

Aus dem Institut für Psychotherapie  
und Medizinische Psychologie  
der Universität Würzburg  
Vorstand: Professor Dr. med. Dr. phil. Hermann Lang

---

**Evaluation von Beobachtungsskalen zur Beurteilung  
musiktherapeutischer Improvisationen**

Inaugural – Dissertation  
zur Erlangung der Doktorwürde der  
Medizinischen Fakultät  
der  
Bayerischen Julius-Maximilians-Universität zu Würzburg

vorgelegt von  
**Niels Schübel**  
aus Gütersloh

Würzburg, Januar 2003

**Referent:** Prof. Dr. Dr. Hermann Lang

**Koreferent:** Prof. Dr. Dr. Gundolf Keil

**Dekan:** Prof. Dr. Silbernagl

**Tag der mündlichen Prüfung:** 19.10.2004

**Der Promovend ist Arzt**

# Inhaltsverzeichnis

	Seite
<b>1 Einleitung</b>	1
<b>1.1 Überblick</b>	1
<b>1.2 Historischer Rückblick</b>	1
1.2.1 Musik und Magie/Ekstase	2
1.2.2 Musik und Harmonie	3
1.2.3 Musik und Iatrophysik	5
<b>1.3 Musiktherapie heute</b>	6
1.3.1 Formen von Musiktherapie	6
1.3.2 Theoretische Grundlagen	7
1.3.2.1 Musik und Kommunikation	8
1.3.2.2 Musik und das Unbewußte	10
<b>2 Stand der Forschung</b>	11
<b>2.1 Einleitung</b>	11
2.1.1 Wahrnehmungsphysiologie	11
2.1.2 Psychosomatische Forschung	12
2.1.3 Beobachtung und Beschreibung von Musik	13
<b>2.2 Aktuelle Forschungsarbeiten</b>	13
2.2.1 Zusammenfassende Beurteilung	28
<b>3 Methodik</b>	29
<b>3.1 Meßinstrumente</b>	29
<b>3.2 Musiksequenzen</b>	32
<b>3.3 Rater und Ablauf des Ratings</b>	32
<b>3.4 Ratertraining</b>	33
<b>3.5 Statistische Auswertungsverfahren</b>	34
3.5.1 Berechnung der Interraterreliabilität	34
3.5.2 Zusammenfassung von Items zu Skalen	35

	Seite	
<b>3.6</b>	<b>Beschreibung der einzelnen Sequenzen</b>	36
<b>4</b>	<b>Ergebnisse</b>	37
<b>4.1</b>	<b>Statistische Auswertung des Ratings</b>	37
4.1.1	Interraterreliabilität für die einzelnen Sequenzen	37
4.1.2	Skalenkonstruktion	38
4.1.3	Vergleich der Sequenzen	39
4.1.4	Interraterreliabilität der einzelnen Items	41
4.1.5	Interraterreliabilität der Adjektivliste zur Interaktion	44
4.1.6	Auswertung der Skalen zur Beurteilung der musikalischen Parameter	45
4.1.7	Auswertung der Skala zur Beurteilung der Differenziertheit des Beziehungsspiels	46
<b>4.2</b>	<b>Beschreibung der Sequenzen 3-12</b>	48
<b>4.3</b>	<b>Zusammenfassende Auswertung</b>	60
4.3.1	Sequenzlänge und Videoeinstellung	60
4.3.2	Probleme bei der Bewertung des Patientinnenspiels	60
4.3.3	Unterschiede zwischen den Ratern	61
<b>5</b>	<b>Diskussion</b>	63
<b>5.1</b>	<b>Diskussion der Ergebnisse</b>	63
5.1.1	Vergleich mit anderen Ratingskalen	63
5.1.2	Beurteilung der Interraterreliabilität	64
5.2	Vorschläge für weitere Untersuchungen	66
<b>6</b>	<b>Zusammenfassung</b>	69
	<b>Literaturverzeichnis</b>	71

	Seite
<b>Anhang</b>	77
Anhang A: Bereiche und Skalen der EWL	77
Anhang B: Skalen aus der Arbeit von Inselmann und Mann	78
Anhang C: Itemanalyse für die Skalen aus der Arbeit von Inselmann und Mann	79
Anhang D: Itemanalyse für die veränderten Skalen	80
Anhang E: Instrumente und Dauer der Sequenzen 3-12	81
Anhang F: Interraterreliabilität der Skalen	81
Anhang G: Semantische Beschreibung der Adjektive der Adjektivlisten zur Beschreibung des Patienten- und Therapeutenspiels	82
Anhang H: Semantische Beschreibung der Adjektive der Adjektivliste zur Beschreibung der Interaktion	84
Anhang I: Vergleich verschiedener statistischer Verfahren zur Berechnung der Interraterreliabilität	85
Anhang J: Interraterreliabilität der Items in der zweiten und dritten Ratingphase	86
Anhang K: Fiktives Beispiel zur Verdeutlichung der Unterschiede zwischen der Interraterreliabilität für Sequenzen und Items	87
Anhang L: Interraterreliabilität der Sequenzen auf Item- und Skalenebene	88

# **1 Einleitung**

## **1.1 Überblick**

Die vorliegende Arbeit beschäftigt sich mit der Möglichkeit, frei improvisierte Musik einer Einzelmusiktherapie mit Hilfe von Ratingskalen abzubilden. Hierbei wird v.a. auf die Interraterreliabilität, d.h. die Übereinstimmung von verschiedenen Raterurteilen, eingegangen, da diese bisher in vielen Studien, trotz ihrer Bedeutung als Objektivitätskriterium, vernachlässigt wurde. Es wird versucht, durch ausführliche Beschreibung der Musiksequenzen und durch Vergleich dieser mit den Ratingergebnissen herauszufinden, welche Faktoren Einfluß auf die Interraterreliabilität haben können. Diese Ergebnisse sollen dazu dienen, Verbesserungsmöglichkeiten für spätere Studien aufzuzeigen.

Im ersten Kapitel wird zunächst ein historischer Überblick über die Verwendung von Musik zu therapeutischen Zwecken gegeben, um darauf aufbauend den heutigen Stand der Musiktherapie zu beschreiben. Dabei wird auf verschiedene Theorien zur Musiktherapie eingegangen. Im zweiten Kapitel wird der aktuelle Stand der Musiktherapie-Forschung dargestellt. Hierbei liegt der Schwerpunkt auf Studien, die sich wie die vorliegende Arbeit mit Ratingskalen beschäftigen. Im dritten Kapitel wird die Methodik der vorliegenden Arbeit ausführlich beschrieben. Das vierte Kapitel umfaßt zwei Hauptteile: zum einen die statistische Auswertung des Ratings, zum anderen die ausführliche Beschreibung der gerateten Musiksequenzen. Im fünften Kapitel werden die Ergebnisse diskutiert, mit anderen Studien verglichen und Vorschläge für weitere Studien dargestellt. Abschließend werden die wichtigsten Aspekte der vorliegenden Arbeit im sechsten Kapitel zusammengefaßt.

## **1.2 Historischer Rückblick**

Um die zahlreichen Formen der heutigen Musiktherapie und die verschiedenen Ansätze der Musiktherapieforschung zu verstehen, ist ein kurzer Überblick über die Geschichte der Musikanschauung und der Verwendung von Musik zu therapeutischen Zwecken hilfreich.

Über den Ursprung der Musik kann man nur spekulieren. C. Darwin schrieb hierzu: „Da wir allen Grund zu vermuten haben, dass die artikuliert Sprache, wie sie sicher die höchste ist, eine der am spätesten vom Menschen erlangte Künste ist, und da das instinctive Vermögen, musikalische Töne und Rhythmen zu producieren, in der Thierreihe sehr weit hinab entwickelt ist, so wäre es durchaus mit dem Principe der Entwicklung im Widerspruch, wenn

wir annehmen sollten, dass die musikalische Fähigkeit des Menschen sich von den in der leidenschaftslosen Rede benutzten Tönen aus entwickelt hätte“ (Darwin 1875, S.315). Betrachtet man Musik als phylogenetisch (und auch ontogenetisch) frühe Kommunikationsform, ist es nicht verwunderlich, daß sie bereits in den ältesten uns bekannten Therapieformen und in allen Kulturen eine wichtige, oft sogar zentrale Rolle gespielt hat und z.T. noch bzw. wieder spielt. Folgende Textstellen gelten als die ältesten schriftlich überlieferten Beschreibungen von Musik zu therapeutischen Zwecken: „Und es geschah, wenn der Geist von Gott über Saul kam, nahm David die Zither und spielte (darauf) mit seiner Hand. Und Saul fand Erleichterung, und es ging ihm besser, und der böse Geist wich von ihm“ (Samuel 16,23). In der Odyssee schrieb Homer: „Und sie verbanden die Wunde kunstgerecht und stillten, Beschwörungen singend, das schwarze Blut“ (Homer, 19.456-458).

### 1.2.1 Musik und Magie/Ekstase

Musik war bereits bei den Naturvölkern ein sehr wichtiger Bestandteil im Heilungsritual. Möller schreibt hierzu: „Zu dem Ritual gehört die Heilmusik: die Heilgesänge des Medizinmannes, der Rhythmus der Rasseln und Klappern, der Klang zahlreicher Glöckchen und Metallstäbchen am Kostüm des Medizinmannes. Die Beschwörungszereemonien werden obendrein begleitet durch Musik und Tanz des gesamten Stammes.“ „Die konstante Wiederholung enger Intervalle und kurzer Motive, die eigensinnige Rückkehr auf einen bestimmten Ausgangston, die Monotonie der Tanzbewegungen... bewirken eine euphorische Bewußtseinseinengung“ (Möller 1974, S. 57-59).

Timmermann bezeichnet die Schamanen als die ersten Künstler und beschreibt ihre Tätigkeit folgendermaßen: „Heilen ist in dieser Phase nicht vom künstlerischen Akt zu trennen: Mythologie bzw. Religion, Kunst und Heilung bilden eine Einheit... Kunst existiert nicht als Selbstzweck,... sondern dient der Psychohygiene und Heilung“ (Timmermann 1999). Ziel sei es, die Harmonie zwischen Individuum, Kollektiv und Kosmos wiederherzustellen. Im Vergleich zu heutigen Vorstellungen von Gesundheit meint Timmermann: „Wenn wir den Menschen als biopsychosoziales Wesen und – ökologisch und spirituell – als Teil seiner Mitwelt betrachten, kommt diese ursprüngliche Definition von Gesundheit... der Sache doch sehr nahe.“ Trotz aller Unterschiede gelingt es ihm, weitere Ähnlichkeiten zwischen schamanistischen und einigen modernen musiktherapeutischen Therapieformen herauszuarbeiten, z.B. bzgl. des Settings, der Instrumente und der Persönlichkeit des „Heilers“.

Aber auch außerhalb dieser schamanistischen Kulturen finden sich magische Vorstellungen bei der medizinischen Anwendung von Musik, so auch im europäischen Kulturkreis von der Antike bis in die Neuzeit hinein, z.B. bei vielen Hebammen des Mittelalters: „Die Hebammen kannten neben zahlreichen Drogen und magischen Praktiken auch Zaubersprüche und Zaubergesänge, mit deren Hilfe sie versuchten, die Wehen einzuleiten, die Schmerzen zu lindern und bei Schweregebärenden die Geburt zu beschleunigen“ (Kümmel 1977, S.346). Weiter schreibt Kümmel, daß „obwohl studierte Ärzte einen magischen Gebrauch der Musik im allgemeinen ablehnten, ihn zumindest nicht empfahlen,... ließen sich Zaubersprüche, Zaubergesänge und magischer Klang bis weit in die Neuzeit nicht verdrängen.“ „Im 15. und 16. Jahrhundert scheint es eine „alltägliche“ Sache gewesen zu sein, bei Fieber und Verwundungen Zaubergesänge anzuwenden“ (Kümmel 1977, S. 215).

### 1.2.2 Musik und Harmonie

Mit der seit dem 7. Jh. v. Chr. durch die Vorsokratiker begründeten neuen Naturphilosophie entstand auch eine neue Sicht der Musik. Von besonderer Bedeutung ist hier Pythagoras, der sich u.a. intensiv mit Zahlen und Musik beschäftigte. So fand er heraus, daß man aus den ersten vier natürlichen Zahlen die harmonischen Verhältnisse der Quart, der Quint und der Oktave konstruieren kann. Diese Vorstellung von Harmonie wurde bei den Pythagoreern zum Ordnungsprinzip der Welt. „Von hier aus wird dann die Ausweitung des Musikbegriffs zu einer ‚musica mundi‘ verständlich, die gesamte Natur ist als Weltenmusik zu begreifen“ (Möller 1974, S.69-70). Aber auch der menschliche Körper unterlag diesem Ordnungsprinzip, so daß der Begriff der „musica humana“ entstand. In der hippokratischen Medizin wurde daher Gesundheit als Harmonie zwischen den vier Körpersäften (Synkresie), Krankheit hingegen als Disharmonie (Dyskresie) bzw. als eine Verstimmung des Körpers oder der Seele verstanden. Musik kann im Rahmen dieses Konzeptes eine harmonisierende Wirkung haben, d.h. sie kann Ordnung bewahren bzw. wieder herstellen. Bei den Pythagoreern wurde daher auch Musik in diesem Sinne verwandt: „Auch war er [Pythagoras] der Auffassung, die Musik trage wesentlich zur Gesundheit bei, wenn man sie in der rechten Weise betreibe... Dabei gab es bestimmte Melodien, die auf die Affekte – etwa auf Anfälle von Mutlosigkeit und nagendem Kummer - zugeschnitten waren... Andere wieder wirkten auf Zornes- und Gemütswallungen und auf jedes Außersichgeraten der Seele, die dafür anfällig ist“ (Jamblichos § 110, S.117). Ähnliche Auffassungen finden sich auch bei Platon: „...und der Einklang, welcher den Bewegungen unserer Seele verwandte Schwingungen in sich schließt, ist demjenigen, welcher vernünftig und nicht zu zweckloser Lust,... sich den Musen hingibt,



von ihnen zum Beistand verliehen, den in uns entstandenen unregelmäßigen Umlauf der Seele zu ordnen und mit sich selbst in Einklang zu bringen. Ebenso verliehen uns dieselben auch das Taktmaß, damit es die in uns in den meisten Fällen stattfindende maßlose und anmutsleere Gemütsstimmung ordnen und bekämpfen helfe“ (Platon, Timaios 47 d-e).

Aus der hippokratischen Humoralmedizin und der pythagoreisch-platonischen Philosophie entstand zunächst im arabischen Kulturkreis, dann im lateinischen Mittelalter ein komplexes System, in dem die Körpersäfte nicht nur bestimmten Charakteren, Elementen, Lebensabschnitten, Himmelsrichtungen, Sternenkonstellationen etc. zugeordnet wurden, sondern auch bestimmten Instrumenten, Melodien, Tönen und Intervallen. Wie in dem folgenden Zitat von dem Gelehrten Johann Knapp aus dem Jahre 1513 n. Chr. deutlich wird, ging man von einer engen Verbindung zwischen Seele und Körper aus, so daß Musik nicht nur bei Erkrankungen des Gemüts eingesetzt wurde, sondern auch bei körperlichen Beschwerden: “Um den verwirrten und kranken Sinnen die wilden und stürmischen Affekte zu nehmen, ist nichts wirksamer und mächtiger als musikalische Klänge. Sie erweisen sich auch oft an kranken Körpern als wirksam, entweder deshalb, weil Körper und Geist demselben Gesetz unterworfen sind, so daß bei Ihnen alles, ob gut oder schlecht, ob sie es wollen oder nicht, wechselseitig zusammenhängt, oder aber richtiger deshalb, weil wir selbst in einer der Musik ähnlichen Weise strukturiert sind...” (zit. n. Kümmel 1977, S.139). Eine ähnliche Auffassung findet sich bereits bei dem arabischen Gelehrten al-Farabi: „Der Körper ist krank, wenn die Seele geschwächt ist, und er ist beeinträchtigt, wenn sie beeinträchtigt ist. Daher geschieht die Heilung des Körpers durch die Heilung der Seele, indem ihre Kräfte wiederhergestellt und ihre Substanz in die rechte Ordnung gebracht wird mit Hilfe von Klängen, die dies bewirken können und die dafür geeignet sind“ (zit. n. Kümmel 1977, S.220). Die praktische Anwendung reichte von dem einfachen ästhetischen Genuß zur Steigerung des Wohlbefindens bis hin zum differenzierten Einsatz der Musik unter Berücksichtigung der Erkrankung, des Charakters des Erkrankten, der Tageszeit, der Sternenkonstellation und vielem anderen mehr. Ab dem ausgehenden 18. Jahrhundert wurden die therapeutischen Möglichkeiten der Musik zunehmend kritischer betrachtet und verlagerten sich zunehmend in den psychiatrischen Bereich. Aus den Jahren 1824/25 sind erste empirische Versuche des französischen Psychiaters Esquirol an Patienten mit psychischen Erkrankungen bekannt, aus denen er schloß, daß die Musik zwar nicht ohne Einfluß sei, eine Heilung oder Besserung der Erkrankung jedoch nicht erreicht werden könnte (Kümmel 1977, S. 276). Mit der Ablösung der Humoralpathologie durch die Zellulärpathologie im 19. Jahrhundert verlor auch die Musik zunehmend ihre Bedeutung in der Medizin.

Aber auch im 20. Jh. gibt es einzelne Autoren, die diese Tradition weiterverfolgen bzw. wiederbeleben wollen. So wendet H. Kayser in seinem Buch „Akroasis – Die Lehre von der Harmonik der Welt“ die Gesetze der Harmonik auf biologische und physikalische Phänomene an und schreibt vom „geheimnisvollen Urgrund tönender Gestalten, welche die Welt im Innersten ordnen und formen.“ (Kayser 1984, S.109). Kaysers Ansichten fanden v.a. über seinen Schüler R. Haase und den Psychotherapeuten A. Pontvik Eingang in die moderne Musiktherapie (Timmermann 1999).

Als besondere Thematik im Rahmen der Theorie einer „musica humana“ soll hier noch kurz der Zusammenhang, der zwischen dem menschlichen Puls und der Musik postuliert wurde, angesprochen werden (Kümmel 1977). Erstmals wandte der Arzt Herophilos um 300 v. Chr. metrisch-musikalische Modelle zur Beschreibung des Pulses an, um eine Pulsdiagnostik und – prognostik aufzubauen. Dieses Konzept wurde bis in die Neuzeit von vielen Gelehrten aufgenommen und modifiziert, dabei wurde insbesondere die Auffassung vertreten, „daß die den musikalischen Intervallen zugrundeliegenden Proportionen auf die Zeitverhältnisse innerhalb des Pulsschlages angewendet werden sollten“ (ebd., S. 77-78). Der Tübinger Medizinprofessor Samuel Hafenreffer (1587-1660) bezeichnete den Puls als süßeste Lebensmelodie und setzte bestimmte Pulsarten mit den vier Stimmlagen, den vier Hauptintervallen und den vier Elementen in Verbindung (ebd., S.38).

### 1.2.3 Musik und Iatrophysik

Insbesondere im 17. Jahrhundert versuchten einige Wissenschaftler – der berühmteste ist wohl René Descartes – physiologische Vorgänge im menschlichen Körper ganz mit mechanischen Gesetzen zu beschreiben. Dementsprechend wurde auch die Musikwirkung mechanisch erklärt: „Die Luftschwingungen werden also auf die Fasern oder Flüssigkeiten des menschlichen Körpers übertragen durch Resonanzvorgänge und erzeugen je nach Art der Musik, somit je nach Qualität und Quantität der Luftschwingungen, Veränderungen in der Zirkulation der Körpersäfte oder im Spannungszustand der Fasern“ (Möller 1974, S.140). Humoralmedizinische Vorstellungen spielten hier weiterhin eine wichtige Rolle, wobei die Musik durch ihre Schwingungen in der Lage sei, die Körpersäfte zu bewegen und zu mischen. Daneben gab es auch die Vorstellung, daß Nervenfasern wie Instrumentensaiten gespannt seien und durch Musik in Schwingungen versetzt werden könnten. Verschiedene Affekte, die als psychophysisches Korrelat zu diesen physiologischen Vorgängen betrachtet wurden, versuchte man durch bestimmte Musik zu beeinflussen. Darüber hinaus gab es aber auch den

Glauben, alle körperlichen Erkrankungen in Zukunft - mit einem größeren Wissen über den Zusammenhang zwischen Musik und seiner physiologischen Wirkung im Körper - gezielt mit Musik heilen zu können.

## **1.3 Musiktherapie heute**

### 1.3.1 Formen der Musiktherapie

Es gibt heute eine Vielzahl verschiedener musiktherapeutischer Verfahren. „Die Vielfalt und Uneinheitlichkeit musikalischer Theorien, Methoden und Anwendungsgebiete spiegelt sich auch in den einzelnen musiktherapeutischen Institutionen, Schulen und Ausbildungsgängen wider, die teilweise verschiedene Richtungen vertreten und sich bevorzugt auf ein bestimmtes Forschungs- und Anwendungsgebiet spezialisieren“ (Tischler 1983).

Man kann die verschiedenen Verfahren nach folgenden Gesichtspunkten grob einteilen:

- nach Anzahl der Patienten: - Einzeltherapie
  - Gruppentherapie
- nach Art des Musizierens: - passiv-rezeptiv (Musik anhören, z.B. als Entspannungsverfahren oder zur Förderung von Assoziationen)
  - aktiv-reproduktiv (Rezitation von Kompositionen)
  - aktiv-produktiv (freie Improvisation)
- nach dem zugrundeliegenden psychotherapeutischen Verfahren:
  - Psychoanalyse
  - Verhaltenstherapie
  - humanistische Psychologie

Bezüglich der Indikation für Musiktherapie gibt es noch keine systematischen Untersuchungen. Allerdings liegen Veröffentlichungen v.a. zu folgenden Erkrankungen vor: psychosomatische Erkrankungen, Borderline, Schizophrenie, Autismus und Eßstörungen (Timmermann et al. 1991). Timmermann bemerkt, daß Musiktherapie auch mit Patienten betrieben werden kann, die sprachlich nicht oder nur schwer erreichbar sind, z.B. Kinder, geistig Behinderte, Autisten oder Mutisten (Timmermann 1990, S.5). Inselmann beschreibt die Arbeit mit psychotischen Patienten und zeigt, wie diese Patienten, die auf sprachlicher Ebene schwere Kommunikationsstörungen aufweisen, über die Musik in die Lage versetzt werden, mit ihrer Umwelt zu kommunizieren (Inselmann 1995). Nach Thaut seien Ansätze der Musiktherapie im europäischen Raum überwiegend auf Psychotherapie bezogen, während man in den USA und Japan ein weit gestreutes Anwendungsgebiet findet, z.B. auch in der

neurologischen Rehabilitation, der Geriatrie oder bei geistig und körperlich Behinderten (Thaut 1997).

### 1.3.2 Theoretische Grundlagen

Bezüglich der Funktion der Musik kann man unterscheiden, ob der Musik selbst eine Heilkraft zugeschrieben wird (siehe Kapitel 1.2.2) oder ob die Musik lediglich als Medium aufgefaßt wird. Mahns bemerkt hierzu: „Die den Definitionen von Musiktherapie folgenden Erklärungen betonen zumeist, daß Musik keine Heilkraft in sich selbst besitzt. Die therapeutischen Effekte seien vielmehr die logische Folge einer professionellen, methodologisch-systematisch begründeten Anwendung von Musik. Die Hervorhebung von ‚Musik als Medium‘ zeigt deutlich, daß moderne Ansätze von Musik in der Therapie nicht länger auf der idealistischen Idee von den ‚heilenden Kräften der Musik‘ beruhen“ (Mahns und Ruud 1991, S.19). Ferner meint er, daß die in dem pythagoreischen Denken wurzelnde Vorstellung von den der Musik innewohnenden harmonisierenden Kräften noch weit verbreitet sei. Er fordert, daß die Musiktherapie, wenn sie als akademische Disziplin ernstgenommen werden will, aus wissenschaftstheoretischen Gründen Abschied nehmen sollte von derartigen Theorien, da sie allein auf metaphysischem oder idealistischem Gedankengut gründen. Er fordert statt dessen, sich an bewährte Therapiemodelle (z.B. psychodynamisches oder lerntheoretisches Modell) zu halten, die auf rationalen Prinzipien basieren würden. Nitzschke warnt hingegen vor zu leichtfertiger Übernahme dem musikalischen Erleben nicht originärer Interpretationsmuster, sondern fordert dazu auf, zunächst danach zu fragen, was das Spezifische der Musiktherapie ist. Hier nennt er v.a. den engen Zusammenhang von Musik und Emotionen und die Möglichkeit zur kathartischen Befreiung der Gefühle durch Musik (Nitzschke 1984). Ähnlich beschreibt auch Priestley das Spezifische der Musiktherapie: „Die Musiktherapie erlaubt und fördert gerade das, was den Analytiker bremst: das Ausagieren von Gefühlen“ (Priestley 1982, S. 5). Thaut, der wie viele andere Autoren ein Theoriedefizit für die Musiktherapie feststellt, kritisiert ebenfalls einen „Theorietransfer“ von verschiedenen psychotherapeutischen Ansätzen auf die Musiktherapie, da dadurch das Verstehen der Wirkungsweise der Musik im therapeutischen Prozeß vernachlässigt und zudem die Verständigung zwischen Therapeuten verschiedener Schulen erschwert werde, was wiederum die Etablierung von Musiktherapie als eigenständige Therapieform behindere (Thaut 1997). Sein Verständnis von Musik als therapeutisches Medium basiert darauf, daß „ästhetische Medien eine verstärkende, erleichternde, verdeutlichende Funktion in der Wahrnehmung und Analyse nicht-ästhetischer Objekte

ausüben können, die zum Verstehen, Erfahren und Erleben von Verhaltenssituationen in unserer Umwelt und zur Entwicklung von angemessenen Reaktionen beitragen.“

Im Folgenden möchte ich noch auf zwei für die Musiktherapie wichtige Aspekte eingehen, die für diese Arbeit von Bedeutung sind: Musik als Form von Kommunikation und als Ausdruck von Unbewußtem.

### 1.3.2.1 Musik und Kommunikation

Watzlawick vertritt die These, daß verbale Kommunikation aus zwei Anteilen besteht, aus einem digitalen und einem analogen (Watzlawick 1996). Unter digitaler Kommunikation versteht er das, was man durch Worte direkt sagt, also das, was man auch durch eine geschriebene Nachricht ausdrücken kann. Dabei beruhen die Bedeutungen der Wörter auf Konvention, d.h. es gibt nur ein semantisches Übereinkommen für die Beziehung zwischen Wort und Objekt. Digitale Sprache besitzt eine logische Syntax und ist in der Lage, komplexe und abstrakte Inhalte zu vermitteln. Nur der Mensch ist zur digitalen Kommunikation fähig.

„In der analogen Kommunikation dagegen finden wir etwas besonders Dingartiges in dem zur Kennzeichnung des Dings verwendeten Ausdrucks“ (ebd., S.62). Analoge Kommunikation findet paralingual über Gestik und Mimik und darüber, *wie* man etwas sagt (Ausdruck, Betonung etc.), statt. Im Gegensatz zu digitalen können analoge Informationen (Gestik, Mimik) von jedem - über alle Sprachgrenzen hinweg - zumindest teilweise verstanden werden. Watzlawick schließt hieraus, daß analoge Kommunikation ihre Wurzeln in viel archaischeren Entwicklungsperioden hat und daher eine weitaus allgemeinere Gültigkeit besitzt als die viel jüngere und abstraktere digitale Kommunikationsweise. Ebenso ist die Kommunikation zwischen Mutter und Kind zunächst eine rein analoge, bis das Kind lernt, sich auch in digitaler Form zu äußern.

In der menschlichen erwachsenen Kommunikation finden beide Kommunikationsformen Anwendung, wobei die digitale Form v.a. für Aussagen auf der Inhaltsebene, die analoge Form auf der Beziehungsebene von Bedeutung ist. „Überall, wo die Beziehung zum zentralen Thema der Kommunikation wird, erweist sich die digitale Kommunikation als fast bedeutungslos“ (ebd., S.64). „Jede sprachliche (also digitale) Auseinandersetzung über eine menschliche Beziehung ist deshalb überaus schwierig, weil sie eine Digitalisierung praktisch rein analoger Phänomene erfordert“ (ebd., S.67).

Es ist offensichtlich, daß zwischen der analogen Kommunikation und der Musik ein enger Zusammenhang besteht, weswegen man auch zur Beschreibung des analogen Anteils der Sprache Wörter verwendet wie Sprachrhythmus, Wortklang, Tonfall oder Satzmelodie.

Mit Musik kann man oft Emotionen weitaus besser ausdrücken als mit (digitaler) Sprache. Ferner wird diese emotionale Botschaft von fast jedem instinktiv relativ gut verstanden. Musik kann somit zur Kommunikation auf emotionaler Ebene und somit auch auf der Beziehungsebene dienen. C. Darwin (1875) sieht genau hier den Ursprung der Musik, nämlich im Balzgesang des männlichen Tieres. Ebenso kann man in der präverbalen (analogen) Kommunikation zwischen Mutter und Kind musikalische Elemente finden. Nitzschke schreibt hierzu: „Ein Indiz für den emotionalen Kommunikationsvorgang zwischen dem Kind und der Mutter sind die lautlichen Äußerungen, die beide austauschen. Es handelt sich dabei zunächst um eine Art primitiven Sprechgesangs... Die Sprechrhythmen von Mutter und Kind werden offenbar durch ein ‚Programm‘ gesteuert, das letztlich der Entwicklung des Kindes förderlich ist“ (Nitzschke 1984, S.177).

Für die Musiktherapie bedeutet dies, daß es möglich ist, mit dem Patienten durch die Musik eine non-verbale Kommunikation direkt auf Beziehungs- bzw. emotionaler Ebene zu führen, ohne daß ein „Umweg“ über digitale Kommunikation, die auf dieser Ebene oft unbefriedigend bzw. umständlich ist, notwendig ist. Es läßt sich ferner vermuten, daß man mit Musiktherapie besser als mit dem gesprochenen Wort problematische oder fehlende frühe Erfahrungen korrigieren bzw. ermöglichen kann (Timmermann 1990). Timmermann meint, daß sich die mit den frühen Bezugspersonen eingeübten repetitiven Beziehungsmuster des Patienten in der musikalischen Beziehungssituation wiederholen müßten. Diese müsse der Therapeut erkennen und dem Patienten bewußt machen.

Der vorliegenden Arbeit liegt die Theorie zugrunde, daß Musik Ausdruck von Emotionen und zugleich Kommunikationsmittel ist. Es wird versucht, mit Fragebögen den emotionalen Ausdruck und die Interaktion zwischen Therapeut und Patient meßbar zu machen. Die Theorie von digitaler und analoger Kommunikation verdeutlicht aber auch eine große Schwierigkeit eines solchen Forschungsansatzes: analoge Informationen (Musik) müssen umgesetzt werden in Sprache, d.h. in meßbare, digitale Informationen, um statistisch verarbeitbares Material zu erhalten. Man muß sich hierbei jedoch bewußt sein, daß bei einer solchen Übersetzung unvermeidbar Informationen verlorengehen.

### 1.3.2.2 Musik und das Unbewußte

In der klassischen psychoanalytischen Literatur findet die Musik kaum Erwähnung. Freud schrieb, daß er bzgl. der Musik fast genußunfähig sei. „Eine rationalistische oder vielleicht analytische Anlage sträubt sich in mir dagegen, daß ich ergriffen sein und dabei nicht wissen soll, warum ich es bin und was mich ergreift“ (Freud 1914, S.172). Hier wird nochmals die innige Beziehung der Musik zum emotionalen Erleben, aber auch zum Unbewußten deutlich. In der Psychoanalyse wird Musik – wie auch andere Kunstformen - meist als Ausdruck eines Transformationsvorganges von Trieben und Wünschen angesehen, die ihren Ursprung im Unbewußten haben (Mahns 1991). Musik sei nach Mahns in der Lage, „tiefste innere Wünsche in symbolischer Form abzubilden. In dieser Weise werden sie vom Ich akzeptiert“ (Mahns 1991, S.54). Dabei gelangen die gleichen Techniken wie im Traum zur Anwendung: Verdichtung, Verdrängung, Umkehr, Auslassung etc. (Mahns 1991). So könne in der Musiktherapie unbewußtes Konfliktmaterial durch Musik aktiviert und ausgedrückt und somit bewußt gemacht werden. Inselmann und Mann (1999) vertreten die Meinung, daß das Unbewußte gleich einer Sprache aufgebaut sei und somit Symbolisierungsfähigkeit besitze. Somit „sind musikalische Improvisationen ebenfalls als dramatische Symbolerzählungen das ins Musikspiel gesetzte unbewußte Denken und somit Ausdruck des emotionalen Erlebens“. Ferner spiegele sich in der aktiven Musiktherapie das unbewußte, emotionale Erleben in der Beziehung zum anderen. „Das bedeutet, daß die Musikimprovisation selbst die Psychopathologie des Patienten und daher die diagnostischen Möglichkeiten und die Therapie spiegelt“ (Inselmann und Mann 1999).

In verschiedenen Arbeiten wurde mit Hilfe von Ratingskalen zur Beurteilung von Musiktherapie untersucht, ob es einen Zusammenhang zwischen der Psychopathologie eines Patienten und seinem musikalischen Ausdruck gibt. Diese werden in Kapitel 2.2 dargestellt. Den in dieser Arbeit verwendeten „Adjektivlisten zum emotionalen Erleben“ liegt die Theorie zugrunde, daß unbewußte Emotionen in der Musikimprovisation zum Ausdruck gelangen und vom Zuhörer bzw. Rater erkannt werden können.

## **2 Stand der Forschung**

### **2.1 Einleitung**

So zahlreich die Anwendungen von Musik zu therapeutischen Zwecken waren und sind, so verhältnismäßig wenig wissenschaftlich fundierte Untersuchungsergebnisse gibt es (Hesse 1974). Strobel und Hubmann schreiben: „Die Durchsicht der Literatur hat nur einige exakte empirische und experimentelle Studien erbracht. Vor allem im kontinentaleuropäischen Raum fällt, entsprechend der wissenschaftlichen Tradition, die Bevorzugung eher theoretischer Arbeiten auf“ (Strobel und Huppmann 1978, S.161). Ebenso meint Steinberg: „But a systematic examination of musical ‚language‘ in musical therapy, what the participants tell one another and how they do it, has never been attempted“ (Steinberg 1991). Timmermann betont, wie wichtig wissenschaftliche Untersuchungen seien, wenn die Musiktherapie als eigenständige Therapieform allgemein anerkannt und nicht lediglich als „medizinisches Adjuvanzverfahren“ angesehen werden soll. „Eine von uns angestrebte Aufwertung der Musiktherapie bzw. deren breite Anerkennung als psychotherapeutisches Verfahren bedarf einer Legitimation, wie sie derzeit nur durch eine wissenschaftliche Fundierung im Sinne einer Grundlagenforschung gegeben ist“ (Timmermann 1990, S.7). Ebenso fordert Tischler: „... die Musiktherapie muß sich – will sie auf der Suche nach ihrem noch keineswegs gefundenen und allseits anerkannten Selbstverständnis ernst genommen werden - der wissenschaftlichen Analyse stellen ... Solange sich die Musiktherapie im unangreifbaren, nicht falsifizierbaren Bereich bewegt, bringt sie sich selbst um ihre empirische Grundlage“ (Tischler 1983). Man müsse sich aber auch darüber bewußt sein, daß durch Quantifizierung qualitativer Vorgänge (Psyche) allein Erkenntnisgewinnung im therapeutischen Bereich begrenzt sei. Empirie könne daher nur eine Ergänzung sein zu hermeneutischen, verstehenden Verfahren (zit. n. Harm 1983).

Einige wichtige empirische Forschungsansätze sollen im folgenden kurz dargestellt werden.

#### **2.1.1 Wahrnehmungsphysiologie**

Die Reizaufnahme und –verarbeitung, die Fähigkeit Schallwellen zu registrieren, zu analysieren und wahrzunehmen, wird u.a. in der Physiologie untersucht. Wichtige Grundlagen legten hier der Anatom Alfonso de Corti und der Physiologe und Physiker Hermann Helmholtz im 19. Jhd (Ackerknecht 1992). Ist die Analyse einzelner Ton- und Klangwahrnehmungen noch relativ einfach, so ist die Untersuchung und Erklärung der



Wahrnehmung von komplexeren Phänomenen, wie z.B. Melodien und Akkorde, sehr kompliziert. Hesse schreibt hierzu, „daß die Eindrucksqualitäten nicht als Summe der konstituierten Elemente zu beschreiben sind, sondern darüber hinaus entstehen Eigenschaften, die dem Zusammenklang als Ganzes zukommen“ (Hesse 1974, S.167). Aus diesem Prinzip der Übersummation folgt also, daß man derartige Phänomene nicht analytisch aufgrund ihrer Einzelelemente vollständig erklären kann.

Ferner spielen bei der Musikwahrnehmung auch Erfahrung und Erwartung des Hörers eine entscheidende Rolle.

### 2.1.2 Psychosomatische Forschung

Ein weiteres großes Forschungsfeld ist die Untersuchung von Veränderungen von somatischen und vegetativen Parametern durch Musik. Bereits 1880 publizierte G. Dogiel seine Arbeit „Über den Einfluß der Musik auf den Blutkreislauf“, in der er Veränderungen von Blutdruck, Atemfrequenz und Muskeltonus untersuchte (vgl. Simon 1974). Inzwischen wurden derartige Experimente durch polygraphische Untersuchungen z.B. mit EKG und EEG erheblich verfeinert. Man geht dabei davon aus, daß psychische Vorgänge parallel mit physischen verlaufen und man somit durch Messungen vegetativer Reaktionen quantitative Aussagen über das psychische Erleben erhalten kann (Möller 1974). Man hoffte, v.a. für die rezeptive Musiktherapie interessante Erkenntnisse zu gewinnen. Die bisherigen Ergebnisse waren jedoch unbefriedigend. Die somatischen Reaktionen können „grundsätzlich nur auf die Intensität, nicht dagegen auf den Inhalt des ablaufenden Erlebnisvorgangs hinweisen“ (Hesse 1974, S.166). Kneutgen schreibt hierzu: „Trotz zum Teil beträchtlichen apparativen Aufwands sind die bisherigen Ergebnisse vor allem deshalb unbefriedigend, weil sie nur allgemeine Schlüsse zulassen und keinen gezielten Einsatz bestimmter Musikstücke in der Therapie ermöglichen“ (Kneutgen 1974, S.24).

Über den Einfluß aktiver Musiktherapie auf vegetative Funktionen gibt es bisher nur wenige Untersuchungen. Neugebauer untersuchte den Einfluß von gemeinsamer musikalischer Improvisation auf die Herzfrequenz und fand heraus, daß es einen eindeutigen Zusammenhang sowohl zwischen der Herzfrequenz von Therapeuten und verschiedenen Probanden, als auch zwischen den Frequenzen und dem musikalischen Geschehen gibt (Neugebauer 1998). Müller et al. (1994) konnten in einer Studie zeigen, daß die IgA-Konzentration im Speichel bei Patienten, die an einer Musiktherapie teilnahmen, nach der Therapie signifikant höher war als vorher. Es zeigte sich allerdings kein Zusammenhang mit psychologischen Parametern.

### 2.1.3 Beobachtung und Beschreibung von Musik

Obwohl die Methode der Beobachtung von einigen Wissenschaftlern als unwissenschaftlich angesehen wird, vertritt Bortz (Bortz 1984, S. 189ff) die Ansicht, „daß die Methode der systematischen Beobachtung zum Arsenal der empirischen Datenerhebungstechniken gehört, weil sie für bestimmte Fragestellungen den einzigen Zugang zu aussagekräftigen Daten darstellt.“ Man muß dabei natürlich beachten, „daß Beobachtung so gut wie nie einer realitätsgetreuen Abbildung des Beobachteten entspricht. Beobachten heißt gleichzeitig, Entscheidungen darüber zu treffen, was ins Zentrum der Aufmerksamkeit rücken soll und wie das Beobachtete zu interpretieren bzw. zu deuten ist“. „Eine Maßnahme, die geeignet ist, das Ausmaß an Subjektivität von Beobachtungen zu kontrollieren, ist der Einsatz mehrerer Beobachter, deren Protokolle verglichen und gegebenenfalls zu einem Gesamtprotokoll zusammengefaßt werden“ (ebd., S. 198). Solche durchschnittlichen Urteile sind reliabler und valider als Individualurteile. Dies sei jedoch nur zulässig, wenn die Übereinstimmung zwischen den Ratern groß genug sei. Die Raterübereinstimmung ist auch deswegen wichtig, da sie als Maß der Objektivität Voraussetzung für Reliabilität und Validität des Meßinstrumentes ist (Bortz 1984). Auf das Problem der Interraterreliabilität wird später noch ausführlich eingegangen.

Man muß sich aber auch der Grenzen eines solchen Verfahrens bewußt sein. „Seelische Erlebniswelten sind nicht nur komplexer als vegetative Vorgänge, sondern auch differenzierter als das an den Tag gelegte, von dem Beobachter registrierte Verhalten des Untersuchungssubjekts“ (Tischler 1983). Eine Verhaltensbeobachtung läßt keine zuverlässigen Aussagen über Intentionen oder Einstellungen des Beobachteten zu, hierüber könnte man nur durch eine zusätzliche Befragung Informationen gewinnen.

Im folgendem werden verschiedene Arbeiten vorgestellt, in denen versucht wird, Meßinstrumente in Form von Ratingskalen zu entwickeln, die verschiedene Aspekte von Musik abbilden sollen und zwar in einer Form, die eine statistische Auswertung der Daten ermöglicht.

## **2.2 Aktuelle Forschungsarbeiten**

### Steinberg 1985, 1991, 1992

Steinberg untersuchte zunächst den Einfluß der Schwere einer psychiatrischen Erkrankung auf das Tempo der Rezitation eines allgemein bekannten einfachen Volksliedes (Steinberg 1985a). Die Schwere der Erkrankung wurde u.a. mit Hilfe der „Brief Psychiatric Rating Scale“ (BPRS) (Overall und Gorham 1962) eingeschätzt. Für alle Patienten (n=53) wurde

zweimal das Tempo bestimmt, einmal zu einem Zeitpunkt mit hohem BPRS-Wert (U1) und einmal mit niedrigem Wert (U2). Trotz großer interindividueller Unterschiede zeigten sich weder Unterschiede im Durchschnittstempo zwischen Patienten und Kontrollgruppe (n=31), noch signifikante intraindividuelle Unterschiede zu den zwei Meßpunkten. Einzige Ausnahme stellte die Untergruppe der Patienten mit endogener Depression (n=13) dar, die bei schwererer Psychopathologie signifikant langsamer spielten (dies gilt jedoch nicht für Patienten mit neurotischer Depression oder Depression bei Schizophrenie).

In einer anderen Studie untersuchte Steinberg die Möglichkeit, Musik von psychiatrischen Patienten und Gesunden mit Hilfe einer Ratingskala abzubilden (Steinberg 1985b,c). Er stellte fest, daß es bisher nur Meßinstrumente zur Beurteilung professioneller Musik (er verweist auf Kleinen 1968 und Batel 1976) gäbe, nicht jedoch ein Meßinstrument, was auch Ausführungen („performances“) mittelmäßiger Musiker oder Amateure erfaßt. Er entwarf daher 12 bipolare, 7-stufig skalierte Items (semantische Differentiale, siehe Osgood 1952), die neben musikalischen Parametern, Merkmale wie Lebendigkeit, Expressivität oder logischen Aufbau messen sollen, wobei die beiden Pole je eine „gute“ und eine „schlechte“ Ausprägung eines Merkmales darstellen sollen. Aus den Items läßt sich ein Gesamtscore berechnen, der angibt, wie hoch die Ausdrucksqualität ist. Aus Audioaufzeichnungen von 61 Patienten und 29 Gesunden wurden kurze Ausschnitte von 20–60 sec Länge mit besonders guter musikalischer Ausführung ausgewählt. Es handelte sich dabei v.a. um die Wiedergabe von Klassik-, Jazz- und Popliedern. Zu jedem Patienten gibt es Aufzeichnungen von einem Zeitpunkt mit hohem (U1) und niedrigem (U2) BPRS-Wert. Geratet wurde von 3 Experten und 50 Musikern. Diese sollten v.a. die Ausführung der Musik und nicht die Musik selbst beurteilen. Steinberg kommt zu folgenden Ergebnissen:

- Die Interraterreliabilität zwischen den 3 Experten beträgt für die einzelnen Items zwischen 0.87 und 0.96 (statistisches Verfahren wurde nicht angegeben), für die anderen 50 Rater gibt es keine entsprechenden Angaben.
- Der Korrelationskoeffizient zwischen 1. Rating und einem 2. Rating nach einem Jahr liegt pro Item für die Ratergruppe der Musiker zwischen 0.54 und 0.83, für die Experten zwischen 0.79 und 0.93 (ermittelt aus den Mittelwerten von den 3 Experten).
- Die Mittelwerte von Experten und Musikern zeigen hohe Übereinstimmungen.
- Teilt man die Spieler nach ihren musikalischen Fähigkeiten in Gruppen ein, so differenzieren die meisten Items gut zwischen diesen.
- Patienten mit endogener Depression (n=8) zeigen signifikante Unterschiede zwischen U1 und U2 in dem Gesamtscore und den Merkmalen Stabilität und Gefühlsausdruck. Für

Patienten mit neurotischer Depression zeigen sich hingegen keine signifikanten Unterschiede zwischen U1 und U2.

- Bei der Untergruppe der Patienten mit Schizophrenie mit hohem BPRS-Wert bei U1 zeigen sich signifikante Unterschiede im Gesamtscore und in den Merkmalen „Gefühlsausdruck“, „Professionalität“ und „logischer Aufbau“ zwischen U1 und U2 (11 von 12 Items zeigten „Verbesserungen“).
- Bei Patienten mit maniformer Störung (n=11) besteht ein signifikanter Unterschied im Merkmal „Stabilität“ zwischen U1 und U2.
- Insgesamt zeigt sich ein hochsignifikanter Zusammenhang zwischen Gesamtscore im Musikrating und BPRS-Werten, dabei sind die Unterschiede im Gesamtscore zwischen U1 und U2 auch bei Patienten mit anfangs schwerer Psychopathologie eher gering (ca. 0.5 auf der 7-stufigen Skala), nur bei einem Patienten waren die musikalischen Fähigkeiten extrem eingeschränkt.

In einer späteren Untersuchung (Steinberg 1992a,b) wurden die Sequenzen nach relativ objektiven Kriterien wie Geschwindigkeitsänderungen, Abbrüche, Fehler im Rhythmus und Melodie (z.B. falsche Noten) ausgewertet. Im Gegensatz zu den semantischen Differentialen zum musikalischen Ausdruck zeigten sich keine oder nur geringe Zusammenhänge zwischen diesen Kriterien und den BPRS-Werten, der Musikfähigkeit oder dem Meßzeitpunkt. Allerdings zeigten sich hohe Korrelationen zwischen den objektiven Kriterien und den semantischen Differentialen.

In einer weiteren Studie (Steinberg 1991) benutzte Steinberg ein ähnliches Studiendesign, diesmal aber mit freier musikalischer Improvisation aus einer Musiktherapie anstatt der Wiedergabe von Kompositionen. Die Interraterreliabilität wird als hoch bezeichnet, es fehlen jedoch genauere Angaben. Die Korrelationskoeffizienten für die Testwiederholung sind ähnlich hoch wie in der vorherigen Studie. Die Auswertung der Untergruppen erbrachte u.a. einige schwer erklärbare Ergebnisse:

- Bei Patienten mit Schizophrenie zeigt sich in allen Items eine „Verschlechterung“ zwischen U1 und U2, diese ist signifikant für das Merkmal „Professionalität“.
- Bei Patienten mit neurotischer Depression zeigt sich zwischen U1 und U2 eine „Verschlechterung“ in 10 von 11 Items, das Spiel ist bei U2 signifikant unrhythmischer und zögerlicher.
- nur für die Gruppe der Patienten mit endogener Depression sind die Ergebnisse ähnlich wie oben. Insgesamt „verbesserte“ sich das Spiel, dies ist signifikant für die Merkmale „Lebendigkeit“ und „Professionalität“. Auch bei einer Untersuchung der Spielgeschwindigkeit ließen sich obige Ergebnisse nicht reproduzieren.

### Beurteilung:

Bemerkenswert ist die hohe Interraterreliabilität, die in anderen Studien nicht erreicht wird. Leider wird das statistische Verfahren nicht angegeben. Ein systematischer Vergleich von Interraterreliabilitäten und deren Beurteilung folgt in Kapitel 2.2.1.

Während die Ergebnisse bzgl. der Rezitation von Kompositionen gut interpretierbar sind, trifft dies nach Ansicht der Autoren für die freie Improvisation nicht zu. Auch ist bei letzterer die Korrelation zwischen Psychopathologie und Ratingergebnissen geringer. Dies könnte darauf hinweisen, daß dieses Meßinstrument für das Rating von freier Improvisation nicht ausreichend valide ist oder daß – wie Steinberg vermutet - die freie Improvisation die Psychopathologie schlechter spiegelt als die Wiedergabe einer Komposition. Um dies zu beurteilen, sollte man untersuchen, ob man mit anderen Meßinstrumenten, die speziell für die freie Improvisation entworfen werden und eine ähnlich hohe Objektivität erreichen, Ergebnisse erhält, die besser interpretierbar sind. Auch die Beurteilung der Merkmalsausprägungen als „schlechter“ und „besser“ ist für die freie Improvisation sehr fragwürdig (v.a. bei Adjektivpaaren wie „friedlich - aggressiv“ oder „seriös - verspielt, oberflächlich“). Erschwerend für die Bewertung des Meßinstrumentes ist, daß die verschiedenen Items teilweise verschiedene Ebenen der Musik bewerten. Während die meisten Items die Ausdrucksqualität messen sollen, beziehen sich einige Items auf musikalische Parameter (Melodie, Rhythmus) oder den emotionalen Gehalt der Musik („aggressiv – friedlich“).

Die hier untersuchten objektiveren Kriterien (z.B. Geschwindigkeit) zeigten im Gegensatz zu den (subjektiveren) semantischen Differentialen geringe oder keine Korrelationen zu den Patientendaten. Dieses könnte ein Hinweis darauf sein, daß Musikgeschehen zu komplex ist, um es hinreichend nur mit einfachen „objektiven“ Parametern zu beschreiben (siehe Kapitel 2.1.1). Die z.T. hohen Korrelationen zwischen „objektiven“ und „subjektiven“ Kriterien lassen es sinnvoll erscheinen, in weiteren Studien beide Kriterien zu berücksichtigen und deren Zusammenhang näher zu erforschen.

### Maler 1992

Maler beschreibt in seiner Arbeit das von ihm entwickelte Lübecker Musiktherapiemodell. Wesentlicher Teil dieses Modells ist eine produktive Gruppen-Musiktherapie, die sich immer in folgende Abschnitte gliedert: Vorgespräch, 1. Spielphase: Spannungsebene, 2. Spielphase: Schwingungs- und Entfaltungsebene, 3. Spielphase: Begegnungsebene, Nachgespräch. Für diese Therapieform hat Maler 8-stufige Ratingskalen entwickelt mit folgendem Inhalt:

Energie, Instrument (Spielmöglichkeiten des gewählten Instrumentes), Tonraum, Lautstärke, Zeitmaß (Langsamkeit des Spiels), Körper-Ich (Körpereinsatz beim Musizieren), Modulation, Stereotypie, Gestalt (plastisches Klangbild), Gruppenorientierung, Gefühlsausdruck (im Nachgespräch), Rationalisierung (im Gespräch).

Mit diesen Skalen soll das musikalische Verhalten möglichst präzise erfaßt werden, ohne es psychologisch zu deuten. Insgesamt wurden 200 Patienten über jeweils 10 Sitzungen beobachtet. Die größten Diagnose-Untergruppen waren: Colitis ulcerosa (n=27), Morbus Crohn (n=27), neurotische Depression (n=26), schizophrene Psychosen (n=25), Anorexia nervosa (n=21) und Bulimie (n=14).

Geratet wurde im Anschluß an die Sitzung durch den Musiktherapeuten selbst und durch eine Medizin-Doktorandin.

Die Interraterübereinstimmung wird als gut bezeichnet, es wird aber kein statistisches Ergebnis genannt. Dafür wird auf eine Voruntersuchung verwiesen, bei der eine mittlere Interraterreliabilität von  $r=0.78$  gefunden wurde (Balck et al. 1987).

Zu jeder Skala wurden für jede Sitzung aus den Werten aller Patienten die Mittelwerte berechnet. Hier zeigten sich in sämtlichen Skalen deutliche Änderungen über die Zeit, die alle signifikant waren, wobei die Skalenwerte für Stereotypie und Gruppenorientierung sanken, alle anderen stiegen an. Insgesamt wurde das Spiel differenzierter. Maler schließt hieraus, daß die Patienten im Laufe der Therapie sich besser und differenzierter wahrnehmen und dies zudem besser ausdrücken können.

Bei der Aufschlüsselung der Werte nach Diagnosegruppen zeigten sich nur geringe Unterschiede zwischen ihnen. Lediglich die Skalen „Energie“, „Gestalt“ und „Körper-Ich“ wiesen signifikant höhere Werte bei den Patienten mit Bulimie auf. Der Verfasser meint, daß der Unterschied möglicherweise deutlicher sei, wenn man nicht die Diagnose, sondern die Persönlichkeitsstruktur zum Vergleich nähme.

### Beurteilung

Wie Maler selbst bemerkt, kann man die deutlichen Veränderungen der Skalen im Verlauf der Therapie nicht mit Sicherheit allein auf die Musiktherapie zurückführen, da parallel auch andere Psychotherapieverfahren stattfanden. Auch müßte der hypothetische Zusammenhang zwischen differenzierterem Spiel und differenzierter Selbstwahrnehmung genauer untersucht werden, so könnte man die Änderung im Spiel auch mit einem Übungseffekt und einem sichereren Umgang mit den Instrumenten erklären. Ferner könnte bei diesem Studiendesign die Erwartungshaltung der Rater die Ergebnisse beeinflusst haben. Um diesen Fehler

auszuschließen, könnte man die Sitzungen aufzeichnen und in randomisierter Reihenfolge raten (siehe Vanger et al. 1995).

Leider wird die Interraterreliabilität dieser Untersuchung nicht angegeben. Diese wäre besonders interessant, da bei dieser - genauso wie bei der vorliegenden Arbeit - das Rating durch zwei Personen mit sehr unterschiedlichen Voraussetzungen durchgeführt wurde: Die eine Person ist Musiktherapeut und zudem selbst der Therapeut in den zu ratenden Sitzungen, während der andere Rater eher als Laie zu bezeichnen ist.

Da die Skalen speziell für diese Therapieformen entworfen wurden, können sie nicht ohne weiteres übernommen werden. Einzelne Skalen finden sich jedoch in ähnlicher Form auch in der quantitativen Analyse der vorliegenden Arbeit wieder: Tonumfang - Tonraum, Lautstärke, Tempo - Zeitmaß; ferner beinhalten Malers Skalen „Modulation“ und „Gestalt“ Aspekte, die in der vorliegenden Arbeit mit den Parametern „Melodie“ und „Rhythmus“ erfaßt werden.

#### Vanger et al. 1995

In dieser Einzelfallstudie einer Morbus Crohn Patientin sollte untersucht werden, ob man anhand einzelner Musikausschnitte, die mit Hilfe von Ratingskalen bewertet und statistisch ausgewertet werden, zu Ergebnissen kommen kann, die mit den Einschätzungen des Therapeuten über die Problematik des Patienten und den Therapieverlauf übereinstimmen.

Als Material dienten Audioaufzeichnungen von 25 Musiktherapie-Sitzungen und einem follow-up. Aus diesen wurden 4 Sitzungen verschiedener Zeitpunkte der Therapie ausgesucht und aus jeder dieser Sitzungen 3 einminütige Sequenzen ausgewählt. Diese 12 Sequenzen wurden randomisiert geordnet und von 4 Ratern ausgewertet.

Als Meßinstrument diente das sogenannte Musiktherapiekodierungssystem (MUSIKOS), welches 5-stufig skalierte semantische Differentiale umfaßt, welche Beziehungsaspekte der therapeutischen Dyade messen sollen (einige Items basieren auf den „styles of relating tables for therapist and patient“ (Orlinsky und Howard 1975)).

Aus den Ergebnissen der 4 Rater wurde der Mittelwert gebildet und eine Faktorenanalyse durchgeführt. Diese ergab die folgenden zwei Faktoren: Aktivität (activity) und Nähe (closeness). Es zeigte sich, daß der Faktor Aktivität signifikant zum Ende der einzelnen Therapiesitzungen abnahm und er im follow-up signifikant höher war als in den anderen Sitzungen. Die Autoren kommen schließlich zu folgendem Schluß: „When comparing the results obtained from experimental analysis of this particular music therapy with the clinical

reports provided by the therapist, there emerges a congruent picture about the patient and about the development of the therapeutic relationship.“

### Beurteilung:

Diese Arbeit ist deswegen interessant, weil hier (wie bei Steinberg) selektiv nur der auditive Kanal untersucht wurde. Kritisch ist jedoch zu bemerken, daß versucht wird, ein Meßinstrument zu validieren, ohne es vorher auf Objektivität zu prüfen. Auch eine Mittelwertbildung aus mehreren Raterurteilen, ohne diese vorher auf Übereinstimmung zu untersuchen, ist nicht zulässig (siehe Kap. 2.1.3 ).

### Pechr 1996

Pechr vergleicht zwei Meßinstrumente, zum einen ein Instrument zur Erfassung musikalischer Parameter, den in Heidelberg und Ulm entwickelten Musikerfassungsbogen (MUERB), zum anderen eine Ratingskala zur Erfassung psychologischer Dimensionen, das von der Studiengruppe Ulm/Stuttgart entwickelte Musiktherapiekodierungssystem (MUSIKOS).

Der MUERB umfaßt 14 bipolar formulierte Items, wobei die deskriptiven Adjektivpaare möglichst konkret die Bestandteile des musikalischen Geschehens darstellen sollen. Solche Paare sind z.B. „laut – leise“, „langsam – schnell“ oder „tonal - nicht tonal“.

Mit dem MUSIKOS soll durch 15 bipolar formulierte Items der emotionale Gehalt von Musik erfaßt werden, es handelt sich um eine um 5 Items erweiterte Form des bei Vanger (1995) vorgestellten Meßinstrumentes. Adjektivpaare sind z.B. „selbstsicher – selbstunsicher“, „untergeordnet – dominant“ oder „entschlossen – unentschlossen“.

Als Untersuchungsmaterial dienten Videoaufnahmen von 9 depressiven Patienten und 9 normalgesunden Personen, die durch matched-pairs Verfahren zugeordnet wurden. Bei der Musik handelte es sich um „experimentelle Therapiemusik“, also nicht um Ausschnitte aus einer Musiktherapie. Geratet wurde von StudentInnen des Fachbereichs Musiktherapie (Anzahl ist nicht angegeben).

Bei den Mittelwerten der einzelnen Items des MUERB ließen sich keine statistisch haltbaren Unterschiede zwischen Depressiven und Normalgesunden finden.

Im MUSIKOS hingegen unterschieden sich die beiden Gruppen in 6 Items signifikant voneinander und zwar wurde das Spiel der Depressiven als selbstunsicherer, zurückhaltender, starrer, bedächtiger, gefühlloser und eingengter bewertet.



Durch eine Faktorenanalyse ließen sich die Items des MUERB zu 4 Skalen zuordnen, die wie folgt benannt wurden: Ordnung, Aktivität, Kraft und Raum. Für das MUSIKOS ergaben sich die Skalen Tatkraft und Affiliation (siehe auch Vanger 1995).

Es ließen sich keine signifikanten Korrelationen zwischen den Einzelitems oder den Skalen der beiden Meßinstrumente finden, d.h., „daß die erwarteten Zusammenhänge zwischen den musikalischen Elementen und der Musik innewohnenden psychologischen Qualitäten nicht sichtbar werden“. Eine mögliche Erklärung hierfür könnte sein, daß bei dem Rating mit dem MUSIKOS zusätzlich zur Musik auch visuelle Eindrücke berücksichtigt wurden.

### Beurteilung:

Auch bei dieser Arbeit fehlt die für die Bewertung eines Meßinstrumentes wichtige Interraterreliabilität.

Wie bei Maler hat sich gezeigt, wie schwierig diagnostische Aussagen mit solchen Meßinstrumenten sind. Diese scheinen jedoch mit Fragebögen zum emotionalen Gehalt von Musik besser möglich, als durch Analyse von musikalischen Parametern (wie bereits die Arbeiten von Steinberg zeigten). Musik läßt sich anscheinend nicht vollständig auf einzelne musikalische Parameter reduzieren, auch wenn ein solcher Ansatz aus wissenschaftlicher Sicht einfacher wäre, da sich diese Parameter leichter objektivieren ließen, während Ratings zur Ausdrucksqualität oder zum emotionalen Gehalt stärker von der subjektiven Einschätzung der Rater abhängig sind. Die Erforschung des Verhältnisses zwischen diesen beiden Ansätzen steht erst am Anfang und bedarf noch weiterer wissenschaftlicher Untersuchungen.

In der vorliegenden Arbeit werden deswegen auch beide Ansätze verfolgt, wobei der Schwerpunkt aber auf den emotionalen Gehalt von Musik gelegt wird. Viele Items des MUSIKOS finden sich auch in den Adjektivlisten der vorliegenden Arbeit, z.B. die Items „selbstsicher“, „selbstunsicher“, „dominierend“, „passiv“, „starr“, „freundlich“ oder „zugewandt“.

### Moreau 1996

Ziel der Arbeit von Moreau war, die von ihr entwickelte Musiktherapeutische Ausdrucks- und Kommunikations-Stufenskala (MAKS) anhand der Testgütekriterien zu evaluieren. Die MAKS umfaßt zwei Skalen: Die Skala zum Ausdrucksverhalten besteht aus den Kategorien „Umgang mit dem Instrument“ (3 Items), „Formgebung, Gestalt“ (3 Items), „Vitalität, Ausdruckskraft“ (5 Items) und „Ausdrucksqualität“ (3 Items). Die Skala zum Kommunikations-

verhalten beinhaltet die Kategorien „allgemeines Engagement“ (5 Items), „Bezug zum Gegenüber“ (4 Items) und „Ausdrucksqualität“ (4 Items).

Jedes Item wurde 7-stufig skaliert und jede Stufe genau operationalisiert. Es werden Verhaltensmerkmale, formale und strukturelle Musikmerkmale sowie Eindrucksqualitäten erfaßt. Bewußt wird auf Beschreibung von Gefühlsinhalten (siehe Timmermann 1990) verzichtet.

Als Untersuchungsmaterial dienten Videoaufzeichnungen von experimentell praktizierter Musik von 12 Jugendlichen (Patienten einer Kinder- und Jugendpsychiatrie und Gesunde). Aus dem Gesamtmaterial wurden für das Rating 10 Solo- und 10 Duoszenen von je einer Länge von ca. 20 sec. herausgeschnitten, die jeweils einen anderen Verhaltensaspekt darstellten.

Die Ratergruppe bestand aus 54 MusiktherapeutInnen bzw. StudentInnen, die z.T. auch an einem Re-Rating nach 4-6 Wochen teilnahmen. Die Rater wurden nach ihrer Dauer der Beschäftigung mit Musiktherapie in 4 Gruppen eingeteilt. Vor jedem Rating fand ein Ratertraining mit je 3 Videoszenen statt.

Zur Beurteilung der Objektivität wird der Interraterkorrelationskoeffizient „tau“ nach Kendall berechnet. Der mittlere Interraterkorrelationskoeffizient liegt für die Ausdrucksskala bei  $\tau=0.4$  und für die Kommunikationsskala bei  $\tau=0.3$ . Die mittleren Raterübereinstimmungen für die einzelnen Szenen schwanken zwischen  $\tau=0.1$  und  $\tau=0.5$ , wobei die Übereinstimmungen für einzelne Raterpaare zwischen  $\tau=1.0$  und  $\tau=-0.8$  (!) liegen. Die mittleren Raterübereinstimmungen für die einzelnen Items schwanken zwischen  $\tau=0.1$  und  $\tau=0.7$ . Die Interraterkorrelationen sind für alle 4 Gruppen und zu beiden Meßzeitpunkten ungefähr gleich. Diese Ergebnisse entsprechen nach Moreau nicht meßtheoretischen Forderungen. Sie hofft, durch Überarbeitung einzelner Items und besserem Ratertraining diese Ergebnisse verbessern zu können.

Zur Beurteilung der Reliabilität wird die Übereinstimmung der einzelnen Rater zwischen den beiden Meßzeitpunkten mit Cohen's Reliabilitätskoeffizient „kappa“ (Cohen 1960) berechnet, zudem wird der prozentuale Anteil der absoluten und der relativen/guten (Abweichung +/- 1) Übereinstimmung angegeben. „Kappa“ lag für die einzelnen Items zwischen 0.21 und 0.42 (Durchschnitt:  $\kappa=0.27$ ). Die relativen Übereinstimmungen für die verschiedenen Items liegen zwischen 75% und 98%. Diese Ergebnisse werden von Moreau als sehr gut angesehen.

Zur Validitätsbestimmung wurden eine Varianz-, eine Konsistenz- und eine Faktorenanalyse durchgeführt. Die Varianzanalyse zeigte, daß die Items hochsignifikant zwischen den einzelnen Szenen differenzieren. Durch die Faktorenanalyse wurde festgestellt, daß es große

Übereinstimmungen mit Faktoren anderer Meßinstrumente gibt. Moreau meint aber, daß die Faktorenstruktur der MAKS noch differenzierter sei.

### Beurteilung:

Diese Arbeit ist besonders interessant, da sie sich intensiv mit den drei Testgütekriterien Objektivität, Reliabilität und Validität auseinandersetzt. Die mittlere Interraterkorrelation ist jedoch mit 0.3–0.4 recht niedrig. Gleiches gilt für den Reliabilitätskoeffizient „kappa“, dennoch wird die Reliabilität aufgrund der hohen absoluten und relativen Übereinstimmung als sehr gut bezeichnet. Hier wäre eine genauere Diskussion über die Anwendung verschiedener statistischer Methoden sinnvoll.

### Timmermann 1990

Diese Arbeit soll hier etwas ausführlicher vorgestellt werden, da sie wesentliche Grundlage für die vorliegende Arbeit ist.

In seiner Studie beschäftigt sich Timmermann mit der Konstruktion eines Versuchsdesigns mit Fragebögen und Ratingskala, um Fremdbeobachtung durch verschiedene Gruppen zu ermöglichen. Dabei stellt er die Hypothese auf, daß die affektiven Inhalte der Beziehungsmuster im musikalischen Dialog zwischen Patient und Therapeut, sowie die Psychopathologie und die Persönlichkeitsstruktur des Patienten auch von nicht informierten Beobachtern (Fremdbeobachtern) nur aufgrund einzelner musiktherapeutischer Szenen identifiziert werden können (Timmermann 1990, S. 68-69). Timmermann versteht diese Studie als Pilotstudie zur Methodenentwicklung und entwarf deswegen ein sehr umfangreiches Studiendesign, um herauszufinden, welche Ansätze sinnvolle Ergebnisse liefern könnten.

Als Material dienten Videoaufzeichnungen von 8 Musiktherapie-Sitzungen eines Patienten mit narzißtisch-schizoider Persönlichkeitsstörung. Zusätzlich schrieb der Therapeut nach jeder Sitzung ein *affektives Spontanprotokoll (a)* und der Verfasser (Timmermann) fertigte anhand der Videoaufzeichnungen ein ausführliches *Stundenprotokoll (b) an*. Darauf folgten noch eine *Zusammenfassung und eine Reflexion (c)* durch den Verfasser. Ferner wurden nach jeder Stunde von Therapeut und Patient spezielle *Fragebögen (d)* ausgefüllt. Als weitere Informationsquellen dienten Aussagen vom Patienten während einer parallel stattgefundenen *analytischen Behandlung (e)* und regelmäßiger *Konsultationsgespräche (f)*.

Aus dem Videomaterial wurden 8 kurze signifikante Szenen herausgesucht. Diese stellten sogenannte „repetitive Muster“ (d.h. sich wiederholende Spielweisen des Patienten, die

typisch für ihn sind) (Szenen 1-5) und sog. „pattern of change“ (d.h. Stellen, an denen eine Veränderung im Spiel auftritt) (Szenen 6-8) dar. Ferner sollten sie einen hohen Informationsgehalt besitzen und repräsentativ für Musiktherapie sein. Die Szenen hatten eine Länge zwischen 0:40 min und 1:55 min.

Die Auswertung dieser Szenen erfolgte durch 3 Ratergruppen: Musiktherapeuten (n=20), Psychotherapeuten (n=10) und Laien (n=20). Es sollte untersucht werden, inwieweit die Ergebnisse dieser Gruppen übereinstimmen.

Zum Rating der einzelnen Szenen wurden folgende Mittel verwendet:

- 1) Adjektivlisten (verkürzte Eigenschaftswörterliste nach Janke und Debus 1978) zur Einschätzung der Stimmungslage von Patient und Therapeut mit nicht unterteilter Skala. Folgende Adjektive wurden vorgegeben: bekloffen, verstört, abgekapselt, starr, wütend, verärgert, aggressiv, froh, selbstsicher, beschwingt, kontaktfreudig, unbekümmert, wehmütig, traurig, ausgeglichen, gelöst, offen, friedlich, einladend, ängstlich, bedroht, mutlos, hilflos, ratlos und überlegt.
- 2) Freie Äußerung zur Einschätzung der Stimmungslage von Patient und Therapeut
- 3) Freie Formulierung zu der Frage: „Was fällt Ihnen am Patienten auf?“
- 4) Adjektivliste zur Einschätzung der musikalischen Interaktion mit folgenden Adjektiven: ergänzend, harmonisch, zusammenschwingend, verschmolzen, gemeinsam, einig, zugewandt, in Kontakt, isoliert, nebeneinander, getrennt, zerrissen, gegenläufig und entspannt.
- 5) Fragen zu Wünschen und Bedürfnissen des Patienten und zur Reaktion des Therapeuten hierauf (Beantwortung mit ja/nein) (modifizierte Standardkategorien zum zentralen Beziehungskonflikt nach Luborsky und Kächele 1988)
- 6) Frage: „Ist die Interaktion konkordant, komplementär oder diskordant?“ (vgl. Aspekte der Rollenbeziehung bei Racker (1968))
- 7) Freie Formulierung zu der Frage: „Was hätten Sie (als Therapeut, Anm. N.S.) gemacht?“

Schließlich sollten die Rater noch eine Gesamteinschätzung frei formulieren bezüglich der Persönlichkeit des Patienten und seiner Hauptproblematik, sowie der technischen Probleme des Therapeuten.

Aus diesen Daten sollte nun u.a. überprüft werden, inwieweit die Einschätzung der Rater bzgl. Patient, Interaktion und Therapeut (Punkte 1-7) mit den Informationen, die oben unter a-f aufgeführt wurden, übereinstimmen, ob die Rater einen signifikanten Unterschied zwischen

den Szenen 1-5 (repetitive Muster) und 6-8 (pattern of change) erkennen und ob es Unterschiede zwischen den Ratergruppen gibt.

Die Adjektive der Adjektivlisten (Punkte 1 und 4) wurden in verschiedene Kategorien eingeteilt und polarisierend gegenübergestellt. Die Skalen wurden zur Auswertung in fünf gleich große Abschnitte unterteilt. Diesen wurden Werte von 1-5 zugeordnet. Für die Adjektive wurden Mittelwerte über alle Rater (n=50) und über die Ratergruppen bzgl. den Szenen 1-5 und 6-8 berechnet.

Timmermann kommt zu folgenden Ergebnissen:

- In den Szenen 1-5 gibt es für fast alle Adjektive einen signifikanten Unterschied zwischen Patient und Therapeut:
  - typisch für den Patienten sind die Adjektive: abgekapselt, verstört, mutlos, starr, bedroht etc.
  - typisch für den Therapeuten sind die Adjektive: einladend, offen, kontaktfreudig, ausgeglichen, selbstsicher etc.
- Es gibt einen deutlichen Unterschied in der Bewertung des Patienten, des Therapeuten und der Interaktion zwischen den Szenen 1-5 und 6-8.
- Es zeigen sich große Übereinstimmungen zwischen den verschiedenen Ratergruppen bezüglich des Patienten- und Therapeutenspiels, wobei sich jedoch folgende Tendenzen herausstellen: Die Laien schätzen das Spiel eher positiv ein (nehmen eher das Gesunde wahr), die Psychotherapeuten sehen eher das Pathologische, die Musiktherapeuten liegen dazwischen. Dieser Unterschied ist jedoch nur bei einigen Adjektiven signifikant. In der abschließenden Gesamteinschätzung wird der Therapeut von den einzelnen Ratern sehr unterschiedlich eingeschätzt (von zu wenig fordernd bis zu dominant). Diese niedrige Raterübereinstimmung deutet Timmermann als Abbild der inneren Ambivalenz des Therapeuten.
- Es zeigen sich große Übereinstimmungen zwischen den verschiedenen Ratergruppen bei der Beurteilung der Interaktion. Hier besteht jedoch die Tendenz, daß die Musiktherapeuten die Interaktion positiver einschätzen als Laien und Psychotherapeuten
- Die freien verbalen Äußerungen der Rater bezüglich Persönlichkeitsstruktur und Pathologie kommen der Einschätzung des Therapeuten und der Selbsteinschätzung des Patienten sehr nahe.

Timmermann kommt zu dem Schluß, daß Persönlichkeit und Problematik von allen drei Ratergruppen adäquat erfaßt und beschrieben werden.

### Beurteilung:

Die Arbeit von Timmermann zeigt auf, daß es prinzipiell mit einem solchen Studiendesign möglich ist, einen musiktherapeutischen Prozeß durch Mikroprozeßanalyse zu studieren. Es ist aber - wie er selbst bemerkt – notwendig, das Design sinnvoll zu reduzieren und einzelne Teile genauer zu untersuchen.

In der vorliegenden Arbeit werden die Adjektivlisten weitgehend übernommen, um sie genauer statistisch auszuwerten. In Timmermanns Auswertung wurde keine Interraterreliabilität zwischen den einzelnen Ratern angegeben, welche für die Aussagekraft eines Meßinstrumentes, wie sie eine Adjektivliste darstellt, von großer Bedeutung ist. Es werden lediglich die Mittelwerte der verschiedenen Ratergruppen miteinander verglichen. Auch der Unterschied, der sich zwischen dem Rating von Szene 1-5 und 6-8 zeigte, wurde lediglich beschrieben, aber nicht statistisch ausgewertet. Die Zuordnung der Adjektive zu Kategorien wurde aufgrund rationaler Überlegungen festgelegt, eine statistische Beurteilung dieser Kategorien (z.B. Berechnung der internen Konsistenz) fehlt.

### Inselmann und Mann 1999

Inselmann und Mann geht es darum, ein Studiendesign zu entwickeln, „welches potentiell geeignet ist, die Psychopathologie eines Patienten abzubilden und den Therapieprozeß darzustellen“. Dem liegt die Theorie zur aktiven Musiktherapie zugrunde, „daß mit ihr das unbewußte, emotionale Erleben eines Patienten zum Ausdruck kommt, welches selbst Symbolisierungsfähigkeit besitzt und daß sich dieses unbewußte emotionale Erleben in der Beziehung zum anderen spiegelt.“ Langfristiges Ziel soll es sein, ein Routineverfahren zu etablieren, das den Therapieprozeß abbilden kann.

Als Material dienten Videosequenzen einer aktiven Einzelmusiktherapie. Diese wurden durch die beiden Verfasserinnen geratet, wobei zunächst 20 Proberatings stattfanden mit dazwischen liegendem Ratertraining. Danach wurden nochmals 10 Sequenzen geratet und diese statistisch ausgewertet.

Zum Rating wurden folgende Auswertungsinstrumente verwendet:

- 1) Adjektivliste zur Einschätzung des emotionalen Erlebens bzgl. Patientinnenspiel, Therapeutinnenspiel und Gefühlserleben des Raters: Die Adjektivliste wurde weitgehend von Timmermann übernommen und noch um Items zur Interaktion ergänzt. Insgesamt umfaßt sie 35 Adjektive.
- 2) Adjektivliste zur Einschätzung der musikalischen Interaktion: Die Adjektivliste wurde vollständig von Timmermann übernommen und umfaßt 14 Adjektive.

- 3) quantitative Analyse des musikalischen Ausdrucks: Die musikalischen Parameter Rhythmus, Lautstärke, Melodik, Tonumfang und Tempo wurden auf einer 3-stufigen Skala geratet.
- 4) musikalische Beziehungsanalyse: Der Grad der Differenzierung des dialogischen Patientenspiels wurde mit einer quantitativen Skala zu erfassen versucht - in Anlehnung an Punkt 6) bei Timmermann (s.o.).

Zu Therapiebeginn wurde folgende Diagnostik durchgeführt:

- a) Freiburger Persönlichkeitsinventar, revidierte Fassung
- b) Modifizierte Form des Fragebogens „Psychiatrischer und Sozialkommunikativer Befund“
- c) Psychoanalytisches Erstinterview

Die Raterübereinstimmung wurde mit Hilfe des Rangkorrelationskoeffizienten nach Spearman berechnet und ergab folgende Ergebnisse:

Improvisationsspiel der Therapeutin:  $\rho=0.82$

Improvisationsspiel der Patientin:  $\rho=0.66$

Interaktion zwischen Therapeutin und Patientin:  $\rho=0.44$

Gefühlserleben des Raters:  $\rho=0.28$

Absolute Übereinstimmungen in der quantitativen Analyse: Rhythmus: 10/10, Tonhöhe: 10/10, Tempo: 8/10, Lautstärke: 7/10, Melodie: 3/5

Absolute Übereinstimmungen in der musikalischen Beziehungsanalyse: 6/10

Im weiteren wurden für die Adjektivliste zur Beurteilung des Therapeutinnen- und des Patientinnenspiels Mittelwerte der zwei Raterwerte für die einzelnen Items berechnet und mittels P-Faktorenanalyse und rationaler Skalenbildung fünf Adjektivskalen entwickelt. Für diese wurde die interne Konsistenz (Cronbachs  $\alpha$ , siehe Cronbach 1951) berechnet, dabei lag  $\alpha$  zwischen 0.82 und 0.96. Die Skalen wurden wie folgt benannt: „selbstbewußt“, „depressiv“, „harmonisch“, „aggressiv“ und „freudig zugewandt“. Für die jeweiligen Skalen wurden die Skalenwerte (Mittelwert aus den Werten der Items der entsprechenden Skala) für die verschiedenen Sequenzen berechnet, um festzustellen, welche Skalen in den einzelnen Sequenzen hohe Werte erreichen. Die so entstehenden Kombinationen von Skalen werden als Spielmuster bezeichnet. Beim Vergleich der Spielmuster mit der quantitativen Analyse zeigte sich, daß depressive Spielmuster - im Gegensatz zu aktiv-kommunikativen - eher mit einem wenig differenzierten Spiel verbunden sind.

Der Vergleich der Raterergebnisse („musikalische Diagnostik“) mit der Diagnostik zu Therapiebeginn zeige laut den Autorinnen eine sehr gute Übereinstimmung.

Zusammenfassend meinen die Verfasser, „daß durch das vorliegende Auswertungsschema eine statistisch abgesicherte Darstellung der Inhalte von aktiver Musiktherapie ... möglich ist, die die Psychopathologie eines Patienten spiegelt.“

### Beurteilung:

Es ist zu begrüßen, daß der Ansatz Timmermanns weiterentwickelt wurde, indem er einerseits reduziert und andererseits weiter standardisiert wurde. Aufgabe der vorliegenden Arbeit soll es sein, zu untersuchen, inwieweit die Ergebnisse dieser Arbeit bezüglich der Interraterreliabilität reproduzierbar sind. Es wäre auch sinnvoll, die Interraterreliabilität nicht nur für die Sequenzen, sondern auch für die einzelnen Items zu berechnen. Hiermit könnte man herausfinden, ob eine hohe Übereinstimmung beim Raten einer Sequenz evtl. nur darauf zurückzuführen ist, daß man sich beim Ratertraining bewußt oder unbewußt auf ein bestimmtes Rating geeinigt hat. Dieses Problem wird in Kapitel 4.1.4 genauer erläutert.

Fraglich ist, ob die Zusammenfassung der Skalen zu Spielmustern sinnvoll ist bzw. ob es zu weiteren Erkenntnissen führt, wenn z.B. die Skalen „selbstbewußt“, „harmonisch“ und „freudig zugewandt“ als Spielmuster „sicher“ bezeichnet werden, während die Skalen „selbstbewußt“ und „harmonisch“ das Spielmuster „harmonisch“ ergeben.

Die innere kriterienbezogene Validierung (Lienert und Raatz 1994, S.222) der „musikalischen Diagnostik“ durch die Diagnostik zu Therapiebeginn geschieht wie bei Timmermann rein deskriptiv. Solange dieser Vergleich nicht genauer standardisiert ist, scheint mir hier die Gefahr zur selektiven Wahrnehmung und subjektiv gefärbter Interpretation sehr groß. So wird die Patientin in der Testdiagnostik zu Therapiebeginn als unzufrieden, ängstlich, emotional labil und depressiv beschrieben, hingegen erreichte bei der musikalischen Diagnostik die Skala „depressiv“ nur in 3 Sequenzen, die Skala „selbtsicher“ jedoch in 7 Sequenzen hohe Werte. Dies wird in der Interpretation aber nicht berücksichtigt. Ferner ist die Validierung hier prinzipiell kritisch zu betrachten, da eine der Raterinnen zusätzlich auch die Therapeutin war und somit ihre Einschätzungen beim Raten eventuell durch die ihr zusätzlichen bekannten Informationen beeinflusst wurden.

Einige dieser Anmerkungen werden in der vorliegenden Arbeit aufgegriffen und in den entsprechenden Kapiteln genauer behandelt.



### 2.2.1 Zusammenfassende Beurteilung:

Zunächst soll auf das Problem der Interraterreliabilität eingegangen werden. Trotz ihrer Bedeutung (siehe Kapitel 2.1.3) wird sie in vielen der besprochenen Arbeiten vernachlässigt. Manchmal wird das rechnerische Verfahren nicht angegeben, manchmal wird kein Wert zur Beobachterübereinstimmung angegeben oder es wird lediglich auf Vorarbeiten verwiesen. Dies scheint nicht nur ein Problem in der Musiktherapieforschung zu sein. Asendorpf bemerkt hierzu: „Der menschliche Beobachter bzw. Beurteiler als Meßinstrument ist aber – wie jedes Meßinstrument – mit Fehlern behaftet, die durch Reliabilitätsmaße kontrolliert werden müssen. Die Festlegung bzw. Berechnung der Reliabilität der Ratings erfreut sich allerdings bisher – der Wichtigkeit des Problems unangemessen – einer gewissen Willkürlichkeit und Beliebigkeit“ (Asendorpf und Wallbott 1979).

Bei dem Vergleich der verschiedenen Arbeiten muß man berücksichtigen, daß es neben der Güte der Ratingskalen noch weitere Faktoren gibt, die Einfluß auf die Interraterreliabilität haben können, wie z.B. das verwendete statistische Verfahren, die Qualifikation der Rater oder das zu ratende Material. So könnte man vermuten, daß durch Selektion von besonders prägnanten Sequenzen die Interraterreliabilität erhöht wird oder daß die Beobachtungsübereinstimmung der Items höher ist, wenn man Sequenzen von verschiedenen anstatt von nur einem Patienten ratet (insbesondere wenn die Variabilität in dessen Spiel gering ist). Schließlich ist auch zu berücksichtigen, was miteinander verglichen wird. Wie wir in Kapitel 4.1.4 sehen werden, kann es erhebliche Unterschiede zwischen der Interraterreliabilität für einzelne Items und einzelne Sequenzen geben. Alle diese Punkte machen es sehr schwierig, die in den verschiedenen Arbeiten angegebenen Werte zur Beobachterübereinstimmung zu vergleichen. Die vorliegende Arbeit befaßt sich daher ausführlich mit der Beobachterübereinstimmung und untersucht folgende Punkte:

- Vergleich verschiedener statistischer Tests (auf die Wahl der verwendeten Tests wird in Kapitel 3.5.1 eingegangen)
- Einfluß von Ratertraining auf die Interraterreliabilität
- Vergleich von Interraterreliabilität zwischen einzelnen Sequenzen und einzelnen Items
- Analyse der einzelnen Sequenzen, um zu untersuchen, ob es bestimmte Faktoren gibt, die Einfluß auf die Interraterreliabilität haben könnten (siehe Kapitel 4.2)

## **3 Methodik**

In diesem Kapitel werden das Ratingverfahren und die Auswertung des Ratings genau dargestellt. Zunächst werden die Meßinstrumente, die Auswahl der zu ratenden Musiksequenzen, das Setting des Ratings und die Rater beschrieben. Darauf folgt eine genaue Darstellung der zur Anwendung kommenden statistischen Verfahren und einiger für die Beschreibung der Sequenzen wichtiger Punkte.

### **3.1 Meßinstrumente**

Die Meßinstrumente sind weitgehend identisch mit denen von Inselmann und Mann (1999) (siehe Kapitel 2.2). Letztere werden im folgenden beschrieben. Wurden in der vorliegenden Untersuchung Veränderungen vorgenommen, werden diese an den entsprechenden Stellen beschrieben und begründet.

Wie in Kapitel 2 erörtert, erscheint es sinnvoll, parallel sowohl den emotionalen Ausdruck als auch die musikalischen Parameter zu messen (siehe Steinberg 1985,1991,1992 und Pechr 1996), da der Zusammenhang zwischen diesen beiden Aspekten noch wenig erforscht ist. Der Schwerpunkt soll jedoch auf der Erfassung des emotionalen Gehaltes von Musik mit Hilfe von Adjektivlisten liegen, da die Arbeiten von Steinberg und Pechr zeigen, daß mit solchen subjektiveren Verfahren die Psychopathologie eines Patienten besser erfaßt werden kann. Die folgenden Meßinstrumente wurden verwendet:

#### 1) Adjektivlisten zur Bewertung des Patienten- und Therapeutenspiels:

diese umfassen folgende 35 Adjektive:

abgekapselt, selbstsicher, froh, hilflos, unbekümmert, zaghaft, energisch, traurig, beschwingt, ängstlich, zurückgezogen, aggressiv, passiv, friedlich, kontaktfreudig, verärgert, starr, offen, teilnahmslos, agil, gelöst, dominant, konkurrierend, abweisend, introvertiert, selbstunsicher, nachgiebig, freundlich, aufdringlich, mutlos, ausgeglichen, wehmütig, verstört, beklommen und bedroht.

Mit diesen Adjektiven soll das emotionale Erleben bzw. der emotionale Ausdruck gemessen werden. 21 der 35 Adjektive sind von Timmermann (1990) übernommen. Die Adjektivliste von Timmermann ist eine verkürzte Version der Eigenschaftswörterliste (EWL) von Janke und Debus (1971), welche ein mehrdimensionales Verfahren zur

quantitativen Beschreibung des momentanen Befindens darstellt. Die 161 Items der Langform EWL-N lassen sich 15 Skalen und folgenden 6 Bereichen zuordnen: Leistungsbezogene Aktivität, allgemeine Desaktivität, Extroversion/Introversion, allgemeines Wohlbefinden, emotionale Gereiztheit und Angst. Timmermann verwendet Items aus 8 Skalen und 4 Bereichen. In der Untersuchung von Inselmann und Mann werden 21 Items der EWL aus 9 Skalen und aus allen 6 Bereichen verwendet (siehe Anhang A). Zusätzlich werden 8 Items verwendet, die die 8 Skalen des Inventars für interpersonelle Probleme beschreiben (Horowitz et al. 1994). Diese sollen Beziehungsprobleme der einzelnen Spieler erfassen.

Die 35 Adjektive sind untereinander angeordnet. Neben jedem Adjektiv befindet sich bei der ersten Ratingphase der vorliegenden Arbeit eine 10 cm lange visuelle Analogskala, die zur Auswertung mit Hilfe einer durchsichtigen Folie in fünf Grade eingeteilt wurde (dies entspricht dem Vorgehen in der Arbeit von Inselmann und Mann). Da es sich herausstellte, daß die Rater in der vorliegenden Untersuchung diese Analogskala unterschiedliche interpretierten und anwendeten, wurde sie für die nachfolgenden zwei Ratingphasen 5-stufig skaliert und folgendermaßen benannt: „gar nicht – kaum – mittelmäßig – ziemlich – außerordentlich“.

Nach Rohrman (1978) werden diese Intensitätsgrade von Urteilern als weitgehend äquidistant aufgefaßt. Zur Anzahl der Stufen meint Bortz – sich auf eine Studie von Matell und Jakoby (1971) beziehend - , daß diese sowohl hinsichtlich Reliabilität als auch der Validität unerheblich sei (Bortz 1984, S.123).

Auf einem zusätzlichen Erläuterungsbogen befindet sich zu jedem Adjektiv eine semantische Beschreibung, um die Adjektive möglichst genau zu definieren (siehe Anhang G).

## 2) Adjektivliste zur Bewertung der Interaktion:

Die Adjektivliste zur Interaktion wurde vollständig von Timmermann (1990) übernommen und beinhaltet folgende 14 Adjektive: isoliert, zerrissen, getrennt, nebeneinander, gegenläufig, in Kontakt, entspannt, gemeinsam, harmonisch, verschmolzen, zugewandt, einig, ergänzend und zusammenschwingend.

Auch hier wurden in der ersten Ratingphase der vorliegenden Arbeit visuelle Analogskalen und ab der zweiten Ratingphase 5-stufige Skalen verwendet. Zudem wurde nach der ersten Ratingphase die Reihenfolge der Adjektive verändert, so daß die Adjektive nach der Ausprägung des Zusammenspiels geordnet sind. Auch hier wurden die Adjektive auf einem zusätzlichen Erläuterungsbogen genauer beschrieben (siehe Anhang H).

- 3) Skalen zur Beurteilung musikalischer Parameter des Patientenspiels: Die musikalischen Parameter Rhythmus, Lautstärke, Melodik, Tonumfang und Tempo wurden mit einer 3-stufigen quantitativen Kodierung entlang ihres Differenziertheitsgrades folgendermaßen beschrieben:

Punktzahl	1	2	3
Rhythmus	einfacher Rhythmus / arhythmisch	ausgefeilter Rhythmus	Rhythmuswechsel
Lautstärke	nur leise / nur mäßig / nur laut	leise / mäßig / laut im Wechsel	an- und / oder abschwellend
Melodik	keine Melodie, chaotische Tonfolge	einfache Melodie	Melodie mit ausgearbeitetem Motiv
Tonumfang	½ Oktave	1 Oktave	gesamter Tonraum
Tempo	gleichförmig langsam / mittel / schnell	langsam / mittel / schnell im Wechsel	fließende Übergänge

Hieraus läßt sich eine Gesamtpunktzahl berechnen, die die Differenziertheit und Expressivität des Spiels wiedergeben soll. Dabei wird von der Hypothese ausgegangen, daß sich die psychopathologische Einschränkung des Patienten im eingeschränkten musikalischen Ausdruck (Differenziertheit und Expressivität) spiegelt, wie z.B. die Arbeiten von Steinberg (1985) oder Maler (1992) vermuten lassen.

- 4) Skala zur Beurteilung der Differenzierung des Beziehungsspiels.

Hierbei wird folgende 6-stufige Skala verwendet:

Punktzahl	1	2	3
Spiel mit Begleitung	nebeneinander	gegenläufig, disharmonisch, zerrissen	Patient läßt Begleitung durch Therapeuten zu
Punktzahl	4	5	6
Spiel im Dialog	Patient läßt Begleitung zu und versucht durch eigene Impulse, sich auf den Therapeuten einzulassen	Patient und Therapeut sind gleichwertig in Beziehung und spielen gleichzeitig / bzw. abwechselnd einfache Themen	gleichzeitiges oder abwechselndes Spielen mit Aufgreifen von musikalischen Themen des jeweils Anderen in redundanter oder weiterentwickelter Form

Auch dieser Skala liegt die Hypothese zugrunde, daß sich eine psychopathologische Einschränkung des Patienten in einer verminderten Fähigkeit zur differenzierten Kommunikation widerspiegelt.

### **3.2 Musiksequenzen**

Die Musiksequenzen wurden den Videoaufzeichnungen von 36 Einzelmusiktherapie-Sitzungen einer Patientin entnommen. Während dieser Sitzungen fanden 70 musikalische Improvisationen statt, die in originaler Reihenfolge auf ein Videoband überspielt wurden. Dieses Band wurde bereits in der Arbeit von Inselmann und Mann (1999) verwendet, es wurden jedoch dort andere Sequenzen ausgewertet als in dieser Arbeit. Die meisten Sequenzen sind zwischen 1 und 3 Minuten lang. Bis auf wenige Ausnahmen spielen die Therapeutin und die Patientin zusammen. Da es - wie bei Inselmann und Mann 1990 dargestellt - das Ziel ist, ein Meßinstrument zu entwickeln, mit dem man Therapieverläufe abbilden kann, erschien es nicht sinnvoll, wie in vielen anderen Studien (z.B. Timmermann 1990 oder Steinberg 1985) eine Selektion von bestimmten Ausschnitten durchzuführen. Eine solche Auswahl von besonders auffälligen oder typischen Szenen erleichtert eventuell das Rating, es besteht jedoch die Gefahr, daß die selektierten Sequenzen nicht mehr repräsentativ sind und somit der Therapieverlauf nicht korrekt wiedergegeben wird.

Die Videoeinstellung war so gewählt, daß die Patientin gut sichtbar ist, die Therapeutin hingegen ist oft nur teilweise oder nicht erkennbar. Bei der Beschreibung der 10 zuletzt gerateten Sequenzen in Kapitel 4.2 werden die Videoeinstellung der einzelnen Sequenzen kurz dargestellt. Ausgewertet wurden folgende Sequenzen:

1. Ratingphase: Sequenz 24-33
2. Ratingphase: Sequenz 34-43
3. Ratingphase: Sequenz 3-12

### **3.3 Rater und Ablauf des Ratings**

Geratet wurde von zwei Personen. Ein Rater (R1) ist Musiktherapeutin und war selbst die Therapeutin in den zu beurteilenden Videosequenzen. Der andere Rater (R2) ist Doktorand der Medizin und bezüglich Musiktherapie ein Laie. Er hatte zum Zeitpunkt des Ratings keine

zusätzlichen Informationen über die Patientin, um möglichst vorbehaltlos zu urteilen. Dies entspricht der Situation in der Arbeit von Inselmann und Mann (1999). Dieser Ansatz mit zwei Ratern mit sehr unterschiedlichen Voraussetzungen erscheint vertretbar, da in allen Studien, in denen verschiedene Ratergruppen (Musiktherapeuten, Studenten, Musiker, Laien etc.) teilnahmen (siehe Steinberg 1985, Timmermann 1990, Moreau 1996), eine hohe Übereinstimmung zwischen den Gruppen bestand.

Für die einzelnen Ratingsitzungen gab es einen festen Ablauf. Zunächst wurde die zu ratende Sequenz einmal angeschaut, um einen ersten Gesamteindruck zu erhalten. Danach wurde vor dem Ausfüllen eines jeden Fragebogens noch einmal diese Sequenz abgespielt, um auf die spezifischen Aspekte, die mit dem Fragebogen bewertet werden sollen, achten zu können. Insgesamt wurde jede Sequenz also sechsmal angesehen. Dieses war notwendig, da sich die verschiedenen Sequenzen oft nur wenig unterschieden und es daher wichtig war, auch geringe Änderungen im Spiel wahrzunehmen. Die Dauer eines Ratings einer Sequenz betrug maximal 30 Minuten. Pro Sitzung wurden maximal vier oder fünf Sequenzen geratet, da danach die Konzentrationsfähigkeit erfahrungsgemäß stark nachläßt.

Insgesamt gab es 3 Ratingphasen zu je 10 Sequenzen. Nach jeder dieser Phasen wurde die Interraterreliabilität für die einzelnen Sequenzen bestimmt. Zwischen den Ratingphasen wurde ein Ratertraining durchgeführt.

### **3.4 Ratertraining**

In der vorliegenden Arbeit fanden zwischen erster und zweiter und zwischen zweiter und dritter Ratingphase jeweils ein Ratertraining statt. Dabei wurden jeweils zwei Sequenzen mit niedriger und eine mit hoher Übereinstimmung besprochen. Die entsprechenden Sequenzen wurden noch einmal angeschaut und die einzelnen Ratingskalen besprochen. Dabei wurde besonders auf die Items der Adjektivlisten eingegangen, die größere Unterschiede zwischen den Ratern aufwiesen. Bestanden zwischen den Ratern unterschiedliche Auffassungen über die genaue Bedeutung einzelner Items, wurden die semantischen Beschreibungen präzisiert (siehe Anhang G und H). Bestanden große Unterschiede in der Gesamtbeurteilung einer Spielweise, versuchte jeder Rater seine Beurteilung zu begründen. Diese Begründungen wurden diskutiert und es wurde versucht, zu einer gemeinsamen Beurteilung zu gelangen.

## **3.5 Statistische Auswertungsverfahren**

### 3.5.1 Berechnung der Interraterreliabilität

Über das Skalenniveau von Ratingskalen und die damit verbundenen Möglichkeiten zur statistischen Auswertung gibt es verschiedene Meinungen. Bortz bemerkt hierzu: „Die meßtheoretischen ‚Puristen‘ behaupten, Rating-Skalen seien nicht intervallskaliert; sie verbieten deshalb die statistische Analyse mittels parametrischer Verfahren... Demgegenüber vertreten die ‚Pragmatiker‘ den Standpunkt, die Verletzung der Intervallskaleneigenschaften seien bei Ratingskalen nicht so gravierend, als daß man auf die Verwendung parametrischer Verfahren gänzlich verzichten müßte“ (Bortz 1984; S. 124; eine ähnliche Meinung wird auch von Tinsley und Weiss (1975) vertreten).

Wenn es also auch vertretbar sein kann, die Beobachtungsübereinstimmung mit statistischen Verfahren für Intervallskalenniveau zu berechnen (z.B. mit dem Intraclass-Korrelations-Koeffizienten (Asendorpf und Wallbott 1979) oder mit der Produkt-Moment-Korrelation nach Pearson (Bortz 1993, S.189)), so ist es jedoch sicherer, von einem Ordinalskalenniveau auszugehen. Hierfür empfiehlt Bortz den Rangkorrelationskoeffizient „rho“ nach Spearman (Bortz 1993, S.215). Als weitere Möglichkeit nennt er noch Kendalls „tau“ (Kendall, 1962). „Da die beiden Methoden auf verschiedenen mathematischen Kalkülen aufbauen, unterscheiden sich die beiden Zusammenhangsmaße ... z.T. erheblich“ (Bortz und Lienert 1990, S.443). Als Faustregel wird für  $-0.8 < \rho < 0.8$  folgende Gleichung angegeben:  $\tau = 2/3 \rho$ . Das zuweilen verwendete Cohen's „kappa“ (Cohen 1960) empfehlen Asendorpf und Wallbott für Nominalskalen, während man für Cohen's weighted „kappa“ (Cohen 1968), wenn man für die Gewichtung die Abweichungsquadrate verwendet, faktisch von einem Intervallskalenniveau ausgeht (Asendorpf und Wallbott 1979). Außerdem bemerken Asendorpf und Wallbott, daß „kappa“ sehr empfindlich auf die Verteilung der Urteile über die Kategorien reagiert, so z.B. zu falschen Ergebnissen führt, wenn viele Zeilen der Übereinstimmungsmatrix unbesetzt sind. Dieses wäre in der vorliegenden Arbeit der Fall gewesen.

In dieser Arbeit wird daher die Interraterreliabilität - wie bei Inselmann und Mann (1999) - mit dem Rangkorrelationskoeffizienten „rho“ nach Spearman berechnet. Um die Ergebnisse besser mit anderen Arbeiten vergleichen zu können, werden für die Patientendaten exemplarisch auch der Pearson-Koeffizient und Kendalls „tau“ berechnet.

Die Korrelationsberechnung erfolgt sowohl über die einzelnen Sequenzen als auch über die einzelnen Items, d.h. zum einen werden die Werte der verschiedenen Items einer Sequenz in





korreliert. Eine Skalenbildung ist um so besser, je höher die korrigierte Trennschärfe und je niedriger die Fremdtrennschärfe der einzelnen Items ist (Diehl und Staufenbiel 1995, S.506ff). Nach Zuordnung einzelner Items zu anderen Skalen werden für die so entstandenen neuen Skalen noch einmal die interne Konsistenz bestimmt und eine Itemanalyse (korrigierte Trennschärfe und Fremdtrennschärfe) durchgeführt. Unter einem Skalenwert wird das arithmetische Mittel der Itemwerte der Items, die zu der entsprechenden Skala gehören, verstanden.

Die Skalen sollen in der vorliegenden Arbeit v.a. dazu dienen herauszufinden, ob es bestimmte Spielweisen gibt, die besonders leicht oder schwer zu raten sind und ob es Skalen gibt, deren Items allgemein eher niedrige oder hohe Interraterreliabilität haben.

### **3.6 Beschreibung der einzelnen Sequenzen**

Abschließend werden die letzten 10 beurteilten Sequenzen vom Autor ausführlich beschrieben. Dabei wird versucht, einen Zusammenhang zwischen bestimmten Spiel- oder Verhaltensweisen und den statistischen Ergebnissen des Ratings herauszufinden. Es sollen hierbei v.a. folgende Punkte berücksichtigt werden:

- Instrumentenwahl
- Länge der Sequenz
- Veränderungen des Spiels während der Sequenz
- Verhalten in Gestik und Mimik
- Beurteilung, ob es im Patientinnenspiel neue und ungewöhnliche Spielmuster (pattern of change) oder eher sich wiederholende, für die Patientin typische Strukturen (repetitive Strukturen, siehe Timmermann 1990) gibt.
- Gesamteindruck der Patientin: worauf beruht dieser, ist dieser Eindruck eher eindeutig oder ist das Verhalten der Patientin schwer interpretierbar oder ambivalent ?
- Unterschiede in der Beurteilung der beiden Rater

Ziel dieser Beschreibung soll es sein herauszufinden, welche Faktoren Einfluß auf die Raterübereinstimmung haben könnten, um hieraus Verbesserungsvorschläge für weitere Ratings abzuleiten. Zu beachten ist, daß der Autor dieser Beschreibung identisch mit Rater 2 ist.

## 4 Ergebnisse

### 4.1 Statistische Auswertung des Ratings

#### 4.1.1 Interraterreliabilität für einzelne Sequenzen

Die Tabelle 1 gibt einen Überblick über die Rangkorrelationskoeffizienten „rho“ nach Spearman der Raterurteile der einzelnen Sequenzen für die Adjektivlisten zum Patientinnen- und Therapeutinnenspiel und zur Interaktion.

Sequenz	Spearman-Rangkorrelationskoeffizient		
	Spiel der Patientin	Spiel der Therapeutin	Interaktion
24-33	0.26 **	0.60 **	0.26 **
34	0.78**	0.56**	0.90**
35	0.32	0.19	0.75**
36	0.60**	0.72**	0.71**
37	0.49**	0.33	0.66**
38	0.57**	0.80**	0.24
39	-0.20	0.67**	0.55*
40	0.34*	0.76**	0.55*
41	0.11	-	-
42	0.42*	0.53**	0.63*
43	0.16	0.54**	0.75**
34-43	0.38**	0.58**	0.59**
3	0.24	0.78**	0.48
4	0.56**	0.79**	0.81**
5	0.63**	0.84**	0.78**
6	-0.15	0.65**	0.30
7	0.16	0.63**	0.50
8	0.70**	0.75**	0.87**
9	0.87**	0.78**	0.72**
10	0.54**	0.86**	0.90**
11	0.81**	0.55**	0.54*
12	0.27	0.32	0.41*
3-12	0.47**	0.70**	0.60**
Inselmann, Mann 1999	0.66	0.82	0.44

Signifikanzniveau: \*<0.05, \*\*<0.01

**Tab. 1:** Rangkorrelationskoeffizient „rho“ nach Spearman der einzelnen Sequenzen, Gesamtwerte für die drei Ratingphasen und Werte aus der Arbeit von Inselmann und Mann. Für die erste Ratingphase liegt nur der Gesamtwert aller 10 Sequenzen vor.

Zum Vergleich wird der Korrelationskoeffizient zwischen den zwei Ratern der Studie von Inselmann und Mann (1999) angegeben, der auf gleiche Weise berechnet wurde. War die Übereinstimmung zwischen den Ratern bzgl. des Patientinnenspiels in dieser Untersuchung mit  $\rho=0.66$  gut, so konnte dieses Ergebnis in der vorliegenden Arbeit nicht reproduziert werden. Der Korrelationskoeffizient in der letzten Ratingphase (Sequenzen 3-12) entspricht mit  $\rho=0.47$  nicht meßtheoretischen Forderungen. Man kann jedoch deutlich eine Verbesserung zwischen den Ratingphasen erkennen, was auf die Effektivität des Ratertrainings hinweist. Zudem sind die großen Unterschiede zwischen den einzelnen Sequenzen auffällig, wobei die Werte für die letzten 10 Sequenzen zwischen  $\rho= -0.15$  und  $0.87$  schwanken. Mit diesem Ratingverfahren ist es also durchaus möglich, bei bestimmten Sequenzen zu sehr hohen Interraterübereinstimmungen zu gelangen. Die Ergebnisse lassen es sinnvoll erscheinen, die einzelnen Sequenzen genauer zu untersuchen, um Ursachen für die großen Unterschiede der Raterübereinstimmungen in den verschiedenen Sequenzen herauszufinden. Dieses geschieht in Kapitel 4.2. Um die statistischen Ergebnisse mit Studien, in denen andere statistische Verfahren verwandt wurden, vergleichen zu können, werden in Anhang I verschiedene Korrelationskoeffizienten gegenübergestellt.

Wie bei Inselmann und Mann (1999) ist der Korrelationskoeffizient für das Therapeutinnenspiel bedeutend besser als für das Patientinnenspiel. Auch wenn das Ergebnis der Arbeit von Inselmann und Mann ( $\rho=0.82$ ) nicht erreicht wurde, wird mit  $\rho=0.70$  in der letzten Ratingphase eine gute Übereinstimmung erzielt. Im Gegensatz zur Bewertung des Patientinnenspiels zeigen die Ergebnisse der einzelnen Sequenzen für das Therapeutinnenspiel nur geringe Schwankungen.

Der Korrelationskoeffizient für die Interaktion konnte zur Voruntersuchung verbessert werden und ist mit  $\rho=0.60$  in der letzten Ratingphase relativ gut. Auffällig ist die starke Verbesserung zwischen erster und zweiter Ratingphase. Dieses ist mit einer Verbesserung der Operationalisierung einzelner Items zu erklären, die zwischen diesen Phasen durchgeführt wurde. Diese Ergänzungen sind in Anhang H dargestellt.

#### 4.1.2 Skalenkonstruktion

Im Anhang B sind die aus der Arbeit Inselmann und Mann (1999) vorgegebenen Skalen dargestellt. Diese wurden, wie oben beschrieben, anhand der Daten der dritten Ratingphase der vorliegenden Arbeit einer Itemanalyse unterzogen (siehe Anhang C). Die Items „traurig“ und „wehmütig“ konnten nicht analysiert werden, da sie keine Varianz aufwiesen. Insgesamt



### 4.1.3 Vergleich der Sequenzen

Für diese Skalen wurde die Interraterreliabilität der 10 Sequenzen der letzten Ratingphase berechnet. Es wurde wieder der Rangkorrelationskoeffizient nach Spearman verwendet, wobei nun für jede Sequenz die fünf Skalenwerte (anstelle der 35 Itemwerte) in einer Rangreihe angeordnet und diese Rangreihen miteinander verglichen wurden. In Anhang L ist die Interraterreliabilität für das Patientinnen- und Therapeutinnenspiel auf Item- und Skalenebene dargestellt.

Es zeigt sich folgende Tendenz für das Patientinnenspiel: Bei Sequenzen, die bereits auf Itemebene eine geringe Interraterübereinstimmung aufwiesen, ist die Übereinstimmung auf Skalenebene noch geringer, oft sogar negativ. Dies kann damit erklärt werden, daß die geringe Übereinstimmung nicht dadurch zustande kommt, daß die Items nicht genau genug definiert sind, sondern daß die beiden Rater das Patientinnenspiel in einigen Sequenzen prinzipiell unterschiedlich wahrnehmen. Das Therapeutinnenspiel zeigt hingegen auf Skalenebene weitgehend hohe Übereinstimmungen.

Um einen ersten Eindruck zu bekommen, ob es bestimmte Spielarten gibt, die bessere Übereinstimmungen zwischen den beiden Ratern aufweisen als andere, sind in Tabelle 2 die Sequenzen nach Höhe der Interraterreliabilität sortiert und zu jeder Sequenz die Skalen mit höchstem (fett gedruckt) und zweithöchstem Skalenwert für beide Rater angegeben.

Betrachtet man die Skalen mit höchsten Skalenwerten, so fällt auf, daß Rater 1 die Skala „selbstbewußt“ häufiger hoch bewertet als Rater 2. Umgekehrt verhält es sich mit der Skala „harmonisch/freundlich“. Hierauf wird in Kapitel 4.3.3 nochmals eingegangen. Bezüglich der Häufigkeit der Skalen „aggressiv“ und „zugewandt“ sind hier keine Unterschiede zu erkennen.

Wie aus entsprechender Tabelle im Anhang E hervorgeht, ist ein Zusammenhang zwischen Raterübereinstimmung und Instrumentenwahl oder Länge der Sequenz nicht erkennbar.

Sequenz	Interraterreliab.		Patientinnenspiel, Skalen mit höchsten Werten	
	nach Skalen	nach Items	Rater 1	Rater 2
9	1.00**	0.87**	<b>selbstbewußt</b> , aggressiv.	<b>selbstbewußt</b> , aggressiv
5	1.00**	0.63**	<b>zugewandt</b> , harmonisch/freundlich	<b>zugewandt</b> , harmonisch/freundlich
11	0.79	0.81**	<b>selbstbewußt</b> , harmonisch/freundlich	<b>harmonisch/freundlich</b> , selbstbewußt, zugewandt
8	0.35	0.70**	<b>selbstbewußt</b> , zugewandt	<b>harmonisch/freundlich</b> , selbstbewußt
4	0.20	0.56**	<b>harmonisch/freundlich</b> , selbstbewußt	<b>zugewandt</b> , depressiv
10	0.10	0.54**	<b>selbstbewußt</b> , aggressiv, zugewandt	<b>harmonisch/freundlich</b> , selbstbewußt.
3	-0.05	0.24	<b>selbstbewußt</b> , harmonisch/freundlich	<b>aggressiv</b> , selbstbewußt
12	-0.60	0.27	<b>aggressiv</b> , selbstbewußt	<b>harmonisch/freundlich</b> , selbstbewußt
7	-0.72	0.16	<b>aggressiv</b> , selbstbewußt	<b>zugewandt</b> , harmonisch/freundlich
6	-0.80	-0.15	<b>zugewandt</b> , harmonisch/freundlich	<b>aggressiv</b> , selbstbewußt

Signifikanzniveau: \* $<0.05$ , \*\* $<0.01$

**Tab. 2:** Darstellung der Skalen mit höchsten und zweithöchsten Werten der einzelnen Sequenzen. Haben verschiedene Skalen den gleichen Wert, können mehr als zwei Skalen aufgeführt sein. Sequenzen sind nach der Höhe der Raterübereinstimmung auf Skalenebene sortiert.

#### 4.1.4 Interraterreliabilität der einzelnen Items

Um die Items zu bewerten, wurde die Interraterreliabilität für die einzelnen Items aus den Ratingergebnissen der letzten Ratingphase berechnet. Zwecks Übersichtlichkeit sind die Items in Tabelle 3 nicht in der Originalreihenfolge, sondern nach Skalen geordnet. Zusätzlich sind noch die Mittelwerte pro Item für beide Rater angegeben. Die letzte Spalte gibt den Korrelationskoeffizienten an, den man erhält, wenn man Patientinnen- und Therapeutinnenspiel zusammen auswertet. Wenn ein Item bei einem oder beiden Ratern keine Varianz aufweist, kann die Interraterreliabilität nicht berechnet werden. Die durchschnittliche Interraterreliabilität der 35 Items beträgt für das Patientinnenspiel  $\rho=0.16$  und für das Therapeutinnenspiel  $\rho=0.38$ . Diese liegt deutlich unter der Interraterreliabilität der einzelnen Sequenzen (siehe Tabelle 1), was damit erklärt werden kann, daß sich die verschiedenen Sequenzen oft nur wenig voneinander unterscheiden und somit viele Items nur geringe Varianz besitzen. Die Items zeigen bei der Bewertung des Therapeutinnenspiels höhere Übereinstimmungen als bei der Bewertung des Patientinnenspiels, was wahrscheinlich daran liegt, daß das Spiel der Therapeutin differenzierter und abwechslungsreicher ist.

Nr.	Item	Skala	Patientinnenspiel			Therapeutinnenspiel			P.+T.	
			Mittel R 1	Mittel R 2	Spearman's rho	Mittel R 1	Mittel R 2	Spearman's rho	Spearman's rho	
2	selbstsicher	1	3.30	2.60	0.53	3.30	3.30	0.48	0.53*	
5	unbekümmert	1	3.50	2.90	0.15	3.40	3.10	0.26	0.16	
7	energisch	1	3.40	2.70	0.14	2.80	2.40	0.78**	0.54*	
20	agil	1	2.90	2.50	0.03	2.80	2.40	0.89**	0.47*	
22	dominant	1	3.00	2.00	0.78**	2.10	1.10	0.50	0.67**	
23	konkurrierend	1	1.30	1.00	-	1.30	1.00	-	-	
1	abgekapselt	2	1.80	1.90	0.19	1.60	1.00	-	0.26	
4	hilflos	2	1.70	2.10	0.42	1.50	1.30	0.04	0.31	
6	zaghaf	2	1.80	2.00	0.47	2.00	2.50	0.90**	0.73**	
8	traurig	2	1.00	1.00	-	1.00	1.00	-	-	
10	ängstlich	2	1.00	1.10	-	1.00	1.00	-	-	
11	zurückgezogen	2	2.00	2.20	0.22	1.60	1.10	0.60	0.59**	
13	passiv	2	1.20	1.50	0.50	1.10	1.00	-	0.40	
17	starr	2	2.20	1.80	0.44	2.10	1.40	0.07	0.27	
19	teilnahmslos	2	1.50	1.60	0.06	1.00	1.00	-	0.33	
25	introvertiert	2	1.30	2.10	0.21	1.70	1.10	0.59	0.13	
26	selbstunsicher	2	1.10	2.00	0.53	1.40	1.10	-0.22	-0.13	
30	mutlos	2	1.20	1.00	-	1.60	1.00	-	-	
32	wehmütig	2	1.00	1.00	-	1.00	1.00	-	-	
33	verstört	2	1.00	1.00	-	1.20	1.00	-	-	
34	bekommen	2	1.00	1.10	-	1.40	1.00	-	-0.08	
35	bedroht	2	1.00	1.20	-	1.00	1.00	-	-	
3	froh	3	2.10	2.00	-0.14	2.10	1.90	-0.41	-0.26	
9	beschwingt	3	2.20	1.90	-0.23	2.10	2.30	0.42	0.11	
14	friedlich	3	2.20	2.10	-0.23	2.10	2.30	0.19	-0.03	
21	gelöst	3	2.70	2.40	-0.46	2.70	2.80	0.20	-0.15	
28	freundlich	3	2.20	2.40	-0.08	2.60	2.70	0.13	0.03	
31	ausgeglichen	3	2.00	2.10	0.14	2.20	2.90	0.07	0.12	
12	aggressiv	4	2.60	1.70	0.41	1.90	1.60	0.91**	0.60**	
16	verärgert	4	1.80	1.30	0.15	1.10	1.20	0.67*	0.25	
24	abweisend	4	1.60	1.50	-0.05	1.00	1.00	-	0.18	
29	aufdringlich	4	2.10	1.60	0.32	1.90	1.30	0.64*	0.44*	
15	kontaktfreudig	5	2.70	2.50	-0.01	3.20	4.00	-	0.32	
18	offen	5	2.90	2.80	-0.21	3.00	3.90	0.00	-0.06	
27	nachgiebig	5	1.80	1.50	0.00	3.50	3.40	0.65*	0.69**	

Signifikanzniveau: \* < 0.05, \*\* < 0.01

**Tab 3:** Pro Item werden die Mittelwerte beider Rater (Mittel R1, Mittel R2) für die Sequenzen 3-12 angegeben, zudem die Rangkorrelationskoeffizienten nach Spearman aus den Werten für das Patientinnen- und das Therapeutinnenspiel und aus den Werten beider Spiele (P. + T.)

Betrachtet man die verschiedenen Skalen, so fällt auf, daß die Items der Skalen 3 („harmonisch/freundlich“) und 5 („zugewandt“) bedeutend niedrigere Übereinstimmungswerte als die Items der Skalen 1 („selbstbewußt“) und 4 („aggressiv“) haben (siehe hierzu auch Anhang F, in dem die Interraterkorrelationen der einzelnen Skalen angegeben sind).

Dieses kann man damit erklären, daß aggressives oder selbstsicheres Spiel meist expressiver und daher relativ eindeutig bewertbar ist. Wie in Kapitel 4.2 genauer dargestellt wird, ist das Spiel der Patientin meist recht monoton, wobei sie oft unbeteiligt wirkt. Es erwies sich daher häufig als schwierig zu entscheiden, ob dieses Spiel friedlich, freundlich und ausgeglichen oder zaghaft und introvertiert ist. Im Anschluß an die Beschreibung der Sequenzen wird auf diese Punkte noch näher eingegangen.

Im Anhang J sind zusätzlich zu den Ergebnissen der dritten Ratingphase auch die Korrelationskoeffizienten der Items für die zweite Ratingphase dargestellt. Während der durchschnittliche Korrelationskoeffizient der einzelnen Sequenzen für das Patientinnenspiel von  $\rho=0.38$  auf  $\rho=0.47$  ansteigt (siehe Tabelle 1), nimmt der durchschnittliche Korrelationskoeffizient der einzelnen Items von  $\rho=0.29$  auf  $\rho=0.16$  ab (siehe Anhang J). Diese Ergebnisse könnten folgendermaßen erklärt werden: Die Zunahme der Übereinstimmung in der Beurteilung der Sequenzen kann dadurch erreicht worden sein, daß die Rater zuvor ähnliche Sequenzen geratet haben und diese im Ratertraining besprochen haben. Die Rater gleichen sich dabei in der Beurteilung der für diese Patientin typischen Spielart einander an und erreichen daher in der nächsten Ratingphase höhere Übereinstimmungswerte für die einzelnen Sequenzen. Dabei wurde aber die Möglichkeit, mit dem Meßinstrument die oft nur geringen Unterschiede zwischen den Sequenzen zuverlässig abzubilden, nicht verbessert, was an den niedrigen Korrelationskoeffizienten der einzelnen Items erkennbar ist. (In Anhang K wird anhand eines fiktiven Beispiels noch einmal verdeutlicht, wie es zu Unterschieden zwischen den Korrelationskoeffizienten der Sequenzen und den der Items kommen kann.)

Auffällig ist ferner, daß trotz teilweise unterschiedlicher Beurteilung der einzelnen Sequenzen die Mittelwerte der Items hohe Übereinstimmungen aufweisen (siehe Tab. 4). (Berechnet man aus den 35 Itemmittelwerten von Rater 1 und Rater 2 für das Patientinnenspiel den Rangkorrelationskoeffizienten nach Spearman, so erhält man mit  $\rho=0.82$  eine sehr gute Übereinstimmung). Insgesamt wird das Spiel der Patientin von beiden Ratern v.a. als „unbekümmert“, „energisch“, „selbstsicher“, „offen“ und „agil“ beschrieben. Es sind jedoch folgende Unterschiede zwischen den Ratern erkennbar: Die Durchschnittswerte aller Items der Skalen „selbstbewußt“ und „aggressiv“ sind bei Rater 1 höher als bei Rater 2, während Rater 2 die meisten Items der Skala „depressiv“ höher ratet. Hier scheint also ein systematischer Fehler vorzuliegen, der möglicherweise durch die unterschiedlichen Voraussetzungen der Rater erklärbar ist (siehe Kapitel 4.3.3).



Das Spiel der Therapeutin wird von beiden Ratern als „kontaktfreudig“, „offen“, „nachgiebig“, „selbstsicher“ und „unbekümmert“ beschrieben, während für Items der Skalen „depressiv“ und „aggressiv“ von beiden niedrige Werte angegeben werden. Kritisch sei hier zu bemerken, daß diese Einschätzungen auch dadurch beeinflußt sein können, daß die Rater bestimmte Vorstellungen davon haben, wie eine Therapeutin sich verhalten bzw. wie sie musizieren sollte. Dieses Wissen könnte v.a. dann die Beurteilung beeinflussen, wenn das Spiel selbst schwierig zu beurteilen ist (dieses gilt selbstverständlich ebenfalls für die Beurteilung des Patientinnenspiels). Wie beim Patientinnenspiel zeigen auch bei der Beurteilung des Therapeutinnenspiels die Skalen „freundlich“ und „zugewandt“ die niedrigsten Übereinstimmungen auf (siehe auch Anhang F).

#### 4.1.5 Interraterreliabilität der Adjektivliste zur Interaktion

Für die Items der Adjektivliste zur Interaktion ergeben sich für die letzte Ratingphase die in Tabelle 4 dargestellten Interraterreliabilitäten. Während die durchschnittliche Beobachterübereinstimmung für die Sequenzen bei 0.60 lag (siehe Tabelle 1), beträgt sie für die Items fast Null.

Nr.	Item	Rater 1		Rater 2		Spearman's rho
		M	S	M	S	
1	isoliert	1.10	0.32	1.30	0.48	-0.22
2	zerrissen	1.00	0.00	1.00	0.00	-
3	getrennt	1.90	0.74	2.00	0.81	-0.30
4	nebeneinander	2.00	0.82	2.10	0.88	-0.15
5	gegenläufig	1.20	0.42	1.00	0.00	-
6	in Kontakt	3.40	0.70	3.70	0.48	0.50
7	entspannt	2.20	0.92	2.40	0.70	-0.52
8	gemeinsam	3.60	0.70	3.60	0.70	0.08
9	harmonisch	2.60	0.97	2.60	0.52	-0.32
10	verschmolzen	3.00	0.82	2.60	0.52	0.00
11	zugewandt	3.30	0.67	3.50	0.53	0.46
12	einig	2.30	0.82	2.20	0.79	-0.07
13	ergänzend	2.10	1.20	1.50	0.53	0.27
14	zusammen schwingend	1.60	0.70	1.00	0.00	-

**Tab. 4:** Dargestellt sind für jedes Item die Mittelwerte (M) und Standardabweichungen (S) für beide Rater aus dem Rating der Sequenzen 3-12, zudem die Rangkorrelationskoeffizienten nach Spearman.

Bei der Interpretation der statistischen Ergebnisse für das Patientinnenspiel (siehe Kapitel 4.1.4) wurde bereits erläutert, wie durch Ratertraining und Anpassung der Raterurteile für eine bestimmte Spielweise es zu einer Steigerung der Interraterreliabilität der Sequenzen kommen kann, während die Interraterreliabilität der Items niedrig bleibt. Wie man an den Mittelwerten sieht, wird das Spiel von beiden Ratern sehr ähnlich eingeschätzt. Beide Rater geben hohe Werte für die Items „in Kontakt“, „gemeinsam“ und „zugewandt“ an, während Items für sehr ausgeprägte und sehr geringe Interaktion niedrig geratet werden. Berechnet man aus den 14 Itemmittelwerte beider Rater die Interraterreliabilität, so erhält man mit  $\rho=0,96$  eine extrem hohe Übereinstimmung. Der Unterschied zwischen den einzelnen Sequenzen ist jedoch sehr gering, was an der niedrigen Standardabweichung der Items sichtbar ist (siehe Tabelle 4). So kann es trotz relativ hoher absoluter Übereinstimmung zu einem schlechten Korrelationskoeffizienten kommen. Item Nr. 1 hat z.B. eine absolute Übereinstimmung von 60%, aber einen negativen Korrelationskoeffizienten. Ferner muß auch berücksichtigt werden, daß hier für jedes Item Rangreihen aus nur je zehn Werten verglichen wurden. Um die Items besser bewerten zu können, wäre es notwendig, Sequenzen zu raten, die größere Unterschiede aufweisen. Die in Anhang H dargestellten Ergänzungen zu Item 6 und 11 erweisen sich als sinnvoll, da diese beiden Items die höchsten Korrelationskoeffizienten aller Items erreichen.

#### 4.1.6 Auswertung der Skalen zur Beurteilung der musikalischen Parameter

Bei den musikalischen Parametern gibt es folgende absolute Übereinstimmungen (Angabe als Anzahl der absoluten Übereinstimmungen / Anzahl der Ratings):

Rhythmus: 8/10

Lautstärke: 4/10

Melodie: 1/4

Tonumfang: 4/4

Tempo: 6/10

Gesamt: 23/38 (entspricht 60.5 %)

Wenn man berücksichtigt, daß die Items nur 3-stufig skaliert sind (siehe Kapitel 3), so ist die Übereinstimmung zwischen den Ratern deutlich zu niedrig. Lediglich der Parameter „Tonumfang“ weist eine hohe Übereinstimmung auf, da dieser eindeutig operationalisiert ist. In der Arbeit von Inselmann und Mann (1999) lag die Gesamtübereinstimmung zwischen den Ratern mit 84.4 % deutlich höher. Dies ist dadurch erklärbar, daß die Rater in jener Arbeit zuvor dieses Meßinstrument selbst entwickelt haben und somit die Operationalisierung für

beide Rater gleichermaßen eindeutig war. Es scheint daher sinnvoll, einzelne Stufen der Meßskalen noch genauer zu beschreiben, so daß auch ungeübte Rater mit diesem Meßinstrument besser arbeiten können. Es wäre z.B. hilfreich, präziser zu definieren, was unter einem „einfachen Rhythmus“ zu verstehen sei, ob man z.B. bereits einen gleichmäßigen Takt ohne Betonungen als einen einfachen Rhythmus bezeichnen kann. Bzgl. der dritten Stufe dieser Skala wäre es sinnvoll anzugeben, ob bereits ein sehr einfacher Rhythmus, der sich kurzzeitig leicht ändert, dieser Stufe zuzuordnen ist oder ob auch hier – wie in Stufe 2 – ausgefeilte Rhythmen vorliegen müssen. Auch ist zu überlegen, ob nicht die erste Stufe dieser Skala mit arhythmischem Spiel bis einfachem Rhythmus ein relativ breites Spektrum abdeckt, besonders wenn man berücksichtigt, daß man es bei der musiktherapeutischen Improvisation meistens mit einer einfachen Spielart zu tun hat. Beim Parameter „Melodie“ sollte genauer definiert werden, was unter einer „einfachen Melodie“ zu verstehen ist. So müßte z.B. geklärt werden, ob bereits das Spiel einer Tonleiter als einfache Melodie bezeichnet werden kann. Wie beim Parameter „Rhythmus“ ist es möglich, daß bei den Parametern „Lautstärke“ und „Tempo“ den jeweils zweiten Stufen höher differenzierte Spielweisen zugeordnet werden als den jeweils dritten Stufen. Spielt z.B. ein Patient in relativ gleichbleibender Lautstärke, ändert diese aber kurzzeitig ein wenig mit fließendem Übergang, so entspräche dieses der dritten Stufe, spielt er hingegen abwechselnd mal sehr laut und mal sehr leise, so entspräche dieses nur der zweiten Stufe. Unter Berücksichtigung dieser Überlegungen wird in Kapitel 5 ein Vorschlag für eine möglicherweise genauere Operationalisierung dieser Skalen dargestellt.

#### 4.1.7 Auswertung der Skala zur Beurteilung der Differenziertheit des Beziehungsspiels

Tabelle 5 zeigt die Ratingergebnisse dieser Skala. Die absolute Übereinstimmung beträgt 4/10, der Korrelationskoeffizient nach Spearman beträgt  $\rho=0,20$ .

Sequenz	Rater 1	Rater 2
3	4	4
4	3	4
5	5	4
6	4	3
7	2	4
8	5	4
9	3	4
10	4	4
11	3	3
12	3	3

**Tab 5:** Ratingergebnisse der letzten 10 Sequenzen der 6-stufigen Skala zur Beurteilung der Differenziertheit des Beziehungsspiels

Trotz recht genauer Operationalisierung der einzelnen Stufen ist die Raterübereinstimmung deutlich zu niedrig. Bereits bei der Analyse der einzelnen Items der Adjektivliste zur Beurteilung der Interaktion zeigt sich die Schwierigkeit, die Interaktion zwischen der Patientin und der Therapeutin einzuschätzen. Wie in Kapitel 4.2 ausführlicher dargestellt wird, ist es schwierig zu beurteilen, wie stark die Patientin das Spiel der Therapeutin wahrnimmt und inwieweit sie sich auf das Therapeutinnenspiel einläßt. Oft wirkt das Patientinnenspiel diesbezüglich sehr ambivalent, was es dem Rater erheblich erschwert, sich auf eine Beurteilung festzulegen.

Ein weiterer Aspekt sei noch festzuhalten: Während Rater 2 alle Sequenzen ähnlich beurteilt, ist die Varianz bei Rater 1 deutlich höher. Dieses kann damit erklärt werden, daß Rater 1 die Musiktherapeutin in den zu ratenden Sequenzen ist und daher das Verhalten der Patientin differenzierter wahrnimmt und beurteilt. Um die „Skala zur Beurteilung der Differenziertheit des Beziehungsspiels“ beurteilen zu können, wäre es sinnvoll, mit dieser Skala Sequenzen zu bewerten, die sich stärker voneinander unterscheiden, z.B. Sequenzen von verschiedenen Patienten mit verschiedenen Diagnosen bzw. Persönlichkeitsstrukturen.

## 4.2 Beschreibung der Sequenzen 3-12

### Sequenz 3

Interraterreliabilität: Patientin (P.): 0.24; Therapeutin (T.): 0.78; Interaktion (IA): 0.48

Dauer: 35 sec

Instrumente: P.: Bongos und Klangschale; T.: Klavier

Bildausschnitt: Patientin ist gut von der Seite sichtbar, Therapeutin ist nicht zu sehen

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

Rater 1 (R1): Skalen: selbstbewußt, harmonisch/freundlich

Items: energisch (Ratingstufe 4), selbstsicher, unbekümmert, aggressiv,  
offen, agil, gelöst, dominant, aufdringlich (je Ratingstufe 3)

Rater 2 (R2): Skalen: aggressiv, selbstbewußt

Items: abgekapselt, hilflos, energisch, zurückgezogen, teilnahmslos,  
abweisend, aufdringlich (je Ratingstufe 3)

Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' - 0'06''	P. schlägt in mittlerem Tempo gleichmäßig abwechselnd auf die zwei Bongos, in gleichen Abständen fügt sie einige langsamere Schläge ein, sie schaut dabei auf das Instrument.	Das Spiel der P. ist schwer einzuschätzen. Das relativ monotone Spiel wirkt etwas lustlos, obwohl es teilweise auch energisch ist. Es ist schwer zu beurteilen, ob das starre Spiel Ausdruck von Gleichgültigkeit oder Selbstbewußtsein ist.
0'06'' - 0'14''	T. setzt ein. Spielt relativ laute, tiefe Töne. Ahmt Patientinnenspiel nach. P. setzt ihr Spiel fort.	T.-Spiel wirkt energisch und bedrohlich. P.-Spiel wie oben, P. läßt sich nicht durch T. beirren, wirkt daher etwas eigensinnig.
0'14'' - 0'19''	P. sieht auf zur T., schlägt langsam noch einige Male auf Bongos, P. nickt T. zu, Pause.	P. sucht Kontakt zur T., P. macht den Eindruck, als ob sie keine Lust mehr hat zu spielen. Der Abbruch wirkt der T. gegenüber abweisend.

0'19'' - 0'27''	P. schaut zur Seite auf die Klangschale (KS), nimmt Schlegel in die Hand, schlägt einmal auf die KS, sieht zur T. hin und sagt etwas (nicht verständlich).	P. scheint nicht sicher zu sein, was sie machen soll.
0'27'' - 0'35''	P. schlägt in mittlerem Tempo einige Male laut auf die KS, dämpft KS mit der Hand, schaut wieder zur T. und sagt etwas.	P. hat anscheinend nun endgültig keine Lust mehr zu spielen.

Gesamtbeurteilung: Die P. ist sehr schwer einzuschätzen. Manchmal erscheint das Spiel recht energisch, dann wieder eher lustlos, manchmal sucht sie Kontakt, dann wirkt sie abweisend, mal spielt sie starr, dann probiert sie etwas aus. Die Wechselhaftigkeit und Kürze erschweren es, sich ein zuverlässiges Urteil bilden zu können.

#### Sequenz 4

Interraterreliabilität: P.: 0.56      T.: 0.79      IA: 0.81

Dauer: 1 min 9 sec

Instrumente: P.: Holzxylophon      T.: Klavier

Bildausschnitt: P. ist gut von der Seite sichtbar, T. nicht sichtbar

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

R1: Skalen: harmonisch/freundlich, selbstbewußt

Items: selbstsicher, froh, hilflos, unbekümmert, energisch, friedlich, offen, gelöst, freundlich (je 3)

R2: Skalen: zugewandt, depressiv

Items: zaghaft, offen (je 4), hilflos, kontaktfreudig, introvertiert, selbstunsicher, freundlich (je 3)

Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' - 0'42''	P. spielt leise abwechselnd mit beiden Schlegeln, erst tiefe, dann hohe Töne. Zwischendurch streicht sie mit den beiden Schlegeln über das Instrument, spielt manchmal kurz nur mit einem Schlegel.  T. nimmt Spiel der P. auf.	P. probiert vorsichtig und zaghaft etwas Neues aus. Insgesamt aber eher einfaches, relativ monotones und gleichförmiges Spiel.

0'42'' - 1'09''	P. hört auf und sagt etwas zur T. . Spielt weiter wie oben und spricht dabei mit T. .	Weiterhin vorsichtig und ausprobierend. Will anscheinend der T. etwas erklären. Macht den Eindruck, als ob sie nicht wüßte, was sie spielen soll.
--------------------	---	---

Gesamtbeurteilung: Das Spiel der P. ist ziemlich gleichbleibend, allerdings ist es wenig expressiv und relativ einfach, so daß es schwierig ist, daß Spiel eindeutig zu beurteilen. Während R1 dieses gleichförmige Spiel eher als freundlich und harmonisch einschätzt, empfindet es R2 eher als depressiv. Die Schwierigkeit, zu einem eindeutigen Urteil zu gelangen, zeigt sich auch darin, daß R1 die Items „selbstsicher“ und „hilflos“ beide relativ hoch bewertet.

#### Sequenz 5

Interraterreliabilität: P.: 0.63      T.: 0.84      IA: 0.78

Dauer: 1 min 45 sec

Instrumente: P.: Metallxylophon    T.: Klavier

Bildausschnitt: P ist von der Seite sichtbar, Gesicht ist kaum zu sehen, T. nicht sichtbar

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

R1:    Skalen: zugewandt, harmonisch/freundlich

Items: unbekümmert, friedlich, kontaktfreudig, offen, gelöst, freundlich (je 4),  
selbstsicher, froh, energisch, beschwingt, agil, nachgiebig, ausgeglichen  
(je 3)

R2:    Skalen: zugewandt, harmonisch/freundlich

Items: kontaktfreudig (4), hilflos, zaghaft, offen, introvertiert, selbstunsicher,  
freundlich (je 3)

Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' - 0'12''	P. spielt abwechselnd mit beiden Schlegeln in mittlerer Geschwindigkeit, relativ leise, tiefer und mittlerer Tonbereich. T. spielt laut und amelodisch, v.a. hohe Töne, wiederholt oft gleichen Ton mehrmals hintereinander.	P. ist vorsichtig, einfaches gleichbleibendes Spiel. T.-Spiel wirkt aggressiv und herausfordernd, P. läßt sich nicht beirren und spielt monoton weiter.

0'12'' - 1'23''	P. schaut mehrmals kurz auf, lächelt manchmal und sagt etwas zur T., spielt dabei weiter wie oben. Bricht einmal kurz ab und sagt etwas. Das Spiel wird zum Schluß hin etwas energischer mit hohen Tönen, ähnlich dem T.-Spiel.	P. spielt weiterhin sehr einfach und monoton. Scheint nicht zu wissen, wie sie auf das Spiel der T. reagieren soll, sucht anstatt auf musikalischer Ebene eher durch Blicke und verbal Kontakt zu T..
1'23'' - 1'45''	P. bricht recht abrupt das Spiel ab. T. spielt noch einmal wenige Töne, P. reagiert nicht darauf. P redet mit T., spielt noch einmal kurz	P. scheint keine Lust mehr zu haben weiterzuspielen und sucht statt dessen das Gespräch mit T..

Gesamtbeurteilung: Das Spiel ist sehr ähnlich wie in Sequenz 4. P. probiert weniger aus, paßt sich aber zum Schluß kurz dem T.-Spiel an. Da außerdem P. immer wieder verbalen Kontakt zu T. sucht, führt dieses dazu, daß sich bei beiden Ratern hohe Skalenwerte für die Skala „zugewandt“ ergeben. Wie bei Sequenz 4 empfindet R1 das Spiel eher selbstsicher und harmonisch, R2 eher selbstunsicher und introvertiert. Auch in dieser Sequenz ist das Spiel der P. nur wenig expressiv, so daß für die Rater viel Raum zur Interpretation gegeben ist: Ist die P. so selbstsicher, daß sie relativ unbeirrt ihr recht monotones Spiel durchhält, oder ist sie selbstunsicher und gehemmt, so daß sie es nicht wagt, zu variieren?

### Sequenz 6

Interraterreliabilität: P.: -0.15      T.: 0.65      IA: 0.30

Dauer: 0 min 58 sec

Instrumente: P.: Metallxylophon    T.: Klavier

Bildausschnitt: P. ist von schräg hinten sichtbar, T. ist nicht zu sehen

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

R1:    Skalen: zugewandt, harmonisch/freundlich

Items: unbekümmert, beschwingt, friedlich, kontaktfreudig, offen, agil, nachgiebig, freundlich (je 4), selbstsicher, froh, gelöst, ausgeglichen (je 3)

R2:    Skalen: aggressiv, selbstbewußt

Items: energisch, starr (je 4), abgekapselt, hilflos, unbekümmert, zurückgezogen, aggressiv, teilnahmslos, abweisend, introvertiert, aufdringlich (je 3)



Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' -0'58''	P. schlägt abwechselnd mit beiden Schlegeln, relativ laut, nur hohe und meist gleiche Töne in mittlerem Tempo. Schaut einmal kurz zur T.. Zum Schluß zwei abrupte Tempowechsel mit kurzzeitig langsamerem Spiel. Abruptes Ende. T. nimmt Spiel der P. auf, variiert aber in der Tonhöhe.	P.-Spiel noch eintöniger und starrer. Außer kurzen Tempowechseln keine Veränderungen im Spiel. Lauteres Spiel wirkt energisch und aggressiv.

Gesamtbeurteilung: Es besteht eine gewisse Ähnlichkeit zu den vorherigen Sequenzen. Das Spiel ist aber noch weniger abwechslungsreich, die P. probiert weniger aus. Die Tatsache, daß die Varianz im Spiel noch geringer ist als in den vorherigen Sequenzen, könnte ein Grund für die erheblichen Unterschiede in den Raterurteilen sein. Erschwerend für das Rating ist auch, daß man das Gesicht der P. kaum sehen kann und die P. nicht wie in den vorherigen Sequenzen wiederholt aufschaut oder verbalen Kontakt mit der T. sucht. Dieses könnte auch die bedeutend niedrigere Übereinstimmung der Beurteilung der Interaktion erklären. Während R1 das Spiel als „gemeinsam“ empfindet, beschreibt R2 es als „nebeneinander“ und „getrennt“. Das P.-Spiel wird von R1 wie in den vorherigen Sequenzen als freundlich, harmonisch und zugewandt beurteilt. R2 empfindet das Spiel wie zuvor als introvertiert und hilflos und gibt zudem höhere Werte für die Items „aggressiv“, „energisch“ und „starr“ an.

### Sequenz 7

Interraterreliabilität: P.: 0.16      T.: 0.63      IA: 0.50

Dauer: 1 min 43 sec

Instrumente: P.: Bongos / Klangschale (KS)      T.: Klavier

Bildausschnitt: P. ist von schräg vorne sichtbar, T. ist nicht zu sehen

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

- R1: Skalen: aggressiv, selbstbewußt,  
Items: aggressiv (4), selbstsicher, unbekümmert, zaghaft, energisch, verärgert,  
agil, dominant, abweisend (je 3)
- R2: Skalen: zugewandt, harmonisch/freundlich  
Items: offen (4), selbstsicher, froh, unbekümmert, energisch, kontaktfreudig,  
agil, gelöst, freundlich (je 3)

Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' - 0'36''	P. schlägt abwechselnd mit beiden Händen auf die zwei Bongos, zwischendurch streicht sie mit einer Hand über das Instrument oder schlägt leise mit einer Hand auf einen Bongo. Lautstärke und Geschwindigkeit wechseln häufig. Insgesamt arhythmisch. T. versucht das Spiel aufzunehmen, ist eher zurückhaltend.	P. probiert verschiedene Spielweisen aus, scheint aber nicht recht zu wissen, wie sie spielen soll. Das Spiel wirkt manchmal zögerlich, manchmal energisch. Für T. scheint es schwierig zu sein, sich dem Spiel der P. anzupassen, da diese ihr Spiel häufig und abrupt ändert.
0'36'' - 1'03''	P. spielt auf der KS. Abwechselnd schlägt sie oder kreiselt mit dem Schlegel in der Klangschale in unterschiedlicher Geschwindigkeit und Lautstärke.	
1'03'' - 1'43''	P. spielt wieder auf den Bongos wie oben.	

Gesamtbeurteilung: Wie in Sequenz 3 ist das Spiel der P. sehr wechselhaft, so daß es schwierig ist, zu einem Gesamteindruck zu kommen. Manche Abschnitte wirken eher zögerlich, andere eher energisch. Das Spiel ist relativ abwechslungsreich, was R2 als „offen“, „gelöst“ und „freundlich“ empfand. Dadurch ist es für T. schwierig, sich dem Spiel der P. anzupassen, was evtl. R1, die zugleich Therapeutin war, dazu veranlaßte, das Spiel als „aggressiv“, „dominant“ und „abweisend“ zu empfinden. Hier zeigt sich, daß die verschiedenen Positionen bzw. Voraussetzungen der beiden Rater anscheinend einen größeren Einfluß auf das Rating haben als angenommen. Die relativ niedrige Interraterreliabilität für die Interaktion könnte wie in Sequenz 6 damit zusammenhängen, daß beide Sequenzen wenig visuelle Informationen enthalten, die P. schaut weder zur T., noch sucht sie verbalen Kontakt.

### Sequenz 8

Interraterreliabilität: P.: 0.70      T.: 0.75      IA: 0.87

Dauer: 3 min 59 sec

Instrumente: P.: Bongos      T.: Bongos

Bildausschnitt: P. und T. sind von der Seite sichtbar, beide sitzen sich gegenüber

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

R1: Skalen: selbstbewußt, zugewandt

Items: selbstsicher, unbekümmert, energisch, aggressiv, agil, dominant (je 4),  
kontaktfreudig, verärgert, offen, gelöst, nachgiebig, aufdringlich (je 3)

R2: Skalen: harmonisch/freundlich, selbstbewußt

Items: unbekümmert (4), selbstsicher, froh, energisch, beschwingt, friedlich,  
kontaktfreudig, offen, agil, gelöst, freundlich, ausgeglichen (je 3)

Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' - 0'58''	P. spielt die Bongos mit zwei Schlegeln. Schlägt abwechselnd mit beiden Schlegeln auf eine Bongo, wechselt mehrmals die Bongos. Tempo variiert, meist abrupter Tempowechsel, gelegentlich kontinuierliche Temposteigerung. Lautstärke variiert. T. spielt ebenfalls mit zwei Schlegeln, nimmt Spiel der P. auf.	P. probiert verschiedene Spielweisen aus. Das Spiel ist durch abrupte Tempowechsel arhythmisch. Es ist manchmal relativ energisch, teilweise wirkt es aber unbeholfen. Für T. scheint es schwierig zu sein, das P.-Spiel aufzunehmen, da P. die Spielweise oft abrupt wechselt.
0'58'' - 1'13''	P. unterbricht das Spiel. P. und T. reden miteinander.	P. scheint nicht sicher zu sein, wie sie weiterspielen soll
1'13'' - 3'30''	Beide spielen mit den Händen. P. schlägt abwechselnd auf beide Bongos. Oft abrupte Tempo- und Lautstärkenwechsel. Gelegentlich kontinuierliche Temposteigerung, zeitweise recht lautes und schnelles Spiel. Kurze Pause, P. streicht über Bongos. Spiel weiter wie vorher. T. nimmt P.-Spiel auf, versucht eigene Impulse zu setzen.	Spieleindruck wie oben. T. und P. scheinen gelegentlich im Spiel zusammenzukommen.
3'30'' - 3'59''	P. bricht Spiel ab, streicht über Bongos, macht noch einige zögerliche Schläge, schaut zur T.. T. und P. sprechen miteinander	P. scheint keine Lust mehr zu haben, weiter zu spielen.

Gesamtbeurteilung: Die P. spielt zwar abwechslungsreich, der Gesamteindruck ist aber trotz der Länge der Sequenz über die Zeit relativ konstant, was das Rating erheblich vereinfacht. Im Gegensatz zur vorherigen Sequenz sind in dieser Sequenz mehr visuelle Informationen vorhanden – die P. schaut zur T. und spricht mit ihr. Auch die Tatsache, daß sich beide Personen gegenüber sitzen und beide auf dem Video sichtbar sind, scheint besonders die Beurteilung der Interaktion zu erleichtern. Das Spiel ähnelt teilweise der vorherigen Sequenz. Die Spielweise der P. wirkt aber in dieser Sequenz selbstsicherer und weniger zögerlich. Beide Rater beschreiben das Spiel als selbstbewußt und kontaktfreudig. Wie in der vorherigen Sequenz empfindet R2 das Spiel jedoch v.a. als freundlich, während R1 es auch als aggressiv und verärgert empfindet. Hier wird auch wieder der ambivalente Eindruck deutlich, den die P. oft hinterläßt. Manchmal scheint sie den Kontakt zur T. zu suchen (oft mehr verbal als musikalisch), manchmal jedoch wirkt das Spiel eher abweisend, indem sie z.B. das Spiel so abrupt ändert, daß es für die T. fast unmöglich ist, mit der P. zusammen zu spielen.

### Sequenz 9

Interraterreliabilität: P.: 0.87      T.: 0.78      IA: 0.72

Dauer: 1 min 5 sec

Instrumente: P.: Klangschale      T.: Bongos

Bildausschnitt: P. und T. sind von der Seite sichtbar, beide sitzen sich gegenüber

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

R1: Skalen: selbstbewußt, aggressiv

Items: energisch, dominant (je 5), selbstsicher, unbekümmert, aggressiv, starr, aufdringlich (je 4), zurückgezogen, kontaktfreudig, verärgert, offen, agil, konkurrierend, abweisend (je 3)

R2: Skalen: selbstbewußt, aggressiv

Items: energisch, agil, dominant (je 4), abgekapselt, selbstsicher, unbekümmert, aggressiv, verärgert, starr, aufdringlich (je 3)

Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' - 1'05''	P. hält mit einer Hand die KS gekippt und schlägt mit dem Schlegel in der KS von einer Seite zur anderen. Erst mittleres Tempo und mittlere Lautstärke, dann Steigerung. Schlägt wenige Male von außen an KS, dann wieder von innen, sehr schnell und laut. Schaut während dessen gelegentlich zur Seite. Sieht schließlich zur T., sagt etwas und beendet das Spiel. T. spielt teilweise eigene einfache Rhythmen, paßt sich in Tempo und Lautstärke der P. an.	Das starre, laute und schnelle Spiel der P. wirkt aggressiv, selbstsicher und dominant. Sie scheint wenig auf T. zu achten, sondern ist ganz mit ihrem eigenen Spiel beschäftigt

Gesamtbeurteilung: Diese Sequenz bietet einen sehr einheitlichen Gesamteindruck, das Spiel der P. verändert sich während der Sequenz fast gar nicht. Die Spielweise der P. unterscheidet sich stark von allen anderen Sequenzen. Sie spielt sehr aggressiv und wütend. Die P. vermittelt nicht wie in vielen anderen Sequenzen einen ambivalenten oder teilnahmslosen Eindruck. Dieses zeigt sich darin, daß viele Items hohe Werte erhielten (nur in dieser Sequenz wurde für das P.-Spiel die Stufe „außerordentlich“ verwendet). Die Interraterreliabilität für das Patientinnenspiel weist in dieser Sequenz mit  $\rho=0,87$  den höchsten Wert auf.

### Sequenz 10

Interraterreliabilität: P.: 0.54      T.: 0.86      IA: 0.90

Dauer: 1 min 18 sec

Instrumente: P.: Holzxylophon      T.: Klavier

Bildausschnitt: P. ist von schräg vorne sichtbar, T. nicht sichtbar

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

R1: Skalen: selbstbewußt, aggressiv, zugewandt

Items: abgekapselt, selbstsicher, unbekümmert, energisch, zurückgezogen, aggressiv, kontaktfreudig, starr, dominant (je 3)

R2: Skalen: harmonisch/freundlich, selbstbewußt

Items: unbekümmert (4), selbstsicher, froh, energisch, beschwingt, friedlich, offen, agil, gelöst, freundlich, ausgeglichen (je 3)

Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' - 1'18''	<p>P. benutzt abwechselnd beide Schlegel. Aufsteigende und absteigende Tonhöhe. Die Lautstärke ist leise bis mittel mit leichten Variationen in der Geschwindigkeit. T. nimmt Spiel der P. auf. Kurzzeitig spielt P. etwas lauter abwechselnd zwei hohe Töne. Dann wieder wie zuvor. Kurzzeitig spielt sie ein langsames Tempo mit abruptem Tempowechsel.</p> <p>Noch einmal nur hohe Töne, etwas lauter, gefolgt von einigen Schlägen mit beiden Schlegeln gleichzeitig. P. beendet das Spiel und schaut zur T..</p>	<p>P.-Spiel ist relativ abwechslungsreich, aber trotzdem einfach. P. wirkt etwas lustlos. Der Umgang mit dem Instrument wirkt sicherer als in früheren Sequenzen. P. spielt weniger zaghaft.</p>

Gesamtbeurteilung: Das Spiel ähnelt früheren Sequenzen mit Holz- oder Metallxylophon, der Umgang mit dem Instrument ist aber sicherer und auch abwechslungsreicher. Beide Rater stufen daher die Items „selbstsicher“, „unbekümmert“ und „energisch“ hoch ein. Wie in Sequenz 7 und 8 bewertet R2 das Spiel als freundlich und harmonisch, während R1 es als „aggressiv“ und „dominant“ empfindet. Auch wird bei R1 der ambivalente Aspekt der P. wieder deutlich, indem R1 sowohl das Item „abgekapselt“, als auch das Item „kontaktfreudig“ hoch ratet. Während die T. versucht, sich auf die P. einzustellen, ist es - wie in den meisten anderen Sequenzen - schwierig zu beurteilen, inwieweit die P. das Spiel der T. wahrnimmt. Manchmal hat man den Eindruck, die P. spiele abgekapselt vor sich hin und registriere die T. nicht, manchmal hat man den Eindruck, sie suche Kontakt, indem sie ihr Spiel variiert.

### Sequenz 11

Interraterreliabilität: P.: 0.81      T.: 0.55      IA: 0.54

Dauer: 1 min 16 sec

Instrumente: P.: Bongos/Tambourin      T.: Bongos

Bildausschnitt: P. und T. sind von der Seite sichtbar, beide sitzen sich gegenüber

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

R1: Skalen: selbstbewußt, harmonisch

Items: selbstsicher, unbekümmert, energisch (je 4), froh, beschwingt, friedlich, kontaktfreudig, offen, agil, gelöst, dominant, freundlich, ausgeglichen (je 3)

R2: Skalen: harmonisch, selbstbewußt, zugewandt

Items: selbstsicher, froh, unbekümmert, beschwingt, friedlich, offen, agil, gelöst, freundlich, ausgeglichen (je 3)

Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' -0'30''	P. schlägt abwechselnd mit beiden Händen auf die Bongos, das Tempo leicht variierend, in mittlerer Lautstärke, zwischendurch spielt sie etwas lauter.  T. nimmt P.-Spiel auf, spielt leise und zurückhaltend, teilweise ist sie kaum hörbar.	Relativ abwechslungsreiches und ruhiges Spiel. P. wirkt gelassen.  Zwischendurch etwas energischeres Spiel.  T. hält sich sehr zurück und ist kaum wahrnehmbar.
0'30'' -1'16''	P. spielt auf dem Tambourin in leiser bis mittlerer Lautstärke und mit Variationen im Tempo. Ein einfacher Rhythmus mit Betonungen ist erkennbar.  P. beendet das Spiel und schaut zur T.	

Gesamtbeurteilung: Das Spiel ist relativ einheitlich über die Zeit. Zum erstenmal ist ein einfacher Rhythmus mit Betonungen erkennbar. Wie in der vorherigen Sequenz geben beide Rater hohe Werte für die Items „selbstsicher“ und „unbekümmert“ an. Das ruhige und recht sanfte Spiel wird zudem von beiden Ratern als freundlich empfunden. Im Gegensatz zur hohen Raterübereinstimmung für das P.-Spiel sind die Interraterreliabilitäten für das T.-Spiel und die Interaktion unterdurchschnittlich. Dieses läßt sich damit erklären, daß die T. sehr zurückhaltend und leise spielt, so daß sie teilweise kaum zu hören ist.

Sequenz 12

Interraterreliabilität: P.: 0.27      T.: 0.32      IA: 0.41

Dauer: 1 min 8 sec

Instrumente: P.: Klangschale      T.: Triangel

Bildausschnitt: P. und T. sind von der Seite sichtbar, beide sitzen sich gegenüber

Skalen und Items (mit Ratingstufe) des Patientinnenspiels mit höchsten Werten:

R1: Skalen: aggressiv, selbstbewußt

Items: dominant (4), selbstsicher, unbekümmert, energisch, aggressiv, starr, teilnahmslos, abweisend, aufdringlich (je 3)

R2: Skalen: harmonisch, selbstbewußt

Items: selbstsicher, unbekümmert, zurückgezogen, friedlich, agil, gelöst, dominant, freundlich, ausgeglichen (je 3)

Zeit	objektive Beschreibung	subjektiver Eindruck
0'00'' -1'08''	P. schlägt mit dem Schlegel in der KS von einer Seite zur anderen. Mittleres Tempo und Lautstärke, relativ konstant. Zwischendurch schlägt sie einige Male von außen an die KS oder kreist mit dem Schlegel in der KS. Zum Schluß nimmt sie die KS in die Hand, dreht sie um und schlägt auf die Unterseite. Sie stellt die KS wieder hin und legt das Unterlegkissen auf die KS, schaut zur T. und sagt etwas. T. spielt sehr leise wenige Töne auf der Triangel.	P. probiert verschiedene Spielweisen aus, findet aber immer wieder zum Grundspiel zurück, welches relativ gleichbleibend ist. P. wirkt eher teilnahmslos und lustlos. T.-Spiel ist sehr zurückhaltend und kaum wahrnehmbar und daher schwer beurteilbar.

Gesamtbeurteilung: P.-Spiel ist relativ gleichbleibend über die Zeit. Die P. probiert verschiedene Spielweisen aus und macht dabei einen recht selbstsicheren Eindruck, weshalb beide Rater die Items „dominant“, selbstsicher“ und „unbekümmert“ hoch bewerten. Ansonsten ist aber die P. schwer einzuschätzen. Wie in den Sequenzen 7 und 10 bewertet R1 das Spiel eher als aggressiv und abweisend, während R2 es als freundlich und ausgeglichen bewertet. Wie in den meisten anderen Sequenzen spielt die P. wenig expressiv. Auch aus der Gestik und Mimik lassen sich wenige Informationen ableiten. Die P. spielt gleichmäßig vor



sich hin und läßt kaum Emotionen erkennen. Es ist ebenfalls schwer herauszufinden, ob bzw. wie stark sie das T.-Spiel wahrnimmt. Wie in vielen anderen Sequenzen kann dieses Verhalten zu verschiedenen Eindrücken bzw. Interpretationen führen. Auch die Interraterkorrelation für das T.-Spiel und die Interaktion sind gering. Dieses läßt sich damit erklären, daß die T. sehr zurückhaltend und kaum wahrnehmbar spielt.

### **4.3 Zusammenfassende Auswertung**

#### **4.3.1 Sequenzlänge und Videoeinstellung**

Wie in Anhang E ersichtlich, ist es nicht möglich, einen direkten Zusammenhang zwischen Sequenzlänge und Höhe der Raterübereinstimmung festzustellen. Wichtig scheint hingegen zu sein, wie konstant das Spiel innerhalb einer Sequenz ist, ob man also einen einheitlichen Gesamteindruck von dem Spiel bekommt oder ob das Spiel sich innerhalb der Sequenz verändert. Im letzteren Fall kann es vorkommen, daß die Rater verschiedene Abschnitte der Sequenz unterschiedlich gewichten und somit zu abweichenden Urteilen gelangen. Wie in Kapitel 4.2 dargestellt, sind die Sequenzen 3 und 7 besonders uneinheitlich, beide Sequenzen weisen mit  $\rho=0.24$  bzw.  $0.16$  sehr niedrige Raterübereinstimmungen für das Patientinnenspiel auf (siehe Tabelle 1). Dagegen ist das Spiel der Patientin in Sequenz 8, der mit Abstand längsten Sequenz, ziemlich konstant; die Interraterreliabilität ist hier mit  $\rho=0.70$  überdurchschnittlich hoch.

Die Videoeinstellungen unterscheiden sich v.a. darin, daß in den Sequenzen 8, 9, 11 und 12 zusätzlich die Therapeutin zu sehen ist. Patientin und Therapeutin sitzen sich in diesen Sequenzen auf dem Boden gegenüber. Wie man aus Tabelle 1 erkennen kann, geht dieses jedoch nicht mit höheren Werten für die Raterübereinstimmung für das Therapeutinnenspiel oder die Interaktion einher.

#### **4.3.2 Bewertung des Patientinnenspiels**

Hier soll noch einmal dargestellt werden, daß es einige Eigenschaften im Patientinnenspiel gibt, die – unabhängig vom Meßinstrument – eine übereinstimmende Beurteilung erschweren. Wie in Kapitel 4.2 mehrfach erwähnt, ist das Patientinnenspiel sehr oft schwer einzuschätzen. Das Spiel ist meist relativ einfach und emotionslos. Die Patientin geht kaum aus sich heraus. Manchmal macht sie den Eindruck, als spiele sie teilnahmslos vor sich hin. Die einzige Ausnahme stellt Sequenz 9 dar, in der sie emotional beteiligt ist und sehr aggressiv spielt.

Diese Sequenz weist mit  $\rho=0.87$  die höchste Interraterreliabilität auf. Alle anderen Sequenzen sind untereinander relativ ähnlich. Daher ist es besonders wichtig, daß man sich vor dem Rating bzw. beim Ratertraining darauf einigt, wie stark man Unterschiede zwischen den einzelnen Sequenzen im Spiel der Patientin bewerten soll, ob man also versucht, jede Sequenz unabhängig von den anderen zu betrachten oder ob man sie mit den anderen in Beziehung setzt und schon geringere Varianzen im Spiel relativ stark unterschiedlich bewertet. Es muß also ein Bezugsrahmen für die Beurteilungen festgelegt werden. In der vorliegenden Arbeit haben sich die Rater daher darauf geeinigt, auch auf Unterschiede zwischen den Sequenzen zu achten. Spielt die Patientin z.B. in den meisten Sequenzen relativ zaghaft, in einer Sequenz hingegen ein wenig energischer, so soll das Item „energisch“ relativ hoch gewertet werden. Dabei besteht jedoch die Gefahr, daß kleine Unterschiede von den Ratern unterschiedlich wahrgenommen werden und somit bei einzelnen Items schnell größeren Abweichungen zwischen den beiden Raterurteilen auftreten. Ferner kommt hinzu, daß das Verhalten bzw. das Spiel der Patientin oft unterschiedliche Interpretationen zuläßt. Es ist z.B. schwer beurteilbar, ob das oft gleichförmige Spiel der Patientin Ausdruck von Unsicherheit und Unbeholfenheit, von Introvertiertheit oder von Gleichgültigkeit ist, ob es eher friedlich und harmonisch oder depressiv klingt. So wird z.B. das sehr eintönige und gleichbleibende Spiel der Patientin in Sequenz 6 von R1 als freundlich und friedlich, von R2 hingegen als abgekapselt und introvertiert empfunden. Wie bereits in Kapitel 4.2 dargestellt, wird das relativ monotone Spiel der Patientin in Sequenz 5 von einem Rater eher als selbstsicher und energisch, vom anderen als selbstunsicher und zaghaft empfunden. Oft erzeugt die Patientin einen ambivalenten Eindruck beim Rater, was daran erkennbar ist, daß in vielen Sequenzen von einem Rater sich anscheinend widersprechende Adjektive hoch gewertet werden, so werden von R1 z.B. in Sequenz 4 „selbstsicher“ und „hilflos“, in Sequenz 8 „dominant“ und „nachgiebig“ und in Sequenz 10 „abgekapselt“ und „kontaktfreudig“ relativ hoch bewertet. Dies bildet wahrscheinlich die Ambivalenz der Patientin gut ab, erschwert aber gleichzeitig eine übereinstimmende Beurteilung durch beide Rater. Die Unsicherheit im Rating zeigt sich auch darin, daß die höheren Stufen der Ratingskalen von beiden Ratern nur selten benutzt werden (zugleich zeigt dies natürlich auch, daß es dem Patientinnenspiel an Expressivität mangelt). So beträgt bei der Beurteilung des Patientinnenspiels die Häufigkeit der vierten und fünften Stufe insgesamt nur 6%. Zusammenfassend läßt sich feststellen, daß aufgrund des variationsarmen und teilweise widersprüchlichen und ambivalenten Spiels der Patientin die Reliabilität des Meßinstrumentes nicht ausreichend zu bestimmen ist. Mit anderen Worten: es kann sein, daß das Instrument

reliabel ist und Unterschiede abbilden kann. Dieses kann jedoch aufgrund der Spielweise der Patientin in dieser Arbeit nicht sicher beurteilt werden.

#### 4.3.3 Unterschiede zwischen den Ratern

Hier soll besonders darauf eingegangen werden, ob Unterschiede im Ratingverhalten mit den unterschiedlichen Voraussetzungen der beiden Rater (siehe Kapitel 3.3) erklärbar sind. Für die Sequenzen 7, 8, 10 und 12 ergeben sich bei R1 hohe Skalenwerte für die Skala „aggressiv“ und niedrige für die Skala „freundlich“, bei R2 verhält es sich genau umgekehrt. Im Kapitel 4.2 werden diese Sequenzen alle ähnlich beschrieben: die Patientin macht einen selbstsichereren Eindruck als in den vorherigen Sequenzen, das Spiel ist insgesamt zwar einfach, aber abwechslungsreicher als zuvor, sie probiert verschiedene Spielweisen aus und wechselt das Spiel manchmal recht abrupt. R2 bewertet die Spielweise in allen vier Sequenzen ähnlich, wobei er neben Adjektiven der Skala „selbstsicher“ auch Adjektive wie „offen“, „freundlich“, „unbekümmert“, „ausgeglichen“ und „kontaktfreudig“ hoch bewertet. R1 schätzt hingegen oft Adjektive wie „aggressiv“, „dominant“, „abweisend“ und „aufdringlich“ hoch ein. Diese Beurteilung könnte damit zusammenhängen, daß R1 die Musiktherapeutin in den zu ratenden Sequenzen war. Wie in Kapitel 4.2 beschrieben, erschweren die teilweise abrupten Wechsel im Patientinnenspiel es der Therapeutin, sich auf das Spiel der Patientin einzustellen und somit mit ihr in musikalischen Kontakt zu treten. Die Patientin scheint dabei wenig Rücksicht darauf zu nehmen, was die Therapeutin spielt. Mit diesem Aspekt der Interaktion läßt sich die Beurteilung von R1 gut erklären. Für die Sequenz 6 ergeben sich genau umgekehrte Ratingergebnisse. R1 bewertet das Patientinnenspiel als kontaktfreudig und freundlich, R2 als aggressiv. Das Spiel ist im Gegensatz zu den oben genannten Sequenzen weniger abwechslungsreich, sowohl Tonhöhe als auch Tempo und Lautstärke sind sehr konstant. Für die Therapeutin ist es daher einfacher, das Spiel der Patientin aufzunehmen und noch zusätzliche Variationen hinzuzufügen. Die Patientin ändert ihr Spiel kaum, läßt also das so entstandene Zusammenspiel zu. Auch hier ist also das Rating von R1 durch Aspekte der Interaktion zwischen Patientin und Therapeutin erklärbar. R2 - als außenstehender Beobachter – nimmt diese Aspekte weniger wahr und berücksichtigt bei der Bewertung des Patientinnenspiels Aspekte der Interaktion in geringerem Umfang als R1.

## **5 Diskussion**

Ziel der vorliegenden Arbeit ist es, ein Meßinstrument zur Erfassung musikalischer Improvisation, wie sie in der Musiktherapie stattfindet, zu evaluieren. Da es beabsichtigt ist, mit den hier untersuchten Ratingskalen die Psychopathologie eines Patienten und den Therapieverlauf abbilden zu können, wurde ein Studiendesign verwendet, welches dem normalen Setting der Musiktherapie möglichst nahe kommt. Es wurden daher nicht wie bei Moreau (1996) oder Pechr (1996) experimentelle Musikimprovisationen ausgewertet, sondern Aufzeichnungen aus Musiktherapiesitzungen. Aus dem gleichen Grund wurde auch nicht wie bei Steinberg (1985) und Vanger (1995) nur der auditive Kanal untersucht oder wie in fast allen vorherigen Studien für das Rating eine Selektion von besonders prägnanten oder auffälligen Ausschnitten vorgenommen. Die Ergebnisse dieses Ansatzes werden diskutiert, wobei der Schwerpunkt auf der Interraterreliabilität liegt.

### **5.1 Diskussion der Ergebnisse**

#### **5.1.1 Vergleich mit anderen Ratingskalen**

Die bisher in Studien verwendeten Ratingskalen kann man in folgende drei Kategorien einteilen:

- 1) Beurteilung musikalischer Parameter (MUERB, siehe Pechr (1996))
- 2) Beurteilung des musikalischen Ausdrucks (MAKS, siehe Moreau (1996), Ratingskalen bei Maler (1992))
- 3) Beurteilung des emotionalen Gehalts von Musik (MUSIKOS, siehe Vanger (1995) und Pechr (1996), Adjektivskalen bei Timmermann (1990) und Inselmann und Mann (1999))

Wie in den Arbeiten von Pechr (1996) und Timmermann (1990) gezeigt wurde, gibt es Korrelationen zwischen der Psychopathologie und dem emotionalen Ausdruck der Musikimprovisation einer Person, dagegen konnte eine Korrelation zwischen Psychopathologie und musikalischen Parametern nicht nachgewiesen werden. Es erscheint daher nicht sinnvoll, ausschließlich musikalische Parameter zu bewerten, auch wenn man annehmen kann, daß eine solche Messung bei guter Operationalisierung der Skalen zu höheren Beobachterübereinstimmungen führt als die subjektiveren Ratingskalen zum musikalischen Ausdruck oder zum emotionalen Gehalt von Musik. Daher wurde in der vorliegenden Arbeit die

Messung der musikalischen Parameter nur ergänzend zu der Messung des emotionalen Gehaltes von Musik mit Hilfe von Adjektivlisten vorgenommen. Für die Messung des emotionalen Gehaltes von Musik werden von Vanger (1995) und Pechr (1996) semantische Differentiale verwendet. Diese haben den Vorteil, daß jedes Item bipolar definiert ist. Ob dieser Ansatz zu höherer Raterübereinstimmung bei der Beurteilung von Musik führt, müßte noch in weiteren Studien geklärt werden. In den Arbeiten von Timmermann (1990), Inselmann und Mann (1999) und in der vorliegenden Arbeit ist jedes Item unipolar durch ein Adjektiv definiert. In der vorliegenden Untersuchung erwies sich die Verwendung einer 5-stufigen Skala gegenüber einer visuellen Analogskala für die Rater als Erleichterung des Ratings. Eine unipolare Definierung der Items hat gegenüber einer bipolaren den Vorteil, daß bei ambivalent erscheinender oder wechselnder Spielweise auch zwei entgegengesetzte Items hoch bewertet werden können, wie es in dieser Untersuchung öfters geschah. Gerade bei einer Meßskala, die wie die Adjektivlisten subjektive Kriterien mißt, ist es notwendig, die Interraterreliabilität genau zu untersuchen, bevor man weitere Untersuchungen – z.B. eine Validierung – vornimmt. Da dieses in der Arbeit von Timmermann nicht berücksichtigt wurde, war es Ziel der Arbeit von Inselmann und Mann und besonders der vorliegenden Arbeit, dieses nachzuholen.

#### 5.1.2 Beurteilung der Interraterreliabilität

Wie in der Arbeit von Inselmann und Mann (1999) wurde die Interraterreliabilität für die einzelnen Sequenzen und Ratingphasen mit dem Rangkorrelationskoeffizienten „rho“ nach Spearman bestimmt. Für das Patientinnenspiel war dieser Koeffizient mit  $\rho=0.47$  deutlich unter dem Ergebnis der Arbeit von Inselmann und Mann ( $\rho=0.66$ , siehe Tabelle 1). Überträgt man die bei Faller (1998) genannte Bewertung des Kappa-Koeffizienten zur Bestimmung der Interraterreliabilität auf die Ergebnisse dieser Arbeit, so ist für ein befriedigendes Ergebnis ein Koeffizient von mindestens  $\rho=0.6$  zu fordern. Trotz kontinuierlicher Steigerung der Interraterreliabilität auf Sequenzebene in den drei Ratingphasen konnte kein befriedigendes Ergebnis erreicht werden. Somit erschien es auch nicht sinnvoll, eine Evaluierung der Ratingskalen auf weiteren Ebenen, z.B. eine Validierung, durchzuführen. Auch in anderen Studien waren die Raterübereinstimmungen nicht befriedigend. Der von D. v. Moreau (1996) verwendete Koeffizient Kendalls „tau“ ist, wie in Anhang I gezeigt, gut vergleichbar mit dem in dieser Arbeit verwendeten Rangkorrelationskoeffizienten nach Spearman, wobei die Werte von „tau“ etwas unter den Werten von „rho“ liegen. Bei Moreau betragen die Werte der Interraterreliabilität für die

Szenen durchschnittlich ungefähr  $\tau=0.4$ , wobei die Werte für die einzelnen Szenen und einzelnen Raterpaare teilweise große Abweichungen von diesem Mittelwert zeigen und von  $\tau=-0.7$  bis  $\tau=0.9$  reichen. Diese Ergebnisse von Moreau liegen somit im gleichen Bereich wie die Ergebnisse der vorliegenden Arbeit, obwohl Moreau absichtlich emotionale Aspekte der Musik nicht bewertet, da diese zu subjektiv seien. Im Gegensatz zu der Arbeit von Inselmann und Mann wird in der vorliegenden Arbeit auch die Interraterreliabilität der einzelnen Items berechnet. Diese ist, wie bereits dargestellt wurde, für die Beurteilung des Meßinstrumentes mindestens so wichtig wie die Interraterreliabilität der einzelnen Sequenzen. Der Rangkorrelationskoeffizient nach Spearman für die einzelnen Items liegt bei der Beurteilung des Patientinnenspiels bei durchschnittlich nur  $\rho=0.16$  in der dritten Ratingphase, wobei die einzelnen Werte zwischen  $-0.46$  und  $0.78$  variieren. Zwischen zweiter und dritter Ratingphase zeigte sich insgesamt eine Verschlechterung der Interraterreliabilität für die Items. Da nur das Spiel *einer* Patientin geratet wurde, ist es nicht möglich zu entscheiden, ob die niedrige Übereinstimmung einzelner Items damit zu erklären ist, daß die Items schlecht operationalisiert sind oder ob nur das Spiel dieser einen Patientin bezüglich dieser Adjektive schwer einzuschätzen ist. Die Beurteilung der Items ist auch schwierig, da sie häufig nur geringe oder keine Varianz aufweisen. Bei Moreau lag die Interraterreliabilität der Items bei durchschnittlich  $\tau=0.4$ . Der höhere Wert gegenüber dem der vorliegenden Arbeit ist damit erklärbar, daß bei Moreau Musik von unterschiedlichen Patienten geratet wurde und somit die Unterschiede zwischen den Sequenzen wahrscheinlich ausgeprägter waren. In der Arbeit von Steinberg (1985) wird mit Werten von  $0.87$  bis  $0.96$  für die einzelnen Items eine bedeutend höhere Interraterreliabilität für die Urteile der drei „Experten“ angegeben, leider fehlen Angaben zum statistischen Verfahren. Zudem ist diese Arbeit nur bedingt mit anderen Studien vergleichbar, da sie sich mit Beurteilungen von Rezitationen von Musikstücken und nicht von musikalischen Improvisationen befaßt. In der vorliegenden Arbeit zeigt die Interraterreliabilität des Therapeutinnenspiels, das mit dem gleichen Meßinstrument wie das Patientinnenspiel beurteilt wurde, mit  $\rho=0.70$  eine gute Übereinstimmung. Mit der angewandten Ratingskala ist es also durchaus möglich, testtheoretische Forderungen zu erfüllen. Die schlechteren Ergebnisse bei der Beurteilung des Patientinnenspiels könnten somit v.a. an den bereits geschilderten Problemen in der Einschätzung dieser Patientin, an den geringen Unterschieden zwischen den verschiedenen Sequenzen und den unterschiedlichen Voraussetzungen der Rater liegen. Es erscheint daher sinnvoll, die Adjektivlisten in weiteren Studien zu untersuchen. Dabei sollte man, wie in den Arbeiten von Steinberg (1985), Pechr (1996) oder Moreau (1996), Musiksequenzen von

unterschiedlichen Personen raten oder wie bei Timmermann (1990) nur besonders prägnante Sequenzen eines Patienten beurteilen oder Sequenzen raten, die im Therapieverlauf zeitlich auseinander liegen (Vanger 1995). Mit einem solchen Forschungsansatz wäre es auch bedeutend besser möglich, die einzelnen Items zu bewerten und anhand dieser Ergebnisse die Adjektivskalen zu optimieren.

Die Ratingskalen zur Beurteilung musikalischer Parameter zeigen zu geringe Interrater-übereinstimmungen. Im folgenden Abschnitt wird daher ein Veränderungsvorschlag für diese Skalen dargestellt.

## **5.2 Vorschläge für weitere Untersuchungen**

Der wichtigste Aspekt, der bei einer weiteren Untersuchung berücksichtigt werden sollte, ist schon mehrfach beschrieben und erläutert worden: Die einzelnen zu ratenden Sequenzen sollten sich stärker voneinander unterscheiden als in der vorliegenden Studie.

In den Studien von Steinberg (1985) und Timmermann (1990) wurden die Unterschiede zwischen verschiedenen Ratergruppen als niedrig beschrieben. Die Interraterreliabilität zwischen einzelnen Ratern wurde allerdings nicht berechnet, es wurden lediglich die Mittelwerte der einzelnen Gruppen verglichen. Die Arbeit von Moreau (1996) zeigte hingegen teilweise große Unterschiede in der Interraterreliabilität verschiedener Raterpaare. Wie im Ergebnisteil der vorliegenden Arbeit beschrieben wurde, ist es möglich, Unterschiede im Rating durch die unterschiedlichen Voraussetzungen der Rater zu erklären. Es wäre daher interessant, eine Untersuchung mit mehreren Ratern durchzuführen und zu untersuchen, inwieweit Unterschiede in der Interraterreliabilität verschiedener Raterpaare mit Unterschieden zwischen den Ratern zusammenhängen, ob z.B. die Übereinstimmung der Urteile zweier Laien größer ist als zwischen einem Laien und einem Musiktherapeuten. Hieraus könnte man Erkenntnisse gewinnen, wie eine Ratergruppe optimal zusammengesetzt sein sollte. Ferner wäre es aus theoretischen Gründen notwendig, Beobachter bzw. Rater und zu beobachtende bzw. zu ratende Personen zu trennen, so daß nicht, wie in der vorliegenden Untersuchung, die Musiktherapeutin zugleich Rater ist. Dieser Rater hat somit zusätzlich zu dem im Video Sicht- und Hörbaren weitere Informationen über die zu ratenden Sequenzen, die sein Ratingurteil beeinflussen können. Will man aber untersuchen, ob es einen Zusammenhang zwischen dem Musikspiel einer Person und seiner Psychopathologie gibt, sollten die Rater nicht über zusätzliche Informationen über diese Person verfügen.

Für das Ratertraining ist zu empfehlen, wie in der vorliegenden Untersuchung geschehen, nach einer Ratingphase die Interraterreliabilität der Sequenzen zu bestimmen und Sequenzen mit besonders niedrigen Werten nochmals zu besprechen. Zudem wäre es aber auch sinnvoll, die Interraterreliabilität der Items zu bestimmen und Items mit geringer Übereinstimmung gesondert zu besprechen und eventuell die Operationalisierung der Items zu präzisieren. Anhand der statistischen Ergebnisse der nächsten Ratingphase könnte man erkennen, ob diese Veränderung erfolgreich war.

Die Adjektivlisten sollten mit den in der vorliegenden Arbeit durchgeführten Veränderungen der Operationalisierungen weiter untersucht werden. Dabei sollten auch die Items mit niedriger Interraterreliabilität übernommen werden, da die niedrigen Werte in der vorliegenden Arbeit mit der Spielweise der untersuchten Patientin zusammenhängen können und bei anderen Personen möglicherweise bessere Ergebnisse erzielt werden. In der Beschreibung der Spielweise der Patientin (siehe Kapitel 4.2) werden teilweise Adjektive verwendet, die nicht in der Adjektivliste zur Beurteilung des Patientinnenspiels vorhanden sind. Es wird daher vorgeschlagen, folgende Adjektive (mit semantischer Beschreibung) in die Liste mit aufzunehmen:

- a) lustlos                    gelangweilt, indifferent, ohne Ausdruck (um dieses Adjektiv besser von dem Adjektiv „teilnahmslos“ abzugrenzen, sollte man bei letzterem in der semantischen Beschreibung noch „abwesend“ hinzufügen)
- b) eigensinnig            spielt ihr eigenes Spiel ohne auf den anderen zu achten
- c) ambivalent            widersprüchliche Spielweise, z.B. manchmal Kontakt suchend, manchmal abweisend

Für die Adjektivliste zur Beurteilung der Interaktion wird vorgeschlagen, für das Adjektiv „verschmolzen“ wie für die Adjektive „einig“, „ergänzend“ und „zusammenschwingend“ ein beidseitiges Zusammenspiel zu fordern, da die semantische Bedeutung dieses Adjektivs einen ähnlich hohen Anspruch an das Zusammenspiel impliziert wie „zusammenschwingend“. Wie man an den Items 6 und 11 gesehen hat, war es sinnvoll, genauer zu definieren, welche Voraussetzungen das Zusammenspiel für die verschiedenen Stufen der Skala erfüllen muß. Es muß v.a. für alle Items genau definiert werden, von wem der Kontakt ausgeht. Die Adjektivliste (siehe Anhang H) könnte daher folgendermaßen ergänzt werden:

Item 1-4:            Die Patientin ist nicht in Kontakt mit der Therapeutin, nimmt die Therapeutin nicht wahr.

Ist die Patientin hingegen in Kontakt mit der Therapeutin, sind für die Items 1-3 die Stufen 1 oder 2 und für das Item 4 die Stufen 1-3 zu raten.



- Item 5: Die Patientin nimmt die Therapeutin wahr, spielt aber bewußt etwas Anderes. Nimmt die Patientin die Therapeutin nicht wahr, so ist die Stufe 1 zu raten.
- Item 6-9,11: Die Patientin ist in Kontakt mit der Therapeutin, toleriert Therapeutinnenspiel (Stufen 2 und 3) oder stellt auch selbst Kontakt her (Stufen 4 und 5).
- Item 10,12-14: Beidseitiger Kontakt. Bei nur einseitigem Kontakt ist die Stufe 1 zu raten.

Wie bereits dargestellt, bedürfen die Skalen zur Beurteilung der musikalischen Parameter einer Überarbeitung. Unter Berücksichtigung der in Kapitel 4.1.6 beschriebenen Probleme werden folgende Operationalisierungen vorgeschlagen:

Punktzahl	1	2	3
Rhythmus	kein Rhythmus, arhythmisch, Spiel ohne Betonungen	einfacher Rhythmus (Takt mit Betonungen)	komplexer Rhythmus mit Variationen
Lautstärke	kontinuierlich fast gleiche Lautstärke	leichte Variationen in der Lautstärke	große Lautstärkeunterschiede
Melodik	keine Melodie, chaotische Tonfolge, nur Tonleiter	einfache Melodie	Melodie mit ausgearbeitetem Motiv
Tonumfang	½ Oktave	1 Oktave	gesamter Tonraum
Tempo	gleichbleibendes Tempo	leichte Variationen im Tempo	große Tempounterschiede

Der Ablauf des Ratings hat sich bewährt. Es hat sich gezeigt, daß es, um alle Items einschätzen zu können, notwendig ist, vor dem Ausfüllen jeder Adjektivliste die entsprechende Musiksequenz erneut anzuschauen. Obwohl das Rating der einzelnen Sequenzen somit je nach Länge der Sequenz bis zu 30 Minuten dauern kann, sollte das Verfahren nicht verkürzt werden. Um die Musik einer Sequenz gut einschätzen zu können, erwies sich eine Länge von ungefähr einer Minute als sinnvoll. Wichtiger aber als die Länge ist die Konstanz des Spiels über die Zeit. Es ist daher bei der Auswahl der Sequenzen darauf zu achten, daß sich der emotionale Ausdruck der Musik nicht innerhalb der Sequenz stark ändert.

Es ist zu erwarten, daß bei Umsetzung dieser Vorschläge eine Verbesserung der Objektivität erreicht wird. Werden für diese testtheoretische Forderungen erfüllt, sollten sich weitere Untersuchungen zur Validierung der Ratingskalen anschließen.

## **6 Zusammenfassung**

Obwohl Musik bereits in den ältesten bekannten Therapieverfahren in unterschiedlicher Weise verwandt wurde, steht die moderne Musiktherapieforschung erst am Anfang. Damit aber die Musiktherapie als eigenständige Therapieform allgemein anerkannt wird, ist eine empirisch-wissenschaftliche Fundierung dieser Methode notwendig. In einigen Forschungsarbeiten wurde daher die Möglichkeit untersucht, verschiedene Aspekte des musiktherapeutischen Geschehens mit unterschiedlichen Ratingskalen abzubilden und die daraus gewonnenen Ergebnisse statistisch auszuwerten (siehe Steinberg 1985, 1991, 1992, Maler 1992, Vanger 1995, Pechr 1996, Moreau 1996, Timmermann 1990, 1991, Inselmann und Mann 1999). Von diesen Arbeiten werden jedoch nur bei Moreau die verschiedenen Testgütekriterien (Objektivität, Reliabilität und Validität) systematisch untersucht. Moreau kommt jedoch dabei hinsichtlich des Objektivitätskriteriums nicht zu zufriedenstellenden Ergebnissen. Die Bestimmung der Interraterreliabilität, als Maß für die Objektivität einer Ratingskala, wird in den meisten Arbeiten vernachlässigt, obwohl sie Voraussetzung für die weitere Evaluierung einer solchen Skala ist (Bortz 1984). In der Arbeit von Inselmann und Mann (1999) liegt der Schwerpunkt auf der Untersuchung von Adjektivlisten zum emotionalen Gehalt von Musik, die in veränderter Form aus der Arbeit von Timmermann übernommen wurden. Mit diesen Adjektivlisten soll das Patienten- und Therapeutenspiel sowie die Interaktion zwischen Patient und Therapeut beurteilt werden. Es geht Inselmann und Mann u.a. darum, die Interraterreliabilität für die Ratingskalen zu bestimmen, da diese bei Timmermann nicht berechnet wurde. Inselmann und Mann kommen dabei für das Patientinnen- und Therapeutinnenspiel zu befriedigenden bzw. guten Ergebnissen ( $\rho=0.66$  bzw.  $\rho=0.82$ ). In der vorliegenden Arbeit wurde untersucht, ob diese Ergebnisse reproduzierbar sind. Wie bei Inselmann und Mann sind Videosequenzen aus einer Musiktherapie einer Patientin von zwei Ratern beurteilt worden. Zunächst wurden zwei Proberatingphasen mit nachfolgendem Ratertraining durchgeführt. Dort, wo es sinnvoll erschien, wurden kleinere Veränderungen der Ratingskalen vorgenommen. Trotz Zunahme der Interraterreliabilität für die einzelnen Sequenzen konnte auch in der letzten Ratingphase für das Rating des Patientinnenspiels ( $\rho=0.47$ ), im Gegensatz zum Rating des Therapeutinnenspiels ( $\rho=0.70$ ) und der Interaktion ( $\rho=0.60$ ), kein befriedigendes Ergebnis erreicht werden. Die Unterschiede zwischen den Interraterreliabilitäten der einzelnen

Sequenzen waren dabei sehr groß, weshalb es sinnvoll erschien, die Sequenzen genauer zu untersuchen, um Ursachen für diese Unterschiede herauszufinden. Zusätzlich zu den bei Inselmann und Mann durchgeführten statistischen Untersuchungen wurde in der vorliegenden Arbeit auch die Interraterreliabilität für die einzelnen Items berechnet. Diese ist für die Beurteilung der Ratingskalen ebenso wichtig wie die Interraterreliabilität der einzelnen Sequenzen, weil mit ihr festgestellt werden kann, ob Abweichungen von der üblichen Spielweise bzw. Unterschiede zwischen den Sequenzen von beiden Ratern in gleicher Weise geratet werden. Hierbei ergaben sich insgesamt niedrige Übereinstimmungswerte zwischen den Ratern, zudem zeigte sich keine Verbesserung der Werte zwischen den Ratingphasen. Dieses wird v.a. mit den geringen Unterschieden der einzelnen Sequenzen erklärt. Daher wird empfohlen, in weiteren Studien Musiksequenzen zu raten, die untereinander größere Differenzen aufweisen. Die Skalenkonstruktion von Inselmann und Mann (1999) konnte in der vorliegenden Arbeit in einer Itemanalyse weitgehend bestätigt werden. Bei fünf Items erschien jedoch eine Zuordnung zu anderen Skalen sinnvoll. Hierdurch verbesserte sich die interne Konsistenz der Skalen und auch die semantische Abgrenzung einzelner Skalen voneinander.

Bei der Beschreibung der Sequenzen zeigte sich, daß Sequenzen, die stärker von der üblichen Spielweise abweichen, leichter zu beurteilen sind. Es wurden Eigenschaften des Patientinnenspiels herausgearbeitet, die das Rating erschweren, wie z.B. eine ambivalente oder emotionslose Spielweise, geringe Expressivität und ein einfaches, wenig differenziertes Spiel. Zudem konnten einige wichtige Unterschiede im Ratingverhalten der Rater mit den unterschiedlichen Voraussetzungen der beiden Rater erklärt werden.

Es werden Verbesserungsvorschläge für weitere Studien beschrieben. Diese umfassen die Auswahl der Sequenzen, die Zusammensetzung der Rater, das Vorgehen beim Ratertraining und Veränderungen der Ratingskalen. Mit diesen Veränderungen erscheint es möglich, die für die Objektivität notwendigen testtheoretischen Forderungen zu erfüllen. Dabei wird empfohlen, die statistischen Methoden dieser Arbeit zu übernehmen. Erst wenn eine ausreichend hohe Interraterreliabilität erreicht wird, ist es sinnvoll, auch die anderen Testgütekriterien zu untersuchen.

## Literaturverzeichnis

Ackerknecht, E.H., Murken, A.H. 1992: Geschichte der Medizin. Stuttgart: Enke

Asendorpf, J., Wallbott H.G., 1979: Maße der Übereinstimmung: Ein systematischer Vergleich. Zeitschrift für Sozialpsychologie, 10, 243-252

Balck, F.B., Jantschek, G., Maler, T., Wilke, E., 1987: Veränderungen und Meßbarkeit des improvisierten Spielausdrucks in der klinischen Musiktherapie. In: Lamprecht, F. (Hrsg.) Spezialisierung und Integration in der Psychosomatik und Psychotherapie. Berlin: Springer

Batel, G., 1976: Komponenten musikalischen Erlebens. Eine experimentell-psychologische Untersuchung. Göttinger musikwissensch. Arbeiten Band 7. Göttingen: Kinzel

Bortz, J., 1984: Lehrbuch für empirische Forschung. Berlin, Heidelberg, New York, Tokyo: Springer

Bortz, J., Lienert, G.A., Boehnke, K., 1990: Verteilungsfreie Methoden in der Biostatistik. Berlin, Heidelberg, New York: Springer

Bortz, J., 1993: Statistik für Sozialwissenschaftler. Berlin, Heidelberg, New York: Springer

Cattell, R.B., Cattell, A.K.S., Rhymer, R.M., 1947: P-Technique demonstrated in determinating psycho-physiological source traits in a normal individuel. Psychometr. 12, 4, 267-288

Cohen, J., 1960: A coeffizient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46

Cohen, J., 1968: Weighted kappa: Nominal scale agreement with provision for scale disagreement or partial credit. Psychol. Bull., 70, 213-220

Cronbach, L.J., 1951: Coeffizient alpha and the internal structure of tests. Psychometr., 16, 297- 334

Darwin, C., 1875: Die Abstammung des Menschen und die geschlechtliche Zuchtwahl.  
Stuttgart

Diehl, J.M., Staufenbiel, T. J.M, 1995: Statistik mit SPSS für Windows Version 6.1. Dietmar  
Klotz

Faller, H., Goßler, S., 1998: Probleme und Ziele von Psychotherapiepatienten. Psychother.  
Psychosom. med. Psychol., 48, 176-186

Freud, S., 1914: „Der Moses des Michelangelo“. Ges. Werke X

Harm, T., 1983: Grenzen der musiktherapeutischen Forschung? Ein Bericht einer  
Arbeitstagung in Oldenburg. Musikther. Umschau, 7, 303- 308

Hesse, H.P., 1974: Musikpsychologische Grundlagenforschung und Musiktherapie. In Revers,  
W. J. (Hrsg.) Neue Wege der Musiktherapie, 163-170. Düsseldorf, Wien: Econ

Homer 1979: Odyssee, Übersetzung von R. Hampe. Stuttgart: Reclam

Horowitz, L.M., Strauß, B., Kordy, H., 1994:: Inventar zur Erfassung interpersoneller  
Probleme, Deutsche Version – (IIP-C). Weinheim: Beltz

Inselmann, U., 1995: Musiktherapeutische Behandlung psychotischer Patienten. Zeitschrift  
für klinische Psychologie, Psychopathologie und Psychotherapie, 249 – 260

Inselmann, U., Mann, S., 1999: Emotionales Erleben, Ausdruck und Kommunikation in  
Musikimprovisationen – Eine qualitativ-quantitative Einzelfallanalyse. Psychother.  
Psychosom. med. Psychol., 49, 1-6

Jamblichos 1963: Pythagoras, gr.-dt. Ausg. v. M.v.Albrecht. Zürich

Janke, W., Debus, G., 1971: Die Eigenschaftswörterliste. Göttingen, Toronto, Zürich: Verlag  
für Psychologie Dr. C. J. Hogrefe

Kayser, H., 1984: Akroasis – Die Lehre von der Harmonik der Welt. Basel, Stuttgart: Schwabe & Co. AG

Kearney, H., 1971: Und es entstand ein neues Weltbild. München: Kindler

Kendall, M.G., 1962: Rank correlation method. London: Griffin

Kleinen, G., 1968: Experimentelle Studien zum musikalischen Ausdruck. Unveröffentlichte Dissertation Hamburg

Kneutgen, J., 1974: Einige Voraussetzungen für eine wirkungsvolle Musiktherapie. In: Revers, W. J. (Hrsg.) Neue Wege der Musiktherapie, 17-50. Düsseldorf, Wien: Econ

Kümmel, W.F., 1977: Musik und Medizin. Freiburg, München: Verlag Karl Alber

Lienert, G., Raatz, U., 1994: Testaufbau und Testanalyse. Weinheim: Beltz, Psychologie-Verl.-Union

Luborsky, L., Kächele, H., 1988: Der zentrale Beziehungskonflikt. Ein Arbeitsbuch. Ulm: PSZ

Mahns, W., Ruud, E., 1991: Meta-Musiktherapie: Wege zu einer Theorie der Musiktherapie. Stuttgart, Jena, New York: G. Fischer

Maler, T., 1994: Beschreibungen des musikalischen Ausdrucks und erste Ergebnisse im Lübecker Musiktherapiemodell. Psychothe. Psychosom. med. Psychol., 44, 122-127

Matell, M.S., Jacoby, J., 1971: Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. Educational and psychological measurement, 31, 657- 674

Möller H.-J., 1974: Psychotherapeutische Aspekte in der Musikanschauung der Jahrtausende. In: Revers, W. J. (Hrsg.) Neue Wege der Musiktherapie, 51-160. Düsseldorf, Wien: Econ

Moreau, D.v., 1996: Entwicklung und Evaluation eines Beschreibungssystems (MAKS) zum Ausdrucks- und Kommunikationsverhalten in der Musiktherapie. Unveröffentlichte Diplomarbeit. Würzburg: Psychologisches Institut I der Universität Würzburg

Müller, A., Hörhold, M., Bösel, R., Kage, A., Klapp, B.F., 1994: Einflüsse aktiver Musiktherapie auf Stimmungen und Immunkompetenz psychosomatischer Patienten. Psychologische Beiträge, 36, 198- 204

Neugebauer, L., 1998: Musik als Dialog – eine Untersuchung zu physiologischen Veränderungen während der Musiktherapie. Musikther. Umschau, 19, 29-43

Nitzschke, B., 1984: Frühe Formen des Dialogs. Musikther. Umschau, 5, 167-187

Orlinsky, D.A., Howard, K.I., 1975: Varieties of psychotherapeutic experience. Multivariate analyses of patient's and therapist's reports. New York: Teachers College Press

Osgood, C.E., 1952: The nature and measurement of meaning. Psychol. Bull., 49, 197ff

Overall, J.E., Gorham, D.R., 1962: The brief psychiatric rating scale. Psychol. Rep., 10, 799-812

Pechr, M., 1996: Musikalische und psychologische Parameter in experimenteller Therapiemusik, Depressive und Normalgesunde im Vergleich zweier Meßinstrumente. Musikther. Umschau, 115-128

Platon 1994: Timaios, Sämtl. Werke 4, 11-104, Übersetzung H. Müller. Hamburg: Rowohlt

Priestley, M., 1982: Musiktherapeutische Erfahrungen. Stuttgart, Kassel: Fischer

Rackert, H., 1968: Transference and Countertransference. New York: Int. Univ. Press

Rohrman, B., 1978: Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. Zeitschrift für Sozialpsychologie, 9, 222-245

Strobel, W., Huppmann, G., 1978: Musiktherapie. Grundlagen, Formen, Möglichkeiten. Göttingen, Toronto, Zürich: Verlag für Psychologie Dr.C.J. Hogrefe

Simon, W.C.M., 1974: Musik und Heilkunst. In: Revers, W. J. (Hrsg.) Neue Wege der Musiktherapie, 9-16, Düsseldorf, Wien: Econ

Steinberg, R., Raith, L., 1985a: Music Psychopathologie, 1: Musical Tempo and Psychiatric Disease. Psychopathologie, 18, 254- 264

Steinberg, R., Raith, L., 1985b: Music Psychopathologie, 2. Assessment of Musical Expression. Psychopathologie, 18, 265- 273

Steinberg, R., Raith, L., Rossnagel, G., Eben, E., 1985c: Music Psychopathologie, 3. Musical Expression and Psychiatric Disease. Psychopathologie, 18, 274- 285

Steinberg, R., 1991: Music Psychopathologie, 4. The Course of Musical Expression during Music Therapy with Psychiatric Inpatients. Psychopathologie, 24, 121-129

Steinberg, R., Fani, M., Raith, L., 1992a: Music Psychopathologie, 5. Objective Features of Instrumental Performances and Psychopathology. Psychopathologie, 25, 266-272

Steinberg, R., Fani, M., Raith, L., 1992b: Music Psychopathologie, 6. The Course of Objektive Instrumental Performance Characteristics with Psychiatric Inpatients. Psychopathologie, 25, 273-282

Thaut, M.H., 1997: Wer und was fordert die Musiktherapie in der Zukunft? Wissenschaftstheoretische Grundlagen, Forschungsmodelle und therapeutische Anwendungen. Musiktherap. Umschau, 18, 15-24

Timmermann, T., 1990: Der Musikalische Dialog – Beziehungsmuster im Spiegel bedeutsamer Momente im Therapieprozeß, Eine musiktherapeutische Einzelfallstudie. Unveröffentlichte Dissertation. Ulm: Fakultät für theoretische Medizin der Universität Ulm



Timmermann, T., Scheytt-Hölzer, N., Bauer, S., Kächele, H., 1991: Musiktherapeutische Einzelfall-Prozeßforschung – Entwicklung und Aufbau eines Forschungsfeldes. *Psychother. Psychosom. med. Psychol.*, 41, 385-391

Timmermann, T., 1999: Musik und Heilung – in den alten Kulturen und in der modernen Musiktherapie. *Musikther. Umschau*, 20, 123- 145

Tinsley, H.E.A., Weiss, D.J., 1975: Interrater Reliability and Agreement of Subjective Judgments. *Journal of Counseling Psychology*, 22, 4, 358-376

Tischler, B., 1983: Ist Musiktherapie empirisch begründbar? *Musikther. Umschau*, 4, 95-106

Vanger, P., Oerter, U., Otto, H., Schmidt, S., Czogalik, D., 1995: The musical expression of the separation conflict during music therapy: a single case study of a crohn´s disease patient. *The Arts in Psychotherapy*, 22, 147-154

Watzlawick, P., Beavin, J.H., Jackson, D.D., 1996: *Menschliche Kommunikation: Formen, Störungen, Paradoxien*. Bern, Göttingen, Toronto, Seattle: Huber

**Anhang A: Bereiche und Subskalen der EWL**

Bereiche nach Janke und Debus	Subskalen nach Janke und Debus	Items der vorliegenden Arbeit
1) Leistungsbezogene Aktivität	A) Aktiviertheit	agil, energisch
	B) Konzentriertheit	-
2) Allgemeine Desaktivität	C) Desaktiviertheit	teilnahmslos, passiv
	D) Müdigkeit	-
	E) Benommenheit	-
3) Extroversion/Introversion	F) Extrovertiertheit	offen, kontaktfreudig
	G) Introvertiertheit	abgekapselt
4) Allgemeines Wohlbehagen	H) Selbstsicherheit	selbstsicher, unbekümmert
	I) Gehobene Stimmung	froh, beschwingt
5) Emotionale Gereiztheit	J) Erregtheit	-
	K) Empfindlichkeit	-
	L) Ärger	verärgert, aggressiv
6) Angst	M) Ängstlichkeit	bekommen, ängstlich, zaghaf
	N) Deprimiertheit	wehmütig, traurig, mutlos, hilflos, verstört
	O) Verträumtheit	-

**Anhang A:** Bereiche und Subskalen der Eigenschaftswörterliste nach Janke und Debus und zugehörige Items, die in die Adjektivlisten der vorliegenden Arbeit übernommen wurden

**Anhang B: Skalen aus der Arbeit von Inselmann und Mann**

<u>Skala 1: Selbstbewußt:</u> ( $\alpha = 0.84$ )	2) selbstsicher 20) agil	5) unbekümmert 21) gelöst	7) energisch 22) dominant
<u>Skala 2: Depressiv:</u> ( $\alpha = 0.78$ )	1) abgekapselt 8) traurig 13) passiv 25) introvertiert 30) mutlos 34) bekloffen	4) hilflos 10) ängstlich 17) starr 26) selbstunsicher 32) wehmütig 35) bedroht	6) zaghaft 11) zurückgezogen 19) teilnahmslos 27) nachgiebig 33) verstört
<u>Skala 3: Harmonisch:</u> ( $\alpha = 0.88$ )	14) friedlich	28) freundlich	31) ausgeglichen
<u>Skala 4: Aggressiv:</u> ( $\alpha = 0.85$ )	12) aggressiv 24) abweisend	16) verärgert 29) aufdringlich	23) konkurrierend
<u>Skala 5: Freudig zugewandt:</u> ( $\alpha = 0.76$ )	3) froh 18) offen	9) beschwingt	15) kontaktfreudig

**Anhang B:** Skalen der Arbeit Inselmann und Mann (1999). Das angegebene „ $\alpha$ “ nach Cronbach gibt die interne Konsistenz der Skalen an. Sie wurde berechnet aus dem Rating beider Rater für das Therapeutinnen- und Patientinnenspiel der letzten 10 Sequenzen der vorliegenden Arbeit.

**Anhang C: Itemanalyse für die Skalen aus der Arbeit von Inselmann und Mann****Skala 1**

	(SKALA 1)	SKALA 2	SKALA 3	SKALA 4	SKALA 5	SKALA 1-ITEM
Item 2	,8240**	-,8128**	,1819	,3152*	,4592**	<b>0.7461</b>
Item 5	,8290**	-,6463**	,3698*	,2133	,4730**	<b>0.7643</b>
Item 7	,7907**	-,4775**	-,3388*	<b>,6986**</b>	-,0018	0.6477
Item 20	,8644**	-,6848**	,1547	,3375*	,3851*	<b>0.7862</b>
Item 21	,5370**	-,6717**	,7107**	-,2220	<b>,7595**</b>	0.4016
Item 22	,7186**	-,2653	-,2723	<b>,7423**</b>	-,1843	0.5225

**Skala 2**

	SKALA 1	(SKALA 2)	SKALA 3	SKALA 4	SKALA 5	SKALA 2-ITEM
Item 1	-,1090	,5622**	-,4499**	,3108	-,7006**	<b>0.4564</b>
Item 4	-,6336**	,7515**	-,3204*	-,1547	-,5864**	<b>0.6770</b>
Item 6	-,7771**	,5744**	,0570	-,4629**	-,2151	<b>0.4026</b>
Item 10	-,0619	,3307*	-,1943	,2389	-,2766	<b>0.3034</b>
Item 11	-,2966	,6596**	-,3785*	,1795	-,6287**	<b>0.5613</b>
Item 13	-,2730	,4160**	-,2202	-,0123	-,4710**	<b>0.3489</b>
Item 17	-,1123	,6253**	-,4116**	,2941	-,5611**	<b>0.4947</b>
Item 19	-,2344	,4294**	-,4413**	,3197*	-,6209**	<b>0.3289</b>
Item 25	-,6677**	,8566**	-,2379	-,2217	-,5180**	<b>0.8073</b>
Item 26	-,6223**	,6258**	-,1025	-,3308*	-,3206*	<b>0.5434</b>
Item 27	-,1985	-,0401	,2750	-,3595*	<b>,4244**</b>	-0.2542
Item 30	-,4866**	,7343**	-,2836	-,1585	-,4285**	<b>0.6762</b>
Item 33	-,3856*	,5769**	-,2593	-,1287	-,2766	<b>0.5362</b>
Item 34	-,4545**	,7518**	-,2972	-,0845	-,3854*	<b>0.7061</b>
Item 35	-,0608	,0977	-,3715*	<b>,2671</b>	-,3512*	0.0545

**Skala 3**

	SKALA 1	SKALA 2	(SKALA 3)	SKALA 4	SKALA 5	SKALA 3-ITEM
Item 14	-,0160	-,1955	,9084**	-,6508**	,6318**	<b>0.7665</b>
Item 28	,0903	-,4114**	,9227**	-,6543**	,7709**	<b>0.8162</b>
Item 31	,1527	-,4457**	,8739**	-,5711**	,7313**	<b>0.7519</b>

**Skala 4**

	SKALA 1	SKALA 2	SKALA 3	(SKALA 4)	SKALA 5	SKALA 4-ITEM
Item 12	,6181**	-,2908	-,6218**	,9204**	-,2704	<b>0.8314</b>
Item 16	,3824*	-,0561	-,6002**	,8113**	-,4142**	<b>0.7144</b>
Item 23	<b>,5015**</b>	-,2235	-,1864	,5896**	,0946	0.4726
Item 24	,0382	,2715	-,6340**	,7100**	-,6353**	<b>0.5713</b>
Item 29	,5400**	-,1589	-,6288**	,8892**	-,3471*	<b>0.7887</b>

**Skala 5**

	SKALA 1	SKALA 2	SKALA 3	SKALA 4	(SKALA 5)	SKALA 5-ITEM
Item 3	,2797	-,4115**	<b>,7596**</b>	-,4186**	,7333**	0.5064
Item 9	,4403**	-,5444**	<b>,7092**</b>	-,2653	,7806**	0.5844
Item 15	,1493	-,5299**	,4175**	-,2380	,7331**	<b>0.4919</b>
Item 18	,1077	-,5793**	,5031**	-,3312*	,8084**	<b>0.6549</b>

Signifikanzniveau: \*<0.05, \*\*<0.01

**Anhang C:** Angegeben sind die Korrelationskoeffizienten zwischen einzelnen Items und den 5 aus der Arbeit von Inselmann und Mann vorgegebenen Skalen. Die letzte Spalte gibt den Korrelationskoeffizienten zwischen Item und der Summe der restlichen Items der entsprechenden Skala an. Fett gedruckt ist jeweils die größte Korrelation pro Item.

**Anhang D: Itemanalyse für die veränderten Skalen****Skala 1**

	(SKALA 1)	SKALA 2	SKALA 3	SKALA 4	SKALA 5	SKALA 1-ITEM
Item 2	,7943**	-,8003**	,3039	,2811	,2875	<b>0.7045</b>
Item 5	,7735**	-,6393**	,4822**	,1689	,2169	<b>0.6916</b>
Item 7	,8203**	-,3788*	-,1502	<b>,7033**</b>	-,2240	0.6908
Item 20	,8296**	-,6484**	,3317*	,3064	,0836	<b>0.7342</b>
Item 22	,7877**	-,1514	-,1372	<b>,7121**</b>	-,4376**	0.6331
<b>Item 23</b>	,6015**	-,2223	-,0668	,4726**	,0902	<b>0.5146</b>

**Skala 2**

	SKALA 1	(SKALA 2)	SKALA 3	SKALA 4	SKALA 5	SKALA 2-ITEM
Item 1	-,0325	,6602**	-,4550**	,3575*	-,7777**	<b>0.5756</b>
Item 4	-,5870**	,7982**	-,4266**	-,1207	-,4766**	<b>0.7397</b>
Item 6	-,7825**	,4988**	-,1167	-,4529**	,0983	<b>0.3193</b>
Item 10	-,0420	,3665*	-,2133	,2693	-,2492	<b>0.3408</b>
Item 11	-,2126	,7287**	-,4245**	,1999	-,6243**	<b>0.6487</b>
Item 13	-,2431	,5009**	-,2895	,0134	-,4967**	<b>0.4416</b>
Item 17	-,0267	,6356**	-,4326**	,3136*	-,4308**	<b>0.5135</b>
Item 19	-,1761	,5423**	-,4825**	,3738*	-,6376**	<b>0.4569</b>
Item 25	-,6087**	,8657**	-,3484*	-,2037	-,3921*	<b>0.8214</b>
Item 26	-,5879**	,6341**	-,1947	-,3266*	-,2591	<b>0.5560</b>
Item 30	-,4286**	,6694**	-,3282*	-,1551	-,1775	<b>0.5999</b>
Item 33	-,3288*	,5164**	-,2875	-,1318	-,0525	<b>0.4716</b>
Item 34	-,3909*	,6905**	-,3383*	-,0790	-,1408	<b>0.6353</b>
Item 35	-,0015	,1600	-,3587*	<b>,3037</b>	-,3571*	<b>0.1199</b>

**Skala 3**

	SKALA 1	SKALA 2	(SKALA 3)	SKALA 4	SKALA 5	SKALA 3-ITEM
<b>Item 3</b>	,1727	-,4003*	,8658**	-,4551**	,1709	<b>0.8006</b>
<b>Item 9</b>	,3408*	-,5606**	,8405**	-,3027	,2918	<b>0.7659</b>
Item 14	-,1294	-,2179	,8696**	-,6827**	,2403	<b>0.7912</b>
<b>Item 21</b>	,3948*	-,6955**	,8104**	-,2709	,4417**	<b>0.7473</b>
Item 28	-,0449	-,4710**	,8844**	-,6960**	,5200**	<b>0.8209</b>
Item 31	,0139	-,4932**	,8453**	-,5796**	,4510**	<b>0.7792</b>

**Skala 4**

	SKALA 1	SKALA 2	SKALA 3	(SKALA 4)	SKALA 5	SKALA 4-ITEM
Item 12	,6941**	-,2124	-,4555**	,9230**	-,2641	<b>0.8226</b>
Item 16	,4662**	,0082	-,5004**	,8319**	-,3477*	<b>0.7329</b>
Item 24	,1501	,3664*	-,6155**	,7433**	-,5774**	<b>0.6013</b>
Item 29	,6474**	-,0888	-,5137**	,8765**	-,2800	<b>0.7505</b>

**Skala 5**

	SKALA 1	SKALA 2	SKALA 3	SKALA 4	(SKALA 5)	SKALA 5-ITEM
Item 15	,0928	-,6483**	,3909*	-,2838	,9079**	<b>0.7920</b>
Item 18	,0240	-,6692**	,5067**	-,3743*	,8382**	<b>0.6948</b>
<b>Item 27</b>	-,2345	-,2542	,2220	-,3976*	,8592**	<b>0.6001</b>

Signifikanzniveau: \*<0.05, \*\*<0.01

Anhang D: Erläuterung siehe Anhang C. Die fett gedruckten Items sind die neu zugeordneten Items.

**Anhang E: Instrumente und Dauer der Sequenzen 3-12**

Sequenz	IRR	Instrumente Pat. / Ther.	Dauer
9	0.87**	Klangschale / Bongos	1'05''
5	0.63**	Metallxylophon / Klavier	1'45''
11	0.81**	Bongos, Tambourin / Bongos	1'16''
8	0.70**	Bongos / Bongos	3'59''
4	0.56**	Holzxylophon / Klavier	1'09''
10	0.54**	Holzxylophon / Klavier	1'18''
3	0.24	Bongos, Klangschale / Klavier	0'35''
12	0.27	Klangschale / Triangel	1'08''
7	0.16	Bongos, Klangschale / Klavier	1'43''
6	-0.15	Metallxylophon / Klavier	0'58''

Signifikanzniveau: \* $<0.05$ , \*\* $<0.01$

**Anhang E:** Musikinstrumente der Patientin und Therapeutin sowie Dauer der einzelnen Sequenzen. Die Sequenzen sind nach Höhe der Raterübereinstimmung auf Skalenebene sortiert.

**Anhang F: Interraterreliabilität der Skalen**

Skala	Patientin	Therapeutin
1) selbstbewußt	0.49	0.90
2) depressiv	0.08	0.68
3) harmonisch/freundlich	-0.21	0.24
4) aggressiv	0.34	0.88
5) zugewandt	-0.09	-0.30

**Anhang F:** Rangkorrelationskoeffizient nach Spearman der einzelnen Skalen für die Raterurteile der letzten 10 Sequenzen bezüglich des Patientinnen- und Therapeutinnenspiels

## Anhang G: Semantische Beschreibung der Adjektive der Adjektivlisten zur Beschreibung des Patienten- und Therapeutespiels

<b>abgekapselt</b>	aktiv sich schützend, sich isolieren, abschotten, aktiv zurückgezogen, da bedrohlich, einschüchternd, ängstigend, bedrängend, aufdringlich
<b>selbstsicher</b>	sehr klares, deutliches, sicheres Spiel, energisch oder differenziert, <i>muß nicht dominant sein</i>
<b>froh</b>	gute Stimmung, freudig, lebendig (Gegensatz zu depressiv)
<b>hilflos</b>	irritiert, verunsichert, „was kann man tun?“, mit wenig Ausdruck, monoton
<b>unbekümmert</b>	unbefangen, sorglos, nicht gehemmt, frei
<b>zaghaf</b>	vorsichtig, ausprobierend, abwartend, schüchtern, behutsam, ängstlich
<b>energisch</b>	mit Kraft, deutlich, durchsetzend, klar
<b>traurig</b>	deprimiert, niedergeschlagen, gedämpft (nicht im Sinne von „unterordnen“, sondern im Affekt gedämpft), <i>sollte deutlich erkennbar sein, nicht nur unbeteiligtes Spiel</i>
<b>beschwingt</b>	fröhlich, mit Elan, energisch, abwechslungsreich
<b>ängstlich</b>	
<b>zurückgezogen</b>	sich entfernen, distanzieren
<b>aggressiv</b>	angriffslustig, eher laut, vereinnahmend
<b>passiv</b>	über sich ergehen lassen, „runterleiern“, unbeteiligt
<b>friedlich</b>	entspannt, harmonisch, [etwas freudig], zufrieden, <i>nicht gleichbedeutend mit freundlich</i>
<b>kontaktfreudig</b>	auf jemanden zugehen, offen, Angebote annehmen, Angebote machen
<b>verärgert</b>	ungehalten, aggressiv, zornig
<b>starr</b>	statisch, ohne Entwicklung, monoton (nicht auf Interaktion bezogen)
<b>offen</b>	aufnehmend, bereit, zugewandt
<b>teilnahmslos</b>	interesselos
<b>agil</b>	tätig, aktiv, <i>lebendig</i>
<b>gelöst</b>	offen, locker, ausgeglichen, entspannt
<b>dominant</b>	im Vordergrund stehen, beherrschen, sich abheben, <i>setzt Kontakt voraus; ist nicht hoch zu raten, wenn Spielende nur vor sich hin spielt</i>
<b>konkurrierend</b>	übertrumpfen wollen, mitstreiten, wettstreiten
<b>abweisend</b>	wegschieben, fernhalten, jemanden zurückweisen

---

<b>introvertiert</b>	in sich verschlossen, nicht aus sich rausgehen (nicht: nachdenklich, kontemplativ)
<b>selbstunsicher</b>	
<b>nachgiebig</b>	sich anpassen, sich unterordnen
<b>freundlich</b>	offen, zugewandt, locker
<b>aufdringlich</b>	eindringlich, angreifen, jemanden unangenehm bedrängen
<b>mutlos</b>	resigniert, aufgebend
<b>ausgeglichen</b>	harmonisch, in sich ruhend, <i>beschwingt, nicht zerrissen</i>
<b>wehmütig</b>	schweremütig, traurig, trübselig, <i>sollte deutlich erkennbar sein, nicht nur unbeteiligt</i>
<b>verstört</b>	[kann nicht so, wie sie will; weiß nicht, was sie will], durcheinander, zerrissen, irritiert, verwirrt
<b>beklommen</b>	befangen, eingeschüchtert
<b>bedroht</b>	das Spiel ist, als fühle er (Therapeut, Patient) sich bedroht

**Anhang G:** Die Beschreibung der Adjektive wurde weitgehend aus der Arbeit von Inselmann und Mann übernommen. Kursiv Geschriebenes wurde während des Ratertrainings der vorliegenden Untersuchung ergänzt, in eckige Klammern Gesetztes gestrichen. Diese Änderungen wurden vorgenommen, wenn die Beschreibungen nicht eindeutig genug waren und es unterschiedliche Auffassungen zwischen den Ratern über die Bedeutung eines Adjektivs gab.



## Anhang H: Semantische Beschreibung der Adjektive der Adjektivliste zur Beschreibung der Interaktion

*Geordnet nach aufsteigenden Gemeinsamkeitswerten*

- 1) **isoliert** (vgl. „abgekapselt“ aus der Adjektivliste), vereinzelt, „für sich“, (aktiv) zurückgezogen
- 2) **zerrissen** Widersprüche im Zusammenspiel, starke Steigerung von „getrennt“, nichts Gemeinsames finden, chaotisch, (aktiver Prozeß)
- 3) **getrennt** nicht zusammenkommend
- 4) **nebeneinander** die Verbindung nicht finden, nicht richtig in Kontakt, passiverer Prozeß als „getrennt“
- 5) **gegenläufig** gegeneinander, Gegensätzliches, aktiver Prozeß; Voraussetzung ist, daß das Spiel des anderen wahrgenommen wird
- 6) **in Kontakt** neutrale Bedeutung, Kontakt zulassen (kann hoch sein, auch wenn andere Gemeinsamkeitswerte niedrig sind)
- 7) **entspannt** ausgeglichene, lockere Atmosphäre, bei der man Entfaltungsmöglichkeiten im Zusammenspiel hat
- 8) **gemeinsam** Zusammenspiel, gemeinsame Elemente in irgendeiner Form (keine hohe Voraussetzung)
- 9) **harmonisch** gemeinsam, muß nicht besonders ausgearbeitet sein, hört sich gut an
- 10) **verschmolzen** ganz gleich, gemeinsam, symbiotisch, auch wenn sich nur einer auf den anderen (maximal) einstellt
- 11) **zugewandt** muß nicht von beiden ausgehen; bereits wenn einer zugewandt ist, kann ein niedriger/mittlerer Wert vergeben werden, sich auf den anderen einlassen, auf jemanden zugehen
- 12) **einig** beide stimmen dem Spiel des anderen zu, gegenseitige Akzeptanz (nicht: nur das Spiel des anderen tolerieren)
- 13) **ergänzend** das Spiel des anderen wahrnehmen und etwas Eigenes/Passendes hinzufügen
- 14) **zusammen-schwingend** hoher Anspruch, beide nehmen das Spiel des anderen wahr und ergänzen/erweitern es auf bewußt harmonischer Basis

**Ergänzungen nach der ersten Ratingphase:**

Adjektive 1-4: kein Kontakt vorhanden, Adjektive 1 und 2: aktiver Prozeß; Adjektive 3 und 4: passiver Prozeß

Adjektive 5-11: setzt Kontakt von mindestens einem Spieler voraus, Adjektive 6 und 11: geht Zusammenspiel nur von einem Spieler aus, darf maximal bis Stufe 3 geratet werden

Adjektive 12-14: setzt Kontakt von beiden Spielern voraus, beide spielen zusammen; ist nur einer der Spieler mit dem anderen in Kontakt, muß Stufe 1 geratet werden

**Anhang H:** Die Beschreibung der Adjektive wurde weitgehend aus der Arbeit von Inselmann und Mann übernommen. Die Ergänzungen nach der ersten Ratingphase wurden vorgenommen, da die Operationalisierung einiger Items nicht eindeutig genug erschien.

### Anhang I: Vergleich verschiedener statistischer Verfahren zur Berechnung der Interraterreliabilität

Sequenz	Spearman	Kendall	Pearson
3	0.24	0.21	0.26
4	0.56	0.49	0.51
5	0.63	0.56	0.58
6	-0.15	-0.13	-0.21
7	0.16	0.13	0.18
8	0.70	0.61	0.65
9	0.87	0.82	0.87
10	0.54	0.46	0.51
11	0.81	0.72	0.81
12	0.27	0.24	0.26
Mittel	0.47	0.41	0.44

**Anhang I:** Vergleich des Rangkoeffizienten („rho“) nach Spearman mit Kendalls „tau“ und dem Pearson- Koeffizienten für das Patientinnenspiel der dritten Ratingphase

**Anhang J: Interraterreliabilität der Items in der zweiten und dritten Ratingphase**

Nr.	Item	Skala	Spearman Patientinnenspiel	
			Rating- phase 2	Rating- phase 3
2	selbstsicher	1	0.40	0.53
5	unbekümmert	1	0.63	0.15
7	energisch	1	0.50	0.14
20	agil	1	0.14	0.03
22	dominant	1	0.72*	0.78**
23	konkurrierend	1	0.33	-
1	abgekapselt	2	0.47	0.19
4	hilflos	2	0.09	0.42
6	zaghaft	2	0.27	0.47
8	traurig	2	-0.11	-
10	ängstlich	2	-	-
11	zurückgezogen	2	0.57	0.22
13	passiv	2	-	0.50
17	starr	2	0.53	0.44
19	teilnahmslos	2	0.71*	0.06
25	introvertiert	2	0.07	0.21
26	selbstunsicher	2	-	0.53
30	mutlos	2	0.27	-
32	wehmütig	2	-	-
33	verstört	2	0.38	-
34	bekloffen	2	-	-
35	bedroht	2	-	-
3	froh	3	-0.18	-0.14
9	beschwingt	3	-0.51	-0.23
14	friedlich	3	-0.32	-0.23
21	gelöst	3	0.20	-0.46
28	freundlich	3	-0.34	-0.08
31	ausgeglichen	3	-0.16	0.14
12	aggressiv	4	0.39	0.41
16	verärgert	4	0.25	0.15
24	abweisend	4	0.52	-0.05
29	aufdringlich	4	0.56	0.32
15	kontaktfreudig	5	0.68*	-0.01
18	offen	5	0.64	-0.21
27	nachgiebig	5	0.57	0.00
	arithmetisches Mittel		0.29	0.16

Signifikanzniveau: \* < 0.05, \*\* < 0.01

**Anhang J:** Vergleich der Rangkorrelationskoeffizienten nach Spearman der Items für die zweite und dritte Ratingphase

## Anhang K: Fiktives Beispiel zur Verdeutlichung der Unterschiede zwischen der Interraterreliabilität für Sequenzen und Items

**Sequenz 1: Rater1**

	Stufe 1	Stufe 2	Stufe 3
Item 1	X		
Item 2			X
Item 3		X	

**Sequenz 1: Rater2**

	Stufe 1	Stufe 2	Stufe 3
Item 1	X		
Item 2			X
Item 3		X	

**Sequenz 2: Rater 1**

	Stufe 1	Stufe 2	Stufe 3
Item 1	X		
Item 2		X	
Item 3		X	

**Sequenz 2: Rater 2**

	Stufe 1	Stufe 2	Stufe 3
Item 1	X		
Item 2			X
Item 3		X	

**Sequenz 3: Rater 1**

	Stufe 1	Stufe 2	Stufe 3
Item 1	X		
Item 2			X
Item 3	X		

**Sequenz 3: Rater 2**

	Stufe 1	Stufe 2	Stufe 3
Item 1	X		
Item 2		X	
Item 3		X	

**Sequenz 4: Rater 1**

	Stufe 1	Stufe 2	Stufe 3
Item 1	X		
Item 2			X
Item 3		X	

**Sequenz 4: Rater 2**

	Stufe 1	Stufe 2	Stufe 3
Item 1	X		
Item 2			X
Item 3	X		

Rangkorrelationskoeffizient nach Spearman für die Sequenzen 1-4: 1.00, 0.87, 0.50, 0.87

Rangkorrelationskoeffizient nach Spearman für die Items: Item 1: nicht berechenbar (fehlende Varianz), Item 2 und 3: je -0.33

**Anhang K:** Fiktives Beispiel für 4 Sequenzen, die sich nur wenig voneinander unterscheiden. Geratet wurden 3 unterschiedliche Items (z.B. Adjektive) auf einer 3-stufigen Skala. Beide Rater raten die einzelnen Sequenzen ähnlich (hohe Interraterreliabilität für die einzelnen Sequenzen). Die Unterschiede zwischen den Sequenzen werden jedoch unterschiedlich eingeschätzt (niedrige Interraterreliabilität für die einzelnen Items). Zur Berechnung der Interraterreliabilität siehe auch S. 33.

**Anhang L: Interraterreliabilität der Sequenzen auf Item- und Skalenebene**

Sequenz	Spiel der Patientin		Spiel der Therapeutin	
	nach Items	nach Skalen	nach Items	nach Skalen
3	0.24	-0.05	0.78**	0.90*
4	0.56**	0.20	0.79**	1.00**
5	0.63**	1.00**	0.84**	0.97**
6	-0.15	-0.80	0.65**	0.60
7	0.16	-0.72	0.63**	0.70
8	0.70**	0.35	0.75**	0.70
9	0.87**	1.00**	0.78**	0.90*
10	0.54**	0.10	0.86**	0.87
11	0.81**	0.79	0.55**	0.90*
12	0.27	-0.60	0.32	0.46
3-12	0.47**	0.39**	0.70**	0.82**

Signifikanzniveau: \* $<0.05$ , \*\* $<0.01$

**Anhang L:** Rangkorrelationskoeffizient nach Spearman für die Sequenzen berechnet auf der Grundlage der Itemwerte und der Skalenwerte der einzelnen Sequenzen.