit.

- a) Verhalten, das in Bezug auf die sozialen Normen als positiv markiert (höflich), nicht markiert (nicht höflich) oder negativ markiert (überhöflich, unhöflich) ist;
- b) Verhalten, das in Bezug auf die Erhaltung der sozialen Zwecke adäquat oder inadäquat ist;
- c) Verhalten, das in Bezug auf die geltenden Werteschemata als „höflich“, „nicht höflich“ und „unhöflich“ gilt.

Die Fragestellungen und Erkenntnisse der Höflichkeitsforschung haben inzwischen auch Eingang in die konversations- und gesprächsanalytische Forschung gefunden. In der ursprünglichen Konversationsanalyse des ganz klassischen Zugangs (vgl. Sacks/Schegloff/Jefferson 1974) ist Höflichkeit keine relevante Kategorie, u. a. weil die Konversationsanalyse bereits vor den Höflichkeitstheorien entstand. Allerdings gibt es Forscher:innen, die sich konversations- und gesprächsanalytisch mit dem Thema Höflichkeit auseinandersetzen und damit Höflichkeit als relevante Fragestellung in die Konversationsanalyse/Gesprächsanalyse einführen.

#### **2.1.7.4 Höflichkeit in der Gesprächsforschung**

In der Gesprächsforschung wird Höflichkeit als theoretischer Beschreibungsrahmen für kommunikative Verhaltensweisen aufgefasst, wobei immer das Prinzip der Interaktion gilt. Durch die Ausrichtung auf den interaktiven und sequenziellen Verlauf einer Gesprächssituation ergänzt die Gesprächsforschung die auf Sprechakte ausgelegten ursprünglichen Höflichkeitsmodelle, wo die Sprecherperspektive stark in den Vordergrund und die Hörerperspektive in den Hintergrund gestellt wird. Außerdem erweitern gesprächsanalytische Arbeiten das Phänomen Höflichkeit um natürliche, authentische Daten; ursprünglich wurden Höflichkeitsstrategien nämlich häufig in isolierten Mini-Kontexten untersucht, oft auch mit erfundenen Beispielen.<sup>46</sup> Interessanterweise stellen bereits Brown & Levinson (1978/1987: 10) fest, dass „some strategies for FTA-handling are describable only in terms of sequences of acts or utterances, strung together as outputs of hierarchical plans“. Damit nähern sie sich an die Betrachtung der lokalen *turn-by-turn*-Organisation von Gesprächen (Sacks/Schegloff/Jefferson 1974) an, die einen Turn als das Ergebnis kollektiver Arbeit der Beteiligten postuliert.

*Politeness* als Begriff ist im konversationsanalytischen Vokabular nicht verankert, auch wenn die Konversationsanalyse/Gesprächsforschung untersucht, wie Interaktionen gestaltet werden, um eine soziale Ordnung herzustellen. Face und

---

<sup>46</sup> Häufig wurde auch gefragt: *Was würdest du in dieser Situation sagen?* Demgegenüber stellen Methode und Ansatz der Konversationsanalyse/Gesprächsforschung einen großen Schritt in Richtung Empirie dar.

Übereinstimmung/Nicht-Übereinstimmung (*alignment/disalignment*) zwischen den Gesprächsbeteiligten werden im sequenziellen Kontext der Interaktion und unter Bezugnahme der Aktivitäten aller Teilnehmer:innen betrachtet (vgl. Hayashi 1996). Mit Übereinstimmung (*alignment*) einher gehen sog. Präferenzstrukturen (vgl. Sacks/Schegloff 1979, Sacks 1992). Für bestimmte Paarsequenzen kommen als *second pair part* zwei (oder auch mehrere) alternative Optionen in Frage, z. B. kann eine Einladung entweder angenommen (präferiert) oder abgelehnt (dispräferiert) werden. Präferenz ist dabei eine Frage der sequenziellen Strukturierung eines Gesprächs und nicht der individuellen Entscheidung der Gesprächsteilnehmer:innen (Kasper 2009: 27). Die präferierte Form kann daran erkannt werden, dass sie unmarkiert ist, die dispräferierte Form ist dagegen markiert, d. h., sie enthält Verzögerungselemente, Pausen, Entschuldigungen, Vorankündigungen, Erklärungen usw. Brown & Levinson (1978/1987) halten fest, dass Präferenz und Dis-Präferenz in vielen Fällen mit *face saving* zusammenfällt bzw. vor dem Hintergrund der Gesichtswahrung determiniert und motiviert ist. Interaktant:innen zeigen zwar generell eine *preference for agreement and contiguity* (Sacks 1973), allerdings ist beispielsweise bei Komplimenten der präferierte zweite Part eine Zurückweisung oder Minimierung (vgl. Gürlich/Mondada 2008: 53).

Sollen Höflichkeit und Konversationsanalyse miteinander in Einklang gebracht werden, so muss das Höflichkeitsmodell in dem Sinne aufgebrochen werden, dass Höflichkeitsstrategien nicht in isolierten Kontexten und nicht nur aus rein sprechakttheoretischer Perspektive betrachtet werden: „the speaker's redressive strategies cannot be explained in terms of individual, or some combination of individual, speech actions alone.“ (Hayashi 1996: 252)

### 2.1.7.5 Höflichkeit in der Orchesterprobe

Höflichkeit ist auch in der Orchesterprobe ein relevantes Thema. Der Instruktionskontext der Orchesterprobe birgt in sich bereits automatisch Face-Bedrohungen: Dirigent:innen sind ständig mit dem Dilemma konfrontiert, ob und wie sie gesichtsbedrohende Handlungen ausführen sollen oder ob sie diese unterlassen sollen. Dabei ist vor allem das negative Face der Musiker:innen bedroht, da der/die Dirigent:in durch seine/ihre Instruktionen die Handlungsfreiheit der Musiker:innen einschränkt. Zugleich kann auch das positive Face der Musiker:innen zum Angriffspunkt werden, wenn der/die Dirigent:in Kritik am Gespielten übt. Daneben können FTAs auftreten, die das positive Face des/der Dirigent:in bedrängen, etwa wenn der/die Dirigent:in entschuldigende Ausdrücke verwendet (z. B. *sorry* bei Unterbrechungen) oder auch wenn die Musiker:innen spielen, sobald der/die Dirigent:in spricht; in diesem Fall sind Autorität und Professionalität des/der Dirigent:in bedroht. Das negative Face des/der Dirigent:in bleibt in den meisten Fällen unberührt.

Gleichzeitig ist für Höflichkeit das Machtgefälle (vgl. die Faktoren *social distance* und *power*) zwischen Dirigent:in und Musiker:innen in Orchesterproben ausschlaggebend (der/die Dirigent:in steht hierarchisch über den Musiker:innen). Die beiden Faktoren *social distance* und *power* prägen sich in der Orchesterprobe folgendermaßen aus:

- a) Der/die Dirigent:in ist alleinige/r Sprecher:in (in den allermeisten Fällen);
- b) der/die Dirigent:in bestimmt den Ablauf der Probe;
- c) der/die Dirigent:in hat die Partitur vor sich liegen, in der alle Musikstimmen gesammelt sind (die Musiker:innen haben nur ihre jeweilige Stimme auf ihren Notenpulten aufliegen);
- d) der/die Dirigent:in kann die Musik zu jeder Zeit unterbrechen;
- e) der/die Dirigent:in instruiert die Musiker:innen;
- f) der/die Dirigent:in bestimmt, was und wie gespielt wird und auch wer spielt, allerdings immer mit Orientierung an der Partitur.

Der Faktor *rank* ist aufgrund der sozialen Distanz und des Machtgefälles zwischen Dirigent:in und Musiker:innen nicht unbedingt ausschlaggebend. Da es in der Orchesterprobe ein Muster gibt, das die Situation definiert bzw. ein institutioneller Kontext gegeben ist, ist das Gewicht von FTAs nicht so groß wie etwa in einer alltäglichen Gesprächssituation. Daneben ist der Faktor *social distance* davon abhängig, wie gut sich der/die Dirigent:in und die Musiker:innen kennen. Deshalb unterliegt dieser Faktor auch einer gewissen Variation/Veränderung, je nachdem, wie vertraut die Beteiligten an einer Orchesterprobe miteinander sind.

Ausgehend von diesen Beobachtungen kann die Face-Bedrohung in Orchesterproben wie folgt dargestellt werden (vgl. Tab. 2):

**Tab. 2:** Face-Bedrohung in Orchesterproben.

	Dirigent:in	Musiker:innen
<b>Positives Face</b>	manchmal gefährdet	gelegentlich gefährdet
<b>Negatives Face</b>	fast nie gefährdet	ständig stark gefährdet

Wie mit Face-Bedrohungen in Orchesterproben effektiv umgegangen wird, zeigt folgendes Beispiel aus einer Probe des *Orchestre de l'Opéra de Rouen* mit dem Dirigenten Antony Hermus:

---

01	MM	♪((spielen))♪
02	D	&%a:h!% & c'est très bon! c'est %très bon! c'est très-% très bon!
	d	& hört auf zu dirigieren&
	mm	%♪((spielen))♪% %♪((spielen))♪----->%

03 D &<sup>o</sup>h:: ta::m ta:: ta:&  
 &blickt auf Partitur&

04 &<sup>o</sup>h: c'est possible (.) eh- de avoir un PEU plus de <> TE:MPS>,  
 &blickt zum Orchester ---->

05 (0.3) dans le NOI:R.&  
 ----->&

---

In Z02 weist der Dirigent gestisch durch die Beendigung seines Dirigats eine Unterbrechung der Musik an. Parallel dazu verwendet er verbal die Interjektion *ah* als Ausruf der (bewundernden) Überraschung sowie der Gutheißen. Diese Interjektion evaluiert gemeinsam mit den Ausdrücken *c'est très bon* und zeigt eine Ratifizierung seitens des Dirigenten an. Damit übt der Dirigent *positive politeness*: Er lobt die Musiker:innen und bemüht sich, eine gemeinsame, emphatische Basis aufzubauen. Gleichzeitig kompensiert er dadurch den gesichtsbedrohenden Akt des Unterbrechens, der eine Attacke gegen das negative Face der Musiker:innen darstellt. In Z03 lokalisiert der Dirigent durch sein Vorsingen eine bestimmte Stelle in der Partitur, zu der er in Z04 und Z05 eine Anweisung gibt. Diese Anweisung äußert er nicht direkt, z. B. in Form eines Imperativs, sondern er verpackt sie in eine indirekte, unaufdringlichere Form: *c'est possible de avoir*. Damit gibt er zu verstehen, dass er die Musiker:innen nicht in ihrer Handlungsfreiheit einschränken möchte – *negative politeness* kommt ins Spiel. Auf ähnliche Weise wirkt das minimierende Adverb *un peu* – ebenfalls in Z04 –, das in Verbindung mit der Takt-Maxime nach Leech (1983) steht: *Minimiere die Kosten von anderen*. Die Verwendung dieser abgeschwächten Formen, um eine bestimmte Stelle in der Partitur anzulegen, geht außerdem mit dem Faktor *size of imposition* einher. Je gesichtsbedrohender eine bestimmte Handlung ist (hier: das Anweisen/Korrigieren), desto indirekter wird sie ausgeführt. Diese Indirektheit bzw. Höflichkeit verschleiert gleichzeitig die tatsächlichen Machtstrukturen in der Orchesterprobe: Zur Schonung des positiven Face der Musiker:innen rückt der Dirigent hierarchische Gefüge in den Hintergrund.

Außerdem veranschaulicht das Beispiel eine Reihe an markiert realisierten Äußerungen, etwa das Lob an die Musiker:innen in Z02 oder die *hedgenden* Formen in Z04. Diese Äußerungen können auch als „face flattering acts“ (FFAs), als positives Pendant zu FTAs bezeichnet werden (Kerbrat-Orecchioni 2010: 37). Unter solche FFAs (oder auch *face enhancing acts*) werden Handlungen gefasst, die gut für das Face sind und dem Face schmeicheln, etwa lobende Ausdrücke, Komplimente oder Ratschläge (z. B. *Könntest du mir einen Ratschlag geben?*).

Im empirischen Teil dieser Studie soll beobachtet werden, wo in der Orchesterprobe Face-Bedrohungen auftauchen und welche logisch kalkulierten positiven (Anerkennung, Respekt, Ehrerweisung) und negativen (Abschwächungen, Aner-

kennung des Rechts auf Handlungsfreiheit der Adressierten, Distanzwahrung) Höflichkeitsstrategien und/oder FFAs zum Einsatz kommen.

## 2.2 Kommunikation in Orchesterproben als Gegenstand der Forschung

Die Orchesterprobe dient der Vorbereitung und Erarbeitung eines musikalischen Werks von Dirigent:in und Orchester. Sie bietet Raum für ‚Fehler‘ oder auch unvollkommene Versuche, für Wiederholungen von musikalischen Passagen und für die Besprechung von spezifischen Stellen in der Partitur. Für Außenstehende ist der Zugang zu diesem Proben und Probieren grundsätzlich versperrt. Erst in der Aufführung des erprobten Werks erhält die Öffentlichkeit bzw. das Publikum Zugang zu dem, was sich als Produkt aus dem Arbeitsprozess zwischen Dirigent:in und Musiker:innen herauskristallisiert hat. In diesem Sinne können das Orchester als Institution und die Orchesterprobe als zu erforschendes Feld als „closed access groups“ (Hornsby-Smith 1993: 53) aufgefasst werden, die nur für Autorisierte zugänglich sind (vgl. auch Schütte 1991: 21). Deshalb ist es nicht verwunderlich, dass bis dato wenig Literatur zu Kommunikations- und Interaktionsmustern in Orchesterproben existiert. Auch in Studien zu Probenstrategien (vgl. z. B. Price/Byo 2002; Biasutti 2012) wird die Kommunikation in der Orchesterprobe nur am Rand behandelt. Biasutti (2012) hält beispielsweise fest, dass

[conductors] tend to sing or play a particular passage instead of explaining verbally what they are looking for. There is also the issue of pacing: speaking too much breaks down the rehearsal rhythm, it interrupts the music flow and the players' concentration, resulting in lower efficiency. However, short and precise instructions were considered efficacious during rehearsing. (Biasutti 2012: 64)

In solchen Studien wird eine vermehrte Kommunikation über Dirigiergesten propagiert; zu viele verbale Anweisungen bzw. Unterbrechungen des/der Dirigent:in werden dahingegen als weniger effizient für einen konstruktiven Fortgang der Probentätigkeit gehandelt. Unsere Daten – so wird sich im empirischen Teil dieser Studie zeigen – beweisen allerdings das Gegenteil. Der verbale Anteil in den Proben ist hoch, d. h., die Dirigent:innen in unserem Korpus unterbrechen oft die Musik, um Musikalisches zu besprechen, zu korrigieren und neu anzugeben. Die vorliegende Arbeit soll dabei helfen, die Lücke, die es in der Untersuchung der Kommunikation zwischen Dirigent:in und Musiker:innen in Orchesterproben gibt, zu schließen. Nachfolgend sollen einschlägige Studien zur Erforschung der Orchesterprobe als Kommunikationssituation skizziert werden. Dabei werden vor allem

solche Studien berücksichtigt, die die Orchesterprobe dezidiert empirisch, d. h. auf der Basis authentischer Daten erschließen.

Bereits Emanuel A. Schegloff – einer der Hauptvertreter der Konversationsanalyse – äußerte in einem Interview eine noch unkonkrete Forschungsidee, wie – in Anlehnung an „Making Music Together“ von Schütz (1964) – Musikmachen untersucht werden könnte:

The initial motivating idea was to examine several distinct orders of interactivity which supply the infrastructure for making music together: the embodied interactive conduct of playing itself, the interaction at rehearsal through which the playing is developed, and the interaction written into the score by the composer. (Prevignano/Thibault 2003: 15–16)

Schegloff bezog sich in seiner Aussage auf gesammelte Videodaten von Proben sowie der anschließenden Aufführung eines Streichquartetts. Die Studien von Haviland (2007, 2011), in denen er sich der konversationsanalytischen Methode folgend der Frage der Koordination in Masterklassen und Ensembleformaten widmet, stellen eine erste Umsetzung der Idee von Schegloff dar. Haviland zeigt, wie professionelle Musiker:innen verbale Kommentare und Vorschläge mit körperlichen Ausdrucksformen (*embodied demonstrations*) abwechseln, um Musikstudent:innen ein Feedback zu ihrer musikalischen Performance zu geben. Auf diese Weise kann illustriert werden, wie eine musikalische Passage mit einem Streichinstrument nicht gespielt oder richtig gespielt werden soll (vgl. Haviland 2007: 155–159). Außerdem konstatiert der Autor eine multimodale Koordination zwischen den beteiligten Musiker:innen (Profis/Master und Lernende). Diese entsteht aus dem komplexen Zusammenspiel der unterschiedlichen Ausdrucksmöglichkeiten von Sprache, Gestik, Gesang, Spiel und Mimik. In musikalischen Masterklassen werden so die Körper der Musiker:innen – laut Haviland (2007: 170–171) – „at once vehicles of performance and meta-performance, means for making music and for communicating about music“. Bedeutung wird demnach auf dynamische und multimodale Art und Weise hergestellt.

Auch Keating (1993) arbeitet mit authentischen Daten, wenn sie Korrektursequenzen in einer Bandprobe untersucht. Die Autorin stellt fest, dass „in the context of the music rehearsal, error correction [...] an important focus of group activity“ ist (Keating 1993: 412). Sie versteht Korrekturen als Turns, die anderen Bandmitgliedern anzeigen, wie sie etwas spielen sollen (z. B. „do it with punch“). Solche Korrekturen tragen dazu bei, dass Aushandlungen stattfinden und auch andere Bandmitglieder ihren Standpunkt einbringen können: „[...] a potentially divisive activity (correction) becomes a source of negotiation and invitation for collaboration.“ (Keating 1993: 419)

Szczepk Reed, Reed & Haddon (2013) verwenden in ihrer Studie zu Gesangs-Masterklassen den Begriff „directives“ anstelle von Korrekturen. Sie

untersuchen Übergänge zwischen Instruktions- und Performanceteilen sowie die Platzierung von Instruktionen in Masterklassen-Sequenzen für Gesang. Unter Instruktionen (oder „directives“) verstehen die Autor:innen „[...] instructional first actions, which initiate and make relevant complying response actions, oriented to putting the directive into practice“ (Szczepk Reed/Reed/Haddon 2013: 26). Solche Instruktionen werden von Masters gegeben, indem sie mehrere semiotische Ressourcen kombinieren (etwa Sprache, Gestik, physische Orientierung und Proxemik), und sind damit Teil des ‚doing being a master‘. Die Gesangsstudent:innen setzen die Instruktionen um, wenn sie performen, und zeigen damit ein ‚doing being a student‘ an („Putting directives into practice is the default and most relevant type of next action for students“, Szczepk Reed/Reed/Haddon 2013: 42). Die Instruktionen des Masters und die Realisierung durch die Student:innen bedingen sich dabei gegenseitig: Eine Instruktion macht eine musikalische (bzw. gesangliche) Umsetzung („embodied musical performance“) relevant. Instruktionen in Masterklassen können in zwei unterschiedliche Typen unterteilt werden: jene, die eine Umsetzung an einem unspezifischen, späteren Punkt relevant setzen („non-local action directives“), und solche, die eine unmittelbare Umsetzung erfordern („local action directives“). Letztere werden von Student:innen als mögliche Gelegenheiten für einen Wiedereinstieg in die Performance interpretiert (das ist Teil von ‚doing being a learner‘). Allerdings ist nicht immer klar (z. B. innerhalb von Instruktionsclustern, d. h. mehreren aneinander gereihten Instruktionen), ob die Instruktionssequenz des Masters bereits abgeschlossen ist und einen Wiedereinstieg initiiert („instruction final“) oder ob noch weitere Instruktionen folgen können („instruction internal“). Die grundsätzliche Frage, die sich hier stellt, ist „NOW or NOT NOW?“. Allerdings entscheidet immer der Master – und nicht der/ die Gesangsstudent:in – wann es zu einem Wiedereinstieg kommt.

Erste ethnomethodologisch ausgerichtete Arbeiten zu Proben von Amateuorchestern, in denen der Autor teils selbst als Musiker mitwirkt, stammen von Peter Weeks (1985, 1990, 1996). In seiner ersten Arbeit (1985) – „Error Correction Techniques and Sequences in Instructional Settings: Toward a Comparative Framework“ – vergleicht Weeks Reparatursequenzen<sup>47</sup> in Vorleserunden im schulischen Unterricht mit Reparaturen/Korrekturen in Orchesterproben. Der Autor arbeitet einen Unterschied zwischen ‚regulären‘ und aufgeführten Interaktionsanteilen heraus: In der ‚regulären‘ Interaktion wird die aufgeführte Aktivität themati-

---

<sup>47</sup> Reparaturen sind hier im Sinne der ethnomethodologischen Konversationsanalyse (KA) zu verstehen, d.h., es handelt sich um Elemente der Interaktion, die von den Gesprächsteilnehmer:innen als ‚reparaturbedürftig‘ behandelt werden. In der Orchesterprobe entscheidet der/die Dirigent:in als „authorized correction-agent“ (Weeks 1996: 253), welche musikalischen Segmente reparaturbedürftig sind, und kommuniziert diese den Musiker:innen.

siert. Das heißt, es gibt zwei Aktivitätssysteme, die miteinander verzahnt sind: die ‚reguläre Interaktion‘ und die ‚Performance‘. Die beiden Settings unterscheiden sich darin, dass in der Schule der/die Lehrer:in den Fehler (*correctable*) identifiziert und die Schüler:innen dazu animiert, sich selbst zu korrigieren (der/die Lehrer:in kennt die richtige Antwort). In der Orchesterprobe bestimmt der/die Dirigent:in das *correctable* und führt außerdem die Korrektur selbst aus (etwa durch verbale Erläuterungen oder durch eine gesangliche Demonstration, wie die betreffende Stelle zu spielen sei). Erst im Anschluss an diese Fremdkorrektur – die so in Vorleserunden nicht vorkommt – führen die Musiker:innen die Selbstkorrektur in einer musikalischen Performance aus.

In seiner Studie „Musical Time as a Practical Accomplishment: A Change in Tempo“ aus dem Jahr 1990 behandelt Weeks die zeitliche Koordinierung und Synchronisation in Orchesterproben. Diese wird durch das Zusammenspiel von verbalen und gesanglich-imitativen Äußerungen, von Partitur und der Aktivität des Musikmachens an sich verwirklicht. Weeks (1990) konstatiert, dass das Singen und das (Mit-)Zählen des Dirigenten während des Dirigierens in der Lage sind, Zeitlichkeit (z. B. wie ein in der Partitur vermerktes Ritardando zu spielen ist) zu verkörpern („embody musical time“) und damit auch besser zu vermitteln. Verbale Instruktionen und die Partitur können dahingegen nur etwas über diese Zeitlichkeit aussagen („refer to musical time“). Um Zeitlichkeit in der Musik auszudrücken und damit verbunden auch Synchronisation herzustellen, eignen sich – so Weeks (1990: 347–348) – am besten Singen bzw. – wie er es nennt – „imitative expressions“: „[...] IEs [imitative expressions] are alternative embodiments of musical time as distinct from the actual playing.“

In der Untersuchung von 1996 – „A Rehearsal of a Beethoven Passage: An Analysis of Correction Talk“ – beschreibt Weeks die Probenarbeit im Orchester als „being oriented to corrections as the major focus-of-activity“ (Weeks 1996: 260). Eine Orchesterprobe besteht aus Korrektursequenzen („correction sequences“), die vom Dirigenten als „authorized correction-agent“ (Weeks 1996: 253) initiiert werden. Dabei steht die grundlegende Frage im Raum: Wie können Korrekturen und Anweisungen eines/einer Dirigent:in an ein Orchester kommuniziert werden? Korrekturen können – nach Weeks’ Beobachtungen – durch zwei kommunikative Formate realisiert werden. Der Autor unterscheidet zwischen verbalen Instruktionen bzw. Spielanweisungen an die Musiker:innen in wortsprachlicher Form (z. B. Erklärungen zu gewünschten musikalischen Effekten oder technische Anweisungen, wie diese erreicht werden können) und sog. „illustrative expressions“ (Weeks 1996: 254). Dabei handelt es sich um Ausdrucksformen wie Singen, Sprechgesang und (Mit-)Zählen, mit denen – neben Sprache – korrigiert und instruiert wird und die auch von Gesten begleitet sein können. Durch Gesang können etwa bestimmte musikalische Eigenschaften markiert werden, wie z. B.

Rhythmus, Tempo oder Tonhöhe. Sehr häufig tauchen auch sog. „contrast pairs“ (Weeks 1996: 269–275) auf, durch die der Dirigent einen Kontrast zwischen einer nicht gewünschten („faulted performed version“) und einer gewünschten Version („the conductor's prescribed version“) einer musikalischen Passage aufzeigt. Dafür imitiert er gesanglich die alte, falsche Version, die vorher von den Musiker:innen gespielt wurde, und stellt sie einer neuen, richtigen Version gegenüber. Kontrastpaare sind dabei zumeist in den verbalen Rahmen „not [IE1] but [IE2]“ (271) eingefasst, der Kontrast selbst wird gesanglich oder lautmalerisch realisiert. Durch Kontrastpaare weist der/die Dirigent:in die Musiker:innen an, wie eine Stelle in der Partitur gespielt werden soll.

Auch Veronesi (2014) beschäftigt sich mit der Organisation von Korrekturen in der Interaktion zwischen Dirigent und Musiker:innen. Die Autorin ergründet in ihrer der konversationsanalytischen Methodik folgenden Untersuchung, wie ein Lexikon an ideographischen Zeichen und Gesten dem Dirigenten ermöglicht, in Verbindung mit Sprache Korrekturen zum Spielen der Musiker:innen anzu bringen und sequenziell zu organisieren. Dieses Lexikon wird unter dem Begriff „Conduction®“ zusammengefasst. Es wurde vom US-amerikanischen Dirigenten, Komponisten und Arrangeur Lawrence D. „Butch“ Morris (1947–2013) ins Leben gerufen und bezeichnet ein Inventar an 55 Dirigiergesten, anhand derer der/die Dirigent:in Harmonie, Melodie, Rhythmus, Tempo, Artikulation, Phrasen u. Ä. an die Musiker:innen vermitteln kann, ohne dabei Noten oder eine Partitur zur Hilfe zu nehmen (vgl. Morris 1995, 2006). Veronesi illustriert, welche verbalen und non-verbalen Mittel zur Koordination genutzt werden, wenn die Beteiligten nicht auf eine Notation zurückgreifen können. Dabei stellt die Autorin fest, dass der Dirigent vor allem gesangliche Demonstrationen einsetzt, um Fehler (oder *correctables*) zu lokalisieren und darüber hinaus eine Möglichkeit zu veranschaulichen, wie der Fehler korrigiert werden kann (u. a. auch innerhalb von Kontrastpaaren, vgl. Weeks 1996). Durch verbale Beschreibungen in Kombination mit Gestik wird angezeigt, wie die vom Dirigenten ausgeführten Dirigiergesten technisch ausgeführt werden sollen, z. B. hält der Dirigent seine linke Hand mit der Handfläche nach oben auf Brusthöhe und erklärt „this means continuous sound“ (Veronesi 2014: 479). Außerdem hebt Veronesi (2014: 485) die Bedeutung des Körpers des Dirigenten als „meaning-making resource“ hervor: „[...] hands and baton – silent instruments through which conductors can shape musical performance – become sound-making tools for the purpose of pedagogical demonstration.“

Mit der interaktiven Wissensvermittlung in Orchesterproben setzt sich die Studie von Parton (2014) auseinander. Die Autorin stellt die Interaktion in der Orchesterprobe als Vermittlung von Wissen des Dirigenten an die Musiker:innen dar. Diese wird durch verbales Feedback und Instruktionen erreicht und modelliert in der Folge das Verhalten bzw. das Spielen der Musiker:innen. Der Dirigent

hat vor der Probe eine Vorstellung davon, „how to play for this performance“ (Parton 2014: 415). Diese Vorstellung gilt es den Musiker:innen während der Probe zu vermitteln: „A conductor moves the players from K- [less knowledge] about his orchestral vision to somewhere further along the gradient towards K+ [more knowledge].“ (Parton 2014: 415) Parton stellt fest, dass Dirigenten ihre Bewertungen in der Kommunikation als subjektiv markieren (durch „I“, „to me“ oder „I hear“), sodass andere Perspektiven offenbleiben. Das bedeutet, dass sich der Dirigent nicht als autoritäre Figur positioniert, sondern „in the moment“ (Parton 2014: 416), d. h., er zeigt an, dass er sich konstant mit dem auseinandersetzt, was das Orchester spielt, und dass seine Redebeiträge interaktional ausgerichtet sind. Das führt dazu – wie Parton (2014: 416) konstatiert –, dass die Orchestermusiker:innen „[...] become part of a collaboration construction, or re-construction, of the final, desired performance vision“.

In den beschriebenen Studien wird wiederholt auf den multimodalen Charakter von Orchesterproben bzw. auf das Zusammenspiel von unterschiedlichen semiotischen Ressourcen hingewiesen, die in der Interaktion zwischen Dirigent:in und Musiker:innen zum Tragen kommen. Arbeiten, die dezidiert spezifische Modalitäten wie Gestik oder Mimik des/der Dirigent:in erörtern, stammen von Parton & Edwards (2009), Bräm & Boyes Bräm (1998) sowie Poggi (2002). Während sich Parton & Edwards (2009) und Bräm & Boyes Bräm (1998) mit Gesten von Dirigenten auseinandersetzen (vgl. dazu auch die Ausführungen in Kap. 2.1.6), erörtert Poggi (2002) die kommunikativen Funktionen des Blicks, der Kopfbewegungen und des Gesichtsausdrucks des Dirigenten und entwickelt daraus ein Lexikon. Dabei unterscheidet die Autorin zwischen semiotischen Ressourcen, die eingesetzt werden, um zu vermitteln, wer spielen soll (z. B. durch einen Blick zu den betreffenden Musiker:innen), wann gespielt werden soll (z. B. durch ein Hochziehen der Augenbrauen vor dem Start der Musik) und wie gespielt werden soll (z. B. bedeutet ein Stirnrunzeln, dass *forte* gespielt werden soll).

Erwähnenswert ist außerdem die Studie von Schütte (1991) zur Scherzkommunikation unter Orchestermusiker:innen. Während in anderen Arbeiten zur Kommunikation in Orchesterproben vor allem dirigentenseitige Anforderungen in den Blick genommen werden, ist in der Untersuchung von Schütte die Perspektive der Orchestermusiker:innen vertreten. Hierbei handelt es sich um eine soziolinguistisch orientierte Arbeit, die die soziale und sprachliche Welt von Musiker:innen im Orchester anhand empirischen Materials (bestehend aus aufgezeichneten Proben- und Pausengesprächen sowie Interviews mit Musiker:innen) beschreibt. Der Autor stellt sprachliche Handlungsmuster heraus, die er unter der Berücksichtigung von beruflich-strukturellen Gegebenheiten, kulturellen Standards und Wertvorstellungen illustriert. Die Musiker:innen bearbeiten Probleme und Widersprüche auf der Grundlage von „joking relationships“ (Schütte 1991: 61): Sie gehen

ernsthaften Austragungen von Konflikten aus dem Weg und greifen stattdessen auf Witze und Scherze zurück. Zur Scherzkommunikation unter den Musiker:innen gehören unterschiedliche Techniken, wie scherzhafte Notengraffiti (z. B. fand der Autor in einer Stimme der ersten Violinen zu Haydns „Die Schöpfung“ auf der vorletzten Seite links oben, wo rechts unten der Schlussstrich beim ‚Fine‘ sichtbar wird: „Land in Sicht!“), Mitteilungen durch Musizieren (z.B. durch den Siegfriedsruf aus Richard Wagners Oper „Siegfried“), die Übertragung von Textzitaten aus musikalischen Werken in den aktuellen Kontext (etwa wenn der zu singende Text „Wacht auf, wacht auf!“ aus der Bachkantate Nr. 20 auf das Stereotyp des Kirchenschlafs während der Predigt bezogen und als Weckruf interpretiert wird), oder auch das Erzählen von Witzen über spezifische Instrumentengruppen (etwa „Was haben die Bratscher und die DDR gemeinsam? – Die kämpfen um die Anerkennung“). Daneben gibt es auch nonverbale Mittel der Scherzkommunikation, wie Manipulationen am Instrument eines/einer anderen Musiker:in, falsche Einsätze sowie Täuschungsmanöver aller Art.

Nicht zuletzt gibt es Studien zu Orchesterproben, die einem pädagogischen Ansatz folgen. Herrle & Egloff (2015) untersuchen beispielsweise ein Laienorchester und beschreiben dieses als Ort, an dem ein Lernprozess durch das gemeinsame Musizieren von Erwachsenen und durch die Vorbereitung auf eine Aufführung in Gang gesetzt wird. Die Erwachsenen lernen ihr Instrument besser zu spielen; darüber hinaus lernen sie über das gespielte Werk und dessen/deren Komponist:in dazu. Ziel der Untersuchung von Herrle & Egloff (2015) ist es, Muster in der Interaktion zwischen Dirigent und Musiker:innen zu identifizieren. Hierfür bündeln die Autor:innen thematisch eingegrenzte Aktivitäten, etwa das Herstellen von Bereitschaft zum gemeinsamen Musizieren, das Unterbrechen der Musik, das Anbringen von Korrekturen, das Anweisen, wie das Spiel verändert werden soll, das Adressieren von bestimmten Instrumentengruppen, das Leiten des Zusammenspiels der Musiker:innen sowie das Überprüfen des Spiels und der Umsetzung der Anweisungen. Alle diese Aktivitäten werden vom Dirigenten ausgeführt und von den Musiker:innen nicht in Frage gestellt. Damit einher geht auch ein asymmetrisches Interaktionsverhältnis: „Führen und Leiten auf der einen Seite, Unterordnen und Relationieren der eigenen Bedeutsamkeit gegenüber dem Gesamtklang auf der anderen Seite bilden ein konstituierendes Moment des Probens.“ (Herrle/Egloff 2015: 266) Die Studie von Herrle & Egloff (2015) hat mit den oben aufgeführten gemeinsam, dass die Autor:innen auch hier das Ineinanderwirken von mehreren Modalitäten, vor allem in den Korrekturen und Anweisungen des Dirigenten, beobachten. Um Anweisungen zu geben, verwendet der Dirigent Sprache, Gestik (die Autor:innen sprechen von „emblematischen Bewegungen“), lautmalerische Beschreibungen in Verbindung mit Vormachen sowie prosodische

Hervorhebungen. Ähnliches konstatiert Adenot (2015) in einer ethnographischen Studie zur Autorität von Dirigenten in Orchesterproben:

[...] the conductor gives instructions, not all of them in formal language: he may use the vocational vernacular (a mix of technical and musical terms), but as music is experienced corporally, he may also resort to singing, miming and onomatopoeia. (Adenot 2015: 4)

## 2.3 Datenbasis und Methode

Das folgende Kapitel widmet sich der theoretisch-methodischen Grundlegung der vorliegenden Arbeit, befasst sich mit technisch-praktischen Aspekten der Datenerhebung und beschreibt das dieser Studie zugrundeliegende Datenkorpus.

### 2.3.1 Methodischer Zugang

Die Untersuchung in der vorliegenden Arbeit stützt sich im Wesentlichen auf zwei Analysemethoden: die ethnomethodologische Konversationsanalyse oder genauer eine multimodal erweiterte Konversationsanalyse<sup>48</sup> sowie ethnographische Beobachtungen und Interviews. Das heißt, die vorliegende Arbeit verfolgt einen gesprächsanalytischen Ansatz, der „im Kern auf den Erkenntnissen und der analytischen Mentalität der Konversationsanalyse“ (Deppermann <sup>2</sup>2001: 10) beruht, allerdings in gewisser Hinsicht über konversationsanalytische Fragestellungen und Vorgehensweisen hinausgeht, indem sie z. B. ethnographische Daten mit in die Analyse integriert. Die ursprüngliche Konversationsanalyse propagiert dagegen eine streng naturalistische und in sich geschlossene Handhabung der Daten. Der/die Forscher:in kann von den Daten ablesen, wie Interaktant:innen Sinn und Ordnung in ihrem Tun herstellen und davon ausgehend ein ereignisgetreues Protokoll bzw. Transkript erstellen. Deppermann (2000: 99) kritisiert dieses „display“-Konzept der klassischen Konversationsanalyse und appelliert an ein gesprächsexternes Wissen (etwa Alltagswissen, ethnographisches Wissen, theoretisches Wissen), ohne das Interaktionen nicht (vollständig) interpretierbar sind. Vor allem ethnographische Daten, wie z. B. Interviews, informelle Gespräche, schriftliche und visuelle Dokumente sowie die teilnehmende Beobachtung, können dabei helfen, möglichst viele Facetten eines sozialen Feldes zu beschreiben, Interpretationslücken zu schließen und vor Fehlinterpretationen zu schützen

---

<sup>48</sup> Für die beiden Analysemethoden vgl. die jeweiligen theoretischen Ausführungen dazu (Kap. 2.1.3 und Kap. 2.1.5).

(vgl. Deppermann 2000). Dadurch ergibt sich ein holistisches Bild von Interaktion. Auch Drew & Heritage (1992) vertreten eine zweigleisige Analyse, vor allem im Hinblick auf institutionelle Kommunikationssituationen, worunter auch die Orchesterprobe fällt:

In the context of institutional talk, this means that empirical analysis must first accomplish the normal CA tasks of analysing the conduct of the participants including their orientations to specific local identities and the underlying organization of their activities. Additionally, however, analysis will normally be concerned to show that the participants' conduct and its organization embody orientations which are specifically institutional or which are, at the least, responsive to the constraints which are institutional in character or origin. (Drew/Heritage 1992: 20)

Nach Drew & Heritage (1992) soll in zwei Analyseschritten vorgegangen werden. Zunächst sollen der Konversationsanalyse entsprechend die formalen Ordnungsprinzipien der Interaktion gefiltert werden. Im Anschluss daran soll der Zusammenhang zwischen den Interaktionsstrukturen und der sozialen Makrostruktur der Institution erforscht werden. Um diesen zweiten Schritt ausführen zu können, bedarf es eines Wissenshintergrunds (entweder Alltagswissen oder ethnographisches Wissen), der über das in der aktuellen Interaktion zugänglich gemachte Wissen hinausgeht (vgl. auch Thörle 2005: 50). Aus den genannten Gründen soll deshalb auch in der vorliegenden Arbeit auf ethnographische Daten (Interviews, informelle Gespräche, teilnehmende Beobachtung) zurückgegriffen werden. Allerdings muss bedacht werden, dass der Einsatz der Ethnographie bei der Konversationsanalyse auch die Gefahr birgt, Gespräche nur noch als „Illustrationen der vorgefertigten Ansichten über das Feld“ (Deppermann 2000: 116; vgl. auch Thörle 2005: 52) zu behandeln. Deshalb empfiehlt es sich, auf ethnographisches Wissen erst dann zu rekurrieren, wenn die Interpretation der manifesten Aktivitäten der Beteiligten völlig ausgeschöpft ist.

### 2.3.2 Einstieg in das Untersuchungsfeld

Orchesterproben sind für die Öffentlichkeit streng genommen nicht zugänglich (mit Ausnahme der Generalprobe vor einem öffentlichen Konzert) und stellen als zu erforschendes Feld – wie bereits weiter oben angeführt (vgl. Kap. 2.2) – sog. „closed access groups“ (Hornsby-Smith 1993: 53) dar, die ausschließlich für Autorisierte (Dirigent:in, Musiker:innen, Zuständige für Administration, Techniker:innen usw.) reserviert sind. Deshalb lag die erste Schwierigkeit darin, geeignete Symphonieorchester in Frankreich und Italien ausfindig zu machen, die bereit waren, ihre Probenarbeit mit Videokameras aufnehmen zu lassen. Hinzu kamen weitere Umstände, die den Zugang zum Untersuchungsfeld erschwerten,

darunter die eventuell zu erwartende Störung der Probezeit, die relativ große Einmischung in die Intimität der Orchesterprobe durch die meist als invasiv wahrgenommene Datenerhebung mittels Videokameras (vgl. Heath/Hindmarsh/Luff 2010: 17) sowie die Auseinandersetzung mit ethischen Prinzipien. So war es für die Orchestermusiker:innen und die jeweiligen Dirigent:innen wichtig, über das Forschungsprojekt und insbesondere die Handhabung der videographicen Daten Bescheid zu wissen. Erst durch die schriftliche Zusicherung der Nicht-Veröffentlichung der Videos und der rein für Forschungszwecke unternommenen Datenaufbereitung konnte auch die Forschungsbeziehung zwischen der Forscherin und den an den jeweiligen Proben Beteiligten in ihrer Rolle als Beforschte aufgebaut werden.

Für das vorliegende Projekt fungierte der Vorsitzende der Vereinigung der französischen Symphonieorchester<sup>49</sup> als sog. *Gatekeeper* (Breidenstein et al. 2013: 52). Er ermöglichte einen ersten Eintritt in das Forschungsfeld durch die Vermittlung von bereitwilligen französischen Orchestern. Mit der Unterstützung des Direktors der Orchestervereinigung im Rücken war es einfacher, das Vertrauen von Orchestern zu gewinnen und sie zur Teilnahme an dem Forschungsprojekt zu motivieren. An der Herstellung einer Forschungsbeziehung zwischen Feldforscherin und Beforschten, dem sog. *Rapport*<sup>50</sup> (Breidenstein et al. 2013: 61), waren daraufhin maßgebend die Verantwortlichen für die Administration der Orchester beteiligt. Für die Vertrauengewinnung des *Haydn Orchester Bozen* stellte sich die bereits durchgeführte Datenerhebung bei französischen Orchestern als vorteilhafte Basis heraus. Ein zusätzliches erklärendes Gespräch mit der Direktorin des Orchesters über Forschungsziele und Methoden sicherte auch hier den Feldzugang.

Das Datenmaterial der vorliegenden Studie besteht aus Videoaufnahmen natürlicher Interaktionen in der Orchesterprobe. „Natürlich“ bedeutet in diesem Zusammenhang, dass die Aufnahmen in Gesprächssituationen stattfanden, die nicht spezifisch für Forschungszwecke arrangiert und zudem so wenig wie möglich durch die Anwesenheit der Forscherin beeinflusst wurden (vgl. Hutchby/Wooffitt 1998: 14; Mondada 2012b: 33). Videoaufnahmen ermöglichen es, nicht nur verbale Interaktion, sondern auch Nonverbales wie Gestik, Mimik, Blicke, Körpertaltung, Körperbewegungen, Positionierung der Teilnehmer:innen und Proxemik, körperlich-räumliche Aspekte usw. festzuhalten (vgl. Gürlich/Mondada 2008: 29). Ein weiterer Vorteil des Mediums Film besteht darin, dass Interaktionssituationen

---

<sup>49</sup> Philippe Fanjas, *Directeur de l'Association Française des Orchestres*.

<sup>50</sup> Der Rapport manifestiert sich nicht nur in der Erlaubnis, Zugang zum Feld zu erhalten, sondern auch in einer auf Vertrauen aufgebauten Beziehung zwischen Forscher:in und Beforschten (Breidenstein et al. 2013: 60).

mit mehreren Teilnehmenden aufgenommen und der Analyse zugänglich gemacht werden können (vgl. Heath/Hindmarsh/Luff 2010: 7–8).

Gleichzeitig bringt die Videographie allerdings auch Schwierigkeiten mit sich, und zwar bereits bei der Aufnahme selbst. Die Videokamera und der/die Forscher:in sind in der Aufnahmesituation ständig – wenn auch mit der Zeit für die Beforschten nur mehr unbewusst – präsent und können demnach auch in das Blickfeld der beforschten Gruppe rücken (vgl. Mondada 2012b: 34). Labov (1980: 17) bezeichnet dieses Phänomen als *Beobachterparadoxon*: „Um die Daten zu erhalten, die am wichtigsten für die linguistische Theorie sind, müssen wir beobachten, wie die Leute sprechen, wenn sie nicht beobachtet werden.“ Mondada (2012b: 34) schlägt vor, aus dieser Eigenheit der Videographie Nutzen zu ziehen und Momente, in denen sich die Orientierung der Forschungsteilnehmer:innen auf die Videokameras richtet, mit in die Analyse einzubeziehen (was allerdings in den Daten für die vorliegende Studie nur äußerst selten vorkommt).

Außerdem ist zu beachten, dass jede Aufnahme nicht nur ein technischer Schritt ist, sondern bereits selektive Schritte miteinschließt, die für die folgende Analyse entscheidend sind (vgl. Heath/Hindmarsh/Luff 2010: 37; Mondada 2012b: 39). Die gewonnenen Videodaten stellen ja nicht die Wirklichkeit dar, sondern repräsentieren diese nur und sind immer hinsichtlich des Forschungsinteresses orientiert (vgl. Hutchby/Wooffitt 1998: 74). Je nach Aufnahmeausschnitt der Kamera/s, Position der Geräte und deren Beweglichkeit oder Statik ändern sich die für die Analyse zugänglichen Daten (vgl. Luff/Heath 2012: 260; Mondada 2012b: 39–41).

Während der Aufnahmen wurde mit zwei Videokameras gearbeitet, die beide fix auf Stativen positioniert waren: eine entweder hinter den Violinist:innen oder den Bratschist:innen mit Ausrichtung auf den/die Dirigent:in, eine zweite vor dem Orchester mit Blick auf die Musiker:innen. Beide Kameras waren situationsbedingt schwenkbar, sodass es möglich war, Positionsveränderungen des/der Dirigent:in oder auch der Musiker:innen zu berücksichtigen sowie Interaktionen zwischen dem/der Dirigent:in und einzelnen Musiker:innen festzuhalten. Zur Erzielung einer akzeptablen Tonqualität wurde außerdem auf einem Notenpult nahe des Dirigentenpults ein externes Audioaufnahmegerät platziert, das per Funkgerät den Ton direkt in die „Dirigent:in-Kamera“ einspeiste.

### **2.3.3 Korpus**

Das Korpus der vorliegenden Studie besteht aus rund 55 Stunden an Videodatenmaterial, das im Zeitraum zwischen Februar und Juli 2016 in den Proben des *Orchestre de Paris*, des *Orchestre de l'Opéra de Rouen* und des *Haydn Orchester*

**Tab. 3:** Datenkorpus.

Orchester	Dirigent:in	Dauer der Aufnahmen	Werke
<i>Orchestre de Paris</i>	Gianandrea Noseda	9:11:05	VERDI, <i>Messa da Requiem</i>
<i>Orchestre de l'opéra de Rouen</i>	Antony Hermus	15:28:24	DVOŘÁK, Symphonie Nr.8; GLANERT, Brahms-Fantasie; BRAHMS, Double Concerto pour violon, violoncello et orchestre;
<i>Haydn Orchester Bozen I</i>	Sybille Werner	06:22:46	MAHLER, Symphonisches Präludium; SMETANA, Vltava „Die Moldau“; MAHLER, Symphonie Nr. 1 D-Dur;
<i>Haydn Orchester Bozen II</i>	Arvo Volmer	12:18:02	ŠOSTAKOVIČ, Concerto Nr. 1 in do minore; STRAWINSKI, Suite (Pulcinella); PROKOFJEW, Symphonie Nr. 1 D-Dur; PÄRT, Cantus;
<i>Haydn Orchester Bozen III</i>	Roberto Molinelli	12:18:05	BOWIE, Let's dance; O'BRENNAN, Orinoco Flow; DE MORAES, La vita e l'arte dell'incontro; BERNSTEIN, Symphonic Dances from West Side Story; PIAZZOLLA, Libertango; MORRICONE, The Good, the Bad and the Ugly; ROSSINI, La danza; BRAHMS, Ungarischer Tanz Nr. 1 & 5; BARTÓK, Rumänischer Tanz;

*Bozen* aufgenommen und gesammelt wurde (vgl. Tab. 3). Es umfasst fünf verschiedene Orchesterprojekte mit fünf unterschiedlichen Dirigent:innen: ein Projekt des *Orchestre de Paris* mit Gianandrea Noseda, ein Projekt des *Orchestre de l'Opéra de Rouen* mit Antony Hermus sowie insgesamt drei Projekte des *Haydn Orchester Bozen* mit den jeweiligen Dirigent:innen Sybille Werner, Arvo Volmer und Roberto Molinelli. Außer Arvo Volmer, der Hauptdirigent des *Haydn Orchester Bozen* ist, waren die vier anderen als Gastdirigent:innen bei und mit den Orchestern tätig.

Die vorliegende Studie war von Anfang an als explorative Untersuchung angelegt, zu keinem Zeitpunkt ging es um eine quantitative Auswertung bzw. die Generierung statistisch relevanter Ergebnisse. Auf der Grundlage von detaillierten Analysen sollen relevante Aspekte rekonstruiert werden, die in analytischer und methodischer Hinsicht für die Interaktion in Orchesterproben konstitutiv

sind. Dafür wurden aus dem Korpus repräsentative Fallbeispiele ausgewählt,<sup>51</sup> die nach GAT 2 (Selting et al. 2009) den verbalen Aspekt und in Anlehnung an Mondada (2016b) auch die nonverbalen Aspekte der Interaktion zwischen Dirigent:in und Musiker:innen erfassen. Ausschnitte aus dieser Interaktion werden in den Analysen jeweils mit einem Verweis auf die entsprechenden Transkriptzeilen und kursiviert zitiert. Außerdem wird mit Screenshots gearbeitet, um Nonverbales besser zu veranschaulichen und in seiner Gesamtheit zu illustrieren.

Das folgende Beispiel zeigt ein nach GAT 2-Konventionen angefertigtes Transkript, das um die Beschreibung einer Zeigegeste und des Blickverhaltens des Dirigenten sowie das gleichzeitige Spielen der Musiker:innen in Z01 erweitert wurde:

---

01	D	&*%(0.6) okay! (.) can we try one time sorry sorry&%	
	d	&streckt Arme nach vorne	----->&
		*Blick-->MM rechts	----->*
	mm	%J((spielen 2.7''))♪	----->%
02		<<rau, gedrückt, tief> !JA:M! (.) !PA:M! (.) !PA:::>	

---

Die Zeigegeste und die Blickrichtung des Dirigenten wurden in einer interlinearen Darstellungsform in eigenen Zeilen (zwischen den Zeilen des Verbaltranskripts) mit „streckt Arme nach vorne“ und „Blick->MM rechts“ paraphrasiert. Auch das Spielen der Musiker:innen wurde in einer eigenen Zeile angeführt. Werden visuelle Phänomene notiert, so werden die Siglen der Interaktionsbeteiligten kleingeschrieben („d“ und „mm“). Die Bewegungs- und Blickphasen des Dirigenten und die musikalische Handlung der Musiker:innen wurden mithilfe von Zeichen („&“ für Handlung A, „\*“ für Handlung B des Dirigenten; „%“ für Handlung A der Musiker:innen usw.) in Bezug auf deren Anfangs- und Endpunkt im Verbaltranskript verankert.<sup>52</sup> Auf diese Weise kann die Synchronisation unterschiedlicher Ausdrucksmodalitäten (hier: Sprache, Gestik, Blick und Musik) dokumentiert werden.

---

51 Die Aufnahmen liegen etwa zu 75% transkribiert vor.

52 Für die unterschiedlichen verwendeten Zeichen und ihre Zuordnung zu den jeweiligen Beteiligten vgl. die Transkriptionskonventionen im Anhang.