9. Interpretation der Ergebnisse

Die Ergebnisse aus den Experteninterviews werden nun mit den Resultaten des quantitativen Strangs zur Beantwortung der qualitativen Forschungsfrage kombiniert, ob die Entwicklungen der Hörfunkprogramme auf Planung oder Zufall beruhen. Dabei wird die Abfolge der quantitativen Untersuchung übernommen. Zunächst werden Struktur und Präsentation verglichen, danach folgen Wortinhalt und Musikprogramm. In Kapitel 9.4. erfolgt die Zusammenfassung und Ergebnisfindung.

9.1 Struktur und Präsentation

Die Auswertung der Programmstruktur hat bei der Berechnung der Relativen Entropie ergeben, dass sich die Sender im Tages- als auch im Wochenablauf überwiegend ähnlich waren. Dies wurde von den Programmdirektoren auch bestätigt, da diese zwischen dem normalen Tagesprogramm, dem Abend- sowie und Wochenendprogramm unterscheiden. Vor allem zu den Randzeiten trauten sich die Sender etwas mehr zu und gingen weg von ihrem Tagesformat, indem sie Spezialsendungen mit Disc Jockeys oder anderen musikalischen Färbungen wie Classic Rock im Programm hatten oder auch auf unkonventionellere Moderatoren wie Matthias Matuschik setzten. Auch hat die Analyse der Playlists ergeben, dass am Wochenende längere Musikstrecken als unter der Woche durchaus üblich waren, was ebenfalls von den Experteninterviews bestätigt worden ist. Da sich die Hörer am Wochenende in einer anderen Realität als unter der Woche befinden und entspannter sind, weshalb sie weniger Wort- und Informationsanteil, sondern mehr Musik wünschen. Sowohl aus der Analyse der Playlists als auch aus der Auswertung der Programmschemata lässt sich in Kombination mit den Interviews erkennen, dass die Sender ganz gezielt ihr Programm ändern. Eine zufällige Entwicklung kann an dieser Stelle ausgeschlossen werden und war auch nicht zu erwarten.

Ähnlich verhält es sich auch beim Einsatz von Blitzermeldungen. Während bei ARD-Anstalten aus rechtlichen Gründen diese nicht gelesen werden dürfen, und Programmdirektoren die Sinnhaftigkeit auf landesweiter Ebene infrage stellen, sehen Privatsender darin eine Serviceleistung. Dieses Serviceelement dient aber weniger dem eigentlichen Warnhinweis, als

vielmehr dem Aufbau von regionaler Kompetenz, durch die die Sender unterschwellig versuchen, in einer Region verwurzelter zu klingen. Auch hier handelte es sich um bewusste Steuerungen der Programmverantwortlichen.

Fast die gleiche Argumentationsweise gilt bei den Privatsendern beim Einsatz von Hörerbeteiligung. Kommerzielle Sender wollen dadurch nahe am Hörer sein und regional kompetenter wirken. Bei den öffentlich-rechtlichen Sendern werden Hörerinteraktionen zwar als positiv empfunden, allerdings sollten diese das Programm bereichern und nicht gesendet werden, wenn die Zuhörer dadurch keinen Mehrwert haben. Die Entscheidung, wie viel Interaktion ein Sender mit seinen Hörern betrieben hat, war also nicht zufällig, sondern unterlag auch den Entscheidungen der Sendeleiter.

Die Wirkung der Nennung des eigenen Sendernamens und der Einsatz der Verpackungselemente sind den Programmachern durchaus bewusst. Die Interviews haben ergeben, dass die Verantwortlichen sehr gezielt ihren eigenen Sender im Rahmen der On-Air-Promotion nennen. Im Regelfall wurden Promos in der Sendeuhr geplant. Dennoch ergaben sich aus den Äußerungen einige Ergebnisse, die nicht unbedingt schlüssig waren. Peter Bartsch von Radio Arabella gab an, dass bewusst weniger Verpackungselemente eingesetzt und die Titel durchaus „back to back“ gespielt wurden, da die Arabella-Hörer zu viele IDs im Programm ablehnen. Diese Aussage ist korrekt und lässt sich auch anhand der Zahlen der eingesetzten Verpackungselemente nachweisen. Allerdings spielte Radio Arabella bewusst einen Musikmix mit großen Sprüngen, die durch Verpackungselemente kaschiert werden sollten. Eine Vermutung ist, dass sich aufgrund der geringen Anzahl von eingesetzten Verpackungselementen die großen Musiksprünge im Programm bemerkbar machen, und hier Hörerverluste die Folge sein könnten, gerade wenn man von einem 50:50 Mix aus Oldies und aktuellen Titeln ausgeht. Falls auf den verstärkten Einsatz von Verpackungselementen verzichtet werden soll, dann müsste die Musik harmonischer geplant und so große Sprünge vermieden werden. Auch Karsten Krögers Äußerung, dass er einen zu hohen Namensdruck kritisch sieht, konnte aus dem Programm abgelesen werden. bigFM hatte zwar die meisten Verpackungselemente, aber auch die wenigsten IDs. Alina Schröders und Stefanie Schäfers Meinung, dass die Zunahme bei DASDING geplant war, verdeutlicht, dass das Setzen von Station IDs nicht dem Zufall überlassen wird.

9.2 Wortinhalt

Überwiegende Einigkeit herrscht bei der Frage nach der zeitlichen Dauer der Moderation. Die quantitative Untersuchung liefert eindeutige Ergebnisse, die die meisten Programmdirektoren auch begründen können. Dennoch sind die Sender nicht immer in der Lage gewesen, die Vorgaben umzusetzen. BAYERN 3 hat den Versuch unternommen die Moderationsdauer zu erhöhen, allerdings gelang dies den „Frühaufdrehern" im Untersuchungszeitraum 2014 nicht. Das ist durchaus erwähnenswert, da die „Frühaufdreher" im Jahr 2008 die Moderatoren Roman Roell und Markus Othmer abgelöst haben und nach sechs Jahren ein eigespieltes Team gewesen sind, welches in der Lage hätte sein müssen, wie gefordert einen hohen Wortanteil zu produzieren – stattdessen reduzierte sich die Moderationsdauer. Bei BAYERN 1 sollten Interviews etwas fokussierter und gezielter auf Kernaussagen ausgerichtet sein, weshalb sich der Wortanteil in der Moderation entsprechend reduziert hat. Dies erklärt auch die gestiegene Musiklaufzeit bei BAYERN 1. Die Begründungen für die Reduzierungen des Moderationsanteils bei den Jugendsendern erscheinen ebenfalls plausibel. Bei bigFM hatte Moderatorin Susanka keinen ebenbürtigen Moderationspartner, bei DASDING griff die gestiegene Formatierung der Sendeuhren.

Vor allem am Beispiel BAYERN 3 lässt sich die gesunkene Laufzeit der journalistischen Inhalte sehr gut nachvollziehen. So sind beispielsweise Beiträge oder Korrespondentenberichte reduziert worden, da es Aufgabe der „Frühaufdreher" gewesen wäre, diese zu transportieren. Aus dem Vergleich der gesendeten Beitragsinhalte geht hervor, dass BAYERN 3 im Jahr 2014, wie von der Programmdirektion festgelegt, weniger journalistischen Content auf Sendung hatte, als noch 2008. Aus der geplanten Reduzierung des journalistischen Inhalts und der unbeabsichtigten Verringerung der Moderationsdauer, lässt sich folglich auch der gesteigerte Musikanteil erklären. BAYERN 3 hat im Jahr 2014 etwa zehn Prozent mehr Musik gespielt als im ersten Untersuchungszeitraum. Aus der überraschten Reaktion von Walter Schmich bei der Vorlage der Ergebnisse während des Interviews, lässt sich zudem schließen, dass es nicht geplant gewesen ist, den Musikanteil zu erhöhen. Es bestätigt sich die Äußerung, dass die „Frühaufdreher" die Vorgaben nicht umgesetzt haben.

Auch die Aussage des ehemaligen bigFM Programmdirektors Karsten Kröger wird bestätigt. Dieser gibt an, dass der Sender stets daran orientiert war, mehr Inhalt im Programm zu bieten als bei einem CHR-Sender üblich. Sowohl 2008 als auch 2014 bot bigFM die höhere journalistische

Laufzeit als DASDING. Im zweiten Untersuchungszeitraum hat der Privatsender zeitlich fast doppelt so viel journalistischen Inhalt auf Sendung gehabt, wie die öffentlich-rechtliche Konkurrenz. Aus der vorausgehenden Äußerung und anhand eines größeren Inhalts lässt sich eine etwas ältere Zielgruppe ableiten, die eher auf junge Erwachsene ausgerichtet gewesen ist als auf Jugendliche. Die Bemerkung Karsten Krögers, dass bigFM eher SWR3 als Konkurrenz gesehen hat als DASDING, erscheint plausibel und anhand der journalistischen Programmausrichtung als nachvollziehbar. Widersprüche ergeben sich hier für das Jahr 2008, allerdings stimmt die Äußerung für die musikalische Auswahl des Jahres 2014. Die Analyse der Erscheinungszeiträume der eingesetzten Titel (Anhang 1, Abbildungen 2 und 3, S. 7) ergibt, dass bei bigFM zwischen den beiden Untersuchungszeiträumen das Programm älter geworden ist. Glaubhaft erscheint zudem die Begründung für die Erhöhung des journalistischen Wortanteils bei Radio Arabella. Dies wird mit der jeweiligen Ausrichtung der wechselnden Redaktionsleiter begründet. Aus eigener Erfahrung des Verfassers kann bestätigt werden, dass der Programmverantwortliche im ersten Untersuchungszeitraum, auch bei anderen Sendern, weniger Wert auf Content als auf Musik legt. Auch hier liegt eine bewusste Steuerung vor, zufällige Ausprägungen können nicht nachgewiesen werden.

Nachvollziehbar sind auch die Erklärungen für die veränderten Laufzeiten der Nachrichten. BAYERN 3 hatte zu wenig Unterhaltungskompetenz und wollte diese mit den „Frühaufdrehern“ erhöhen. In der Folge sind „harte“ Informationen in die Nachrichten verschoben worden, weshalb sich die Laufzeiten erhöhten. Dies lässt sich, wie bereits beschrieben, auch aus der Reduzierung der journalistischen Inhalte und deren Laufzeiten ableiten. Auch bei den Oldie-Wellen gibt es nachvollziehbare Gründe für die Verkürzung der Nachrichten. Bei BAYERN 1 wurde eine regionale Topmeldung eingeführt, weshalb sich die Weltnachrichtendauer verringerte, die Gesamtlaufzeit zur vollen Stunde aber fast gleichgeblieben ist. Ein weiterer Grund war, dass der Serviceblock mit Wetterpresenter sehr ausgedehnt gewesen ist, und man, um konkurrenzfähig zu bleiben, bewusst die Nachrichten kürzte, um rascher Musik spielen zu können. Peter Bartsch von Radio Arabella begründet die Kürzung ähnlich, verweist aber zudem auf die langen Werbeblöcke, weshalb eine Reduzierung der Nachrichtenlänge nötig war. Auch bei bigFM ist die Begründung, dass sich durch die Übernahme der Nachrichten von externen Zulieferern die Laufzeit erhöhte, nachvollziehbar. Es kann also formuliert werden, dass auch die Änderungen der Laufzeiten der Nachrichten bewusst gesteuert werden und nicht dem Zufall überlassen sind.

9.3 Musikprogramm

Walter Schmich bezeichnete BAYERN 3 als „Hot AC“ (Bayern 3: 37). Diese Einordnung konnte anhand der Musikanalyse aus dem quantitativen Teil der Untersuchung nachgewiesen werden. BAYERN 3 ist in beiden Untersuchungszeiträumen der schnellere Sender, bezogen auf die Geschwindigkeit nach Beats per Minute, gegenüber antenne bayern gewesen. Hinzu kommt auch, dass BAYERN 3 mehr aktuelle Hits in der Playlist hatte als der Privatsender. Die Aussagen von Walter Schmich über die musikalische Programmierung des Senders stimmen mit den Ergebnissen der quantitativen Untersuchung überein. Auch der Beginn der Programmreform von BAYERN 1, hin zu einem jüngeren Programm, lässt sich aus dem Datensatz ableiten, da die Schlager aus der Rotation genommen wurden. Selbstverständlich waren die Formatänderungen von BAYERN 1 und BAYERN 3 geplant und es wird eine langfristige Strategie verfolgt.

Möglicherweise soll sich die Zielgruppe von antenne bayern altermäßig von zwei Seiten verkleinern: BAYERN 3 nimmt dem Privatsender die jüngeren Hörer, während BAYERN 1 parallel die älteren Hörer des Privatsenders abwirbt. antenne bayern wird von den Musiksendern des Bayerischen Rundfunks, umgangssprachlich ausgedrückt, „in die Zange genommen“. DASDING hat ebenfalls bewusst seine Musikfarbe zwischen 2008 und 2014 geändert, um neue Zielgruppen zu erschließen. Die Auswahl der Musikgenres ist also wegen der Konkurrenzsituation geplant und gesteuert. Dabei steht natürlich der Erfolg des eigenen Senders im Vordergrund – diese Ergebnisse waren zu erwarten und sind nicht überraschend.

Äußerst verwunderlich ist hingegen, dass die Sender zwar viel Wert auf die Auswahl der Titel legen und auch Research betreiben, welche Lieder besonders gut bei den Zielgruppen ankommen – die Songabfolge aber in allen Fällen Mängel aufweist. Wie bereits angeschnitten, hat das Ergebnis der Untersuchung des Musikflusses gezeigt, dass die Sender keinen Wert auf diesen legen. Auch wenn die Stationen mithilfe von Musikplanungssoftware ihre Rotationen berechnen lassen, so ist das Ergebnis in allen Fällen eher dürftig und würde ohne den Einsatz von Verpackungselementen kaum funktionieren. Für Hörer ist keinerlei harmonischer Fluss bei der Songauswahl erkennbar. Es wird bewusst ein großer, abwechslungsreicher Musikmix ausgestrahlt. Ob dieser allerdings erfolgreicher ist, sei dahingestellt, da keine öffentlich zugänglichen Untersuchungen verfügbar sind, die eine höhere Verweildauer durch einen stark abwechslungsreichen Musikmix belegen. Es kommt hinzu, dass die Geschwindigkeit der Musik von allen Musikredaktionen nur geschätzt wird, und die Geschwindigkeitser-

mittlung nur auf dem Gefühl der Musikredakteure basiert. Radiosender überlassen, was sich sowohl aus den Interviews, als auch aus den quantitativen Ergebnissen ableiten lässt, so gut wie nichts dem Zufall. Es ist also etwas paradox, dass das, was für einen Sender mit am wichtigsten ist, nämlich der Musikfluss, aus eher zufällig aneinander gereihten Liedern besteht. Gerade bei einem Nebenbeimedium sollte diese unterschwellig beim Hörer wahrgenommene Komponente nicht unterschätzt werden. Noch schwieriger wird es für die Sender, wenn der Hörer aktiv und bewusst Radio hört und ihm dann diese starken Schwankungen deutlich auffallen können.

9.4 Ergebnis qualitative Forschungsfrage

Aus dem Vergleich und der Diskussion der quantitativen und der qualitativen Ergebnisse ergibt sich für die die Forschungsfrage

> *„Waren die strukturellen, präsentatorischen, inhaltlichen und musikalischen Veränderungen der Sender bigFM, DASDING, antenne bayern, BAYERN 3, Radio Arabella und BAYERN 1 zwischen den Jahren 2008 und 2014 bewusste Entscheidungen der Programmverantwortlichen, oder handelte es sich um zufällige Entwicklungen? Und was waren gegebenenfalls die Beweggründe?“*

folgendes Ergebnis:

> **Die untersuchten Radiosender planten ihre Programme in den meisten Fällen sehr sorgfältig, und es gab im Normalfall keine zufälligen Entwicklungen.**

Allerdings kann es durchaus vorkommen, dass nicht alle Vorgaben der Programmdirektion umgesetzt werden. Dies liegt möglicherweise an mangelnder interner Kontrolle des eigenen Programms oder an der Personalbesetzung, die die Vorgabe nicht umsetzen kann. Das wird beim Versuch der Steuerung des Wortanteils erkennbar, so konnten beispielsweise die Co-Moderatoren von bigFM mit Susanka nicht mithalten, und die „Frühaufdreher“ haben den ihnen zur Verfügung gestellten Raum nicht voll ausgenutzt. Dies mag allgemein an der Ausbildung der Moderatoren liegen, die im Volontariat auf 1:30 Minuten Sprechzeit geeicht werden und bei kleinen Sendern meistens als Solomoderatoren zu hören sind. Kommen diese dann in einer Sendung mit Doppel- oder Dreifachmoderation zum Einsatz, ist es durchaus möglich, dass die Charaktere nicht ganz harmo-

nieren und die maximale Zeit eines Breaks von 1:30 so verinnerlicht ist, dass Formatmoderatoren diesem Korsett nur schwer entfliehen können. Hinzu kommt, dass Programmdirektoren im Aircheck Moderatoren oft eindringlich darauf hinweisen, sanfter und kürzer zu moderieren. Zudem sollen sie keine „Ecken und Kanten" haben.

Ein überraschendes Ergebnis ist, dass bei der Musikzusammenstellung bewusst auf viel mehr Abwechslung als auf Flow Wert gelegt wird. Es wäre sicherlich eine interessante Forschungsfrage, ob eine harmonische Musikabfolge Einfluss auf die Verweildauer hätte.

9.5 Limitationen

Nachdem die Untersuchungsergebnisse in den vorherigen Kapiteln dargelegt wurden, sind nun mögliche Schwachstellen der Methodik offenzulegen. Wie bei fast jeder größeren Untersuchung gibt es gewisse Herangehensweisen, die sich im Laufe der Arbeit als wenig zielführend, unnötig zeitintensiv oder nicht durchführbar erweisen.

9.5.1 Quantitative Methode

In Analogie zur Abfolge der vorangegangenen Kapitel wird zunächst die quantitative Methode hinsichtlich ihrer Limitationen analysiert. Die Methodenkritik soll auf mögliche Fehlerquellen der Arbeit hinweisen, die das Ergebnis verfälscht haben könnten.

Auswahl der Untersuchungszeiträume und künstliche Wochen. Ursprünglich war für die Untersuchung der Hörfunksender vorgesehen, dass neben der ersten Untersuchungswoche 2008 ein Jahr später, 2009, erneut im März eine zweite Datenerhebung stattfinden sollte. Aufgrund diverser Verzögerungen und um den Datensatz etwas aktueller zu gestalten sowie dem möglichen Kritikpunkt zu entgehen, die Abstände zwischen den Messzeitpunkten seien zu kurz gewählt, wurde der ursprünglich zweite Erfassungszeitraum, obwohl er vorliegt, aus der Untersuchung gestrichen und gegen eine Woche im März 2014 ersetzt. Die Begründung für eine Märzwoche liegt darin, dass die Erhebungen in den Zeitraum der Funkanalyse Bayern und der Mediaanalyse in Baden-Württemberg fallen sollten. Aus heutiger Sicht wäre es aus Aktualitätsgründen sinnvoller gewesen, zuerst das Jahr 2008 zu codieren, und nach Abschluss der ersten Codierarbeiten eine zweite Woche zu erheben, die nicht zwingend im März

hätte aufgezeichnet werden müssen. So hätte ein Vergleich zwischen den Jahren 2008 und 2018 stattfinden können, der die Ergebnisse vermutlich weiter ausgeprägt hätte, gerade vor dem Hintergrund, dass BAYERN 1 mit einer umfassenden Programmreform im Jahr 2014 begonnen hatte, deren Anfänge sich in der Playlist 2014 bereits abzeichneten, die aber erst zwei bis drei Jahre später abgeschlossen war. Allerdings ist mit der Aufzeichnung ein nicht unerheblicher technischer und damit finanzieller Aufwand verbunden. Waren 2008 und 2014 noch analoge Radioreceiver im Einsatz, so hätte man nach der Abschaltung der analogen Kabelfrequenzen durch Kabel Deutschland im Jahr 2018, in neue Receiver investieren müssen, um die Programme zu empfangen.

Des Weiteren wäre es aus heutiger Sicht auch sinnvoller, nicht die Woche vor der Kommunalwahl in Bayern zu erheben, da es zu inhaltlichen Verschiebungen im Bereich der Politik gekommen sein könnte. Allerdings besteht dieses Problem auch, wenn ein Politikskandal beispielsweise bei einer Aufzeichnung im Sommer aufgedeckt würde. Ähnlich verhielt es sich mit der Urteilsverkündung gegen Uli Hoeneß wegen Steuerhinterziehung. Diese lag in der Untersuchungswoche im März 2014. Hier wurde der Inhalt sicherlich in Richtung der journalistischen und moderativen Inhalte „Politik & Recht“ sowie „Verbrechen“ verzerrt. Ohne einen wissenschaftlichen Beleg zu haben, war es allerdings sehr erkenntnisreich, dass bei den Jugendsendern in Baden-Württemberg das Thema eher humoristisch begleitet wurde, wohingegen in Bayern bei den AC- und Oldie-Stationen der ehemalige FC Bayern Präsident nicht als Vorlage für Gags diente.

Eine Methode, um Tendenzen in der Berichterstattung oder bei Inhalten zu umgehen, ist der Einsatz von „künstlichen Wochen“. Allerdings wäre auch hier der Arbeits- oder besser zwischenmenschliche Aufwand entsprechend hoch gewesen. Hintergrund ist, dass bigFM nur über das Kabelnetz in Baden-Württemberg stabil empfangbar war. Infolge dessen musste eine Aufnahmeeinheit in Ulm bei einer Privatfamilie positioniert werden. Es wäre wohl kaum vertretbar gewesen, den Computer mit Receiver über fast zwei Monate tageweise in einem fremden Haushalt zu verwenden. Auch um die Gefahr zu umgehen, dass das Aufnahmeprogramm seitens der Personen vor Ort nicht korrekt gestartet wird oder auch Urlaubs- und Ferienzeiten in die acht Wochen mit eingerechnet werden müssen, hätte einen erhöhten Mehraufwand bedeutet.

Ein Kritikpunkt mag sicherlich auch das Alter der untersuchten Daten sein. Allerdings ist es berufsbegleitend nur sehr schwer möglich, aktuelle Daten zu liefern, da die Codierung und das Verfassen der Arbeit einen ge-

wissen Zeitraum in Anspruch nehmen. Ältere Daten bieten aber auch den Vorteil, dass in Experteninterviews mit ehrlicheren Antworten zu rechnen ist, als wenn mit einem aktuellen Datensatz operiert wird. Hintergrund mag sicherlich sein, dass Programmdirektoren vermutlich nicht über aktuelle Strategieausrichtungen sprechen möchten, aber über Ereignisse, die einen gewisse Zeit zurückliegen, wahrscheinlich offenere Aussagen treffen.

Limitierung des Codeplans und der Codierung. Der Hauptkritikpunkt bei der Umsetzung des Codeplans und der damit verbundenen Dateneingabe in SPSS ist nicht, dass die Datenmatrix zu wenige Optionen hatte, sondern eher zu umfangreich gestaltet war. Ein Teil der erhobenen Daten konnten bei der Auswertung nur am Rande berücksichtigt werden. So wurde zum Beispiel ermittelt, ob eine Moderation „mit Witz" stattgefunden hat oder nicht. Dieses Kriterium erwies sich im Nachgang als zu subjektiv und konnte nicht in die quantitative Hauptuntersuchung einfließen. Auch wurde für jede Moderation erhoben, egal ob es sich um einen Wortbeitrag mit einem oder zwei Co-Moderatoren handelte. Es wäre ausreichend gewesen, dies in der allgemeinen Beschreibung der Frühsendung zu vermerken, als diesen Punkt für jeden „Modbreak" erneut zu ermitteln. Neben den Erhebungen der Titel und Interpreten wurden zudem noch das Geschlecht und die Sprache bestimmt. Allerdings konnten diese Erkenntnisse, aufgrund des Umfangs der vorliegenden Arbeit, nicht mit einfließen. Mit etwas mehr Überlegung im Vorfeld, hinsichtlich der Auswertung der Daten, wäre an dieser Stelle eine effektivere, zeitsparendere Codierung möglich gewesen.

Anhand des umfangreichen Datenmaterials wurde, obwohl für alle Sender eine Vollerhebung zur Verfügung steht, lediglich die Frühsendung inhaltlich zwischen 6 Uhr und 9 Uhr detailliert ausgewertet. Dafür benötigte eine Person etwa neun Monate in Teilzeit. Für eine genauere Datenanalyse von 6 Uhr bis 24 Uhr wären beim gleichen Zeitaufwand sechs Personen benötigt worden. Um eine empirische Vollerhebung (von 6 Uhr bis 24 Uhr) in drei Monaten abzuschließen, hätte man ein Team von etwa zwölf bis 15 Personen in Festanstellung benötigt – ein Aufwand, der finanziell kaum zu leisten ist. Machbar war hingegen die Auswertung der Playlists von 6 Uhr bis 24 Uhr.

Die wohl größte Limitierung des Codeplans ergibt sich für den Bereich Moderation. Diese lässt sich inhaltlich in ein Hauptthema sowie zwei Nebenthemen aufteilen. Eine solche Einteilung richtet sich in der Regel nach der Laufzeit. Es ist allerdings so gut wie nicht möglich, diese Laufzeiten zusätzlich zu erfassen. Im Regelfall stellt dies auch kein Problem dar, da es aus dem Hörbild deutlich hervorgeht, was das Hauptthema ist. Es kann

allerdings bei kleinen Sendeplätzen, die einen Überblick über das Tagesgeschehen geben sollen, vorkommen, dass der Moderator beispielsweise über das Wetter in gleichem Maße spricht, wie zum Beispiel über ein verschwundenes Flugzeug und eine Senderaktion mit nachgeschaltetem Promo. Dann ist es nicht mehr möglich eindeutig zu unterschieden, welches Thema Haupt- beziehungsweise Nebeninhalt war. Eine Lösung für dieses Problem wäre nur über den Faktor Zeiterhebung möglich, was aber unter Umständen auch kein Ergebnis liefern würde, denn: Welches Thema ist bei einer Überblicksrubrik des Tagesgeschehens wie „Drei Gesprächsthemen in drei Sätzen" von größerer Bedeutung? Bei den journalistischen Themen beziehungsweise Inhalten stellte sich dieses Problem nicht, da hier üblicherweise nur ein Thema behandelt wird. Hinzu kommen die Themenbereiche „Schule, Ferien und Urlaub". Diese werden im Codeplan (Anhang 2, Tabelle 2, S. 162 bis S. 165) inhaltlich als „Sonstiges" (V201/13, V402/13, V403/13 und V404/13) codiert. Die Auswertung hat jedoch gezeigt, dass diese Themenfelder, vor allem bei den CHR- und AC-Formaten überraschend oft thematisiert werden. Folglich wäre eine eigene Kategorie für die Erfassung sinnvoll gewesen.

Ebenfalls liegt es im Bereich des Möglichen, dass es Schwachstellen bei der Auswertung der Chi-Quadrat-Tests hinsichtlich der nominalen Variablenausprägungen gegeben haben könnte. Manche Variablen (zum Beispiel journalistische Genres) mussten aufgrund mangelnder Fallzahlen unberücksichtigt bleiben, wodurch kleinere Verzerrungen bei der Auswertung entstehen können. Hinzu kommt, dass aufgrund der unterschiedlichen Ausprägungen zwischen 2008 und 2014, auch zwischen den Sendergruppen, ein Vergleich derselben Variablenausprägungen nicht möglich war. Allerdings bedeutet das Nichtauswerten von Variablen, die keine Werte haben, dass die Sender eher unähnlich sind, da nicht beide Stationen keine Fallzahlen aufweisen, sondern im Regelfall nur eine.

Eine weitere Limitierung, die zunächst nicht erwähnenswert erscheint, ist die Möglichkeit des Fehlcodierens durch Tippfehler. Trotz intensiver Kontrolle ist die Eventualität der Ergebnisverschiebung durch versehentliche Eingabefehler vorhanden. Dies wird bereits bei der Differenz zwischen der theoretischen Sekundenzahl als auch der tatsächlich codierten Zeit deutlich. Rein rechnerisch beträgt die maximal codierbare Zeit der Frühsendungen 907.200 Sekunden. Tatsächlich wurden aber bei der Gegenrechnung mit SPSS 907.223 Sekunden codiert. Es liegt im Bereich des Möglichen, dass sich innerhalb der Datenerhebung weitere Codierfehler befinden, die einen Einfluss auf das Ergebnis haben könnten, und die sich vermutlich auch gegenseitig egalisieren. Die Eingabefehler sollten also,

aufgrund der großen Datengrundlage von zwölf Wochen zu analysierender Morning Shows, nur wenig ins Gewicht fallen – auch bezogen auf die Gesamtdifferenz von lediglich 23 Sekunden.

Playlists. Die Erfassung und Analyse der Playlists stellte einen nicht unerheblichen mehrwöchentlichen zeitlichen Aufwand dar. Um ein möglichst umfassendes Bild der Musikrotationen der Sender zu erhalten, wurden diese sowohl 2008 als auch 2014, von Montag bis Sonntag eine Woche lang zwischen 6 Uhr und 24 Uhr erhoben.

Trotz aller technischen Hilfsmittel, wie der App Shazam oder dem Abrufen der Playlist über die Senderhomepages, konnten nicht alle gesendeten Titel erfasst werden. Als besonders problematisch erwiesen sich abendliche Spezialsendungen für elektronische als auch Volksmusik. Dieses Problem ließe sich nur dadurch lösen, falls man die Sender vor der Datenerhebung informieren und um die Bereitstellung der Playlists bitten würde. Allerdings wäre dadurch auch eine Ergebnisverschiebung möglich, da die Sender unter Umständen während des Untersuchungszeitraums ihre Rotationen im Sinne der Untersuchung manipulieren könnten. Auch ist es möglich, dass Sender die Playlists nicht zur Verfügung stellen, um Datendiebstahl und missbräuchliche Verwendung zu verhindern. Es ist davon auszugehen, dass, vor allem bei den Jugendsendern, pro Woche etwa 20 bis 30 Titel nicht zugeordnet werden konnten – vor allem im Bereich der DJ-Mix-Sendungen mit Basis elektronischer Musik, da hier oft inoffizielle Remixe oder Mash-Ups zum Einsatz kamen. Bei rund 1600 bis 1800 Titeln pro Woche ein zu vernachlässigender Wert. Bei den AC- und Oldie-Formaten stellte sich dieses Problem nicht.

Neben der Datenerhebung der Playlists ergaben sich aus den verschiedenen Schreibweisen der Interpreten Schwierigkeiten bei der Erfassung. Ein gutes Beispiel hierfür ist der Song „All For Love“ von „Rod Stewart, Bryan Adams & Sting“ sein. Die Schreibweisen unterscheiden sich zwischen „All 4 Love“ und „Bryan Adams Feat. Rod Stewart with Sting“ erheblich. Je nach Label, Interpret und Album gibt es an dieser Stelle sowohl unterschiedliche Schreibweisen als auch die damit verbundene alphabetische Reihenfolge bei der Interpreten- und Titelermittlung. Daraus resultiert, dass unter Umständen einige wenige Titel der Playlist doppelt erfasst wurden. Weitere Erhebungsunterschiede ergeben sich aus dem Worten und Symbolen wie „and“ und „&“, als auch zwischen den umgangssprachlichen, apostrophischen Schreibweisen geschriebener Wörter wie beispielsweise „playin´“ im Gegensatz zu „playing“ oder „singin´“ verglichen mit „singing“. Gleiches gilt für „gefeaturte“ Künstler. Gerade im R'n'B und Hip-Hop Bereich gibt es im Regelfall einen Hauptact, der von weiteren

Künstlern unterstützt wird. Häufig ist es mehr wie ein zusätzlicher Musiker, der im Refrain oder bei der Produktion mitgewirkt hat. Es ist nicht immer möglich, alle Mitwirkenden zu erfassen. Folglich sind vor allem beim senderübergreifenden Playlistvergleich unterschiedliche Schreibweisen und neben dem Hauptkünstler andere „gefeaturte" Künstler feststellbar, obwohl es sich um denselben Song handelt. Eine Korrektur und intensive Prüfung der Schreibweise hätte, außer einem einheitlicheren Schriftbild, keinen zusätzlichen wissenschaftlichen Informationsgewinn bedeutet.

Intracoder-Test. Um die Reliabilität des Testverfahrens zu bestätigen, wird ein Intracoder-Test vorgenommen (Früh, 2011, S. 188f). Der größte Teil des ersten Codiervorgangs fand im Herbst und Winter des Jahres 2018 statt, der Intracoder-Test folgte in zeitlichem Abstand von etwa eineinhalb Jahren im August 2020. Zunächst wurde die Basiscodierung für Nachrichten, Moderation, Musik, redaktionelle Inhalte, Werbung und Verpackungselemente geprüft. Danach folgten die Einzelvariablen der Vertiefungsanalyse. Als statistische Tests kamen Cohen's Kappa für kategoriale Variablen und die Pearson-Korrelation für metrische Daten zum Einsatz (Brosius, 1998, S. 422). Diese Vorgehensweise ermöglichte, dass unterschiedliche Variablentypen mit in die Gesamtauswertung einfließen konnten – sie wurden standardisiert. Um ein belastbares und wiederholbares Ergebnis des eigenen Codiervorgangs außerhalb der Basiscodierung auszuweisen, musste der aus allen Berechnungen gebildete Durchschnittswert, der im Normalfall zwischen null und eins liegt (Brosius, 1998, S. 422), ein Resultat über .80 erreichen, da kleinere Werte als weniger reliabel bis kritisch einzustufen sind (Medistat, Cohen's Kappa Koeffizient, Ohne Jahr).

Vom Intracoder-Test ausgeschlossen wurden beispielsweise Filenamen, Sendernamen, Wochentage oder Sendungsnamen, da es sich pro Sendestunde und Item um einen „Copy-Paste-Vorgang" handelte, der keinen Einfluss auf das Codierverhalten hat. Ebenfalls nicht geprüft wurden Songtitel und Interpreten. Alle anderen Variablen flossen in den Test mit ein.

Für den Intracoder-Test wurde die Sendestunde des Senders BAYERN 1 von Mittwoch den 12. März 2014 zwischen 7 Uhr und 8 Uhr herangezogen. Früh fordert als Mindestgröße für Textauswertungen circa 30 bis 50 Nennungen pro Variable, optimal wären seiner Auffassung nach 200–300 (Früh, 2011, S. 189). Diese Zahlen können auf eine Hörfunkforschung nur schwer übertragen werden. Würde man die geforderten 200 Variablen alleine auf Nachrichten beziehen, so wären 100 Stunden Hörfunk (bei durchschnittlich zwei Nachrichtenausgaben zur vollen und zur halben

Stunde) neu zu codieren, was etwa der Hälfte der untersuchten Daten entspricht. Da die Codierung einer Hörfunkstunde aus über 300 Einzeloptionen besteht, wird diese Vergleichsgrundlage als ausreichend erachtet, da sie der numerischen Vorgabe von Früh folgt.

Auswertung kategoriale Retest-Reliabilität. Die Basiscodierung der Variablen V100 bis V600 ergab, dass der Kappa-Wert nicht berechnet werden konnte (Anhang 1, Tabelle 216, S. 128), da die Werte konstant waren. Dies ergibt sich, wenn die beiden zu vergleichenden Codierer ohne Abweichung exakt das Selbe codieren. IBM erklärt dazu: „SPSS will not calculate kappa statistic (…) If one rater scores every subject the same, the variable representing that rater's scorings will be constant and SPSS will produce the above message. (…) If the two raters assign every subject the same rating, then the denominator becomes 0 as well, in which case kappa is indeterminate“ (IBM Support, Ohne Jahr).

Dies bedeutet, dass eine Übereinstimmung von 100 Prozent vorlag. An dieser Stelle hätte es auch keine Abweichung geben dürfen, da sonst die Reliabilität des Codierers stark infrage stünde. Träten hier Fehlcodierungen auf, die nicht auf Tippfehler zurückzuführen wären, würde dies bedeuten, dass der Zweitcodierer beispielsweise nicht in der Lage gewesen wäre, zwischen Nachrichten, Moderation oder Werbung zu unterscheiden. Für einen Intracoder-Retest würde dies bedeuten, dass die gesamte Untersuchung fraglich wäre.

Die gleichen Ergebnisse ergaben sich für die Variablen V209, V406 und V501. Die Begründung für die hohen Übereinstimmungswerte lag in der Einfachheit der Auswahlmöglichkeiten, da bei V209 und V406 lediglich bestimmt werden musste, ob die Moderation, oder der redaktionelle Beitrag von einem Musikbett unterlegt war, oder nicht. Auch Variable 501 war an dieser Stelle sehr leicht nachzuvollziehen, da es sich bei der Codierung um klassische Werbeblöcke mit verschiedenen Werbespots handelte.

Keine Ergebnisse lieferten die Variablen V205, V206, V207, V208, V210, V212, V313, V402, V403, V407, V408 und V409. Der Grund hierfür ist, dass die Variablen in der Sendestunde nicht vorkamen und folglich auch nicht codiert wurden. Das Ergebnis spricht also für den Retest und kann mit einer Übereinstimmung von ebenfalls 100 Prozent (beziehungsweise 1) gewertet werden – es wurde nichts codiert, was nicht gesendet worden ist.

Kappa-Ergebnisse konnten für die Variablen V101, V201, V202, V203, V204, V304, V307, V308, V309, V310, V311, V312, V401 und V402 berechnet werden. Von den 14 erneut getesteten Variablen lagen neun bei einem Wert um eins. Größere Abweichungen ergaben sich bei V312 (Charter-

folg), V307 (Geschlecht), V204 (Musikinformation) und den Inhalten der Moderation V201 und V202. Der niedrige Wert von $k_{V312} = .305$ ergab sich aus einer Fehlcodierung von insgesamt 13 Fällen. Bei den Moderationsinhalten gab es ebenfalls Abweichungen. Es trat bei der erneuten Codierung der Sendestunde der Fall ein, dass Teile der Moderation nicht immer ganz eindeutig Haupt- und Nebeninhalt zugeordnet werden konnten.

Aus den berechenbaren Kappa-Ergebnissen (Anhang 1, Tabelle 216, S. 128) ergab sich eine durchschnittliche Übereinstimmung von gerundet .89. Bezöge man die konstanten und nicht berechenbaren (keine Fallzahlen) Variablen mit jeweils eins in die Berechnung ein, so ergäbe sich ein Wert von gerundet .96. In beiden Fällen liegt eine sehr hohe Übereinstimmung vor, und das Ergebnis ist als reliabel einzustufen.

Auswertung metrische Retest-Reliabilität. Für die Auswertung der metrischen Variablen wird der Pearson Korrelationskoeffizient verwendet. Brosius schreibt: „Pearsons Korrelationskoeffizient liegt stets zwischen +1 und -1. Das Vorzeichen gibt die Richtung, der Betrag die Stärke des Zusammenhangs“ (1998, S. 510). Auch hier sollte am Ende ein Resultat von über .80 erreicht werden (Medistat, Korrelationskoeffizient nach Pearson, Ohne Jahr). Da beide Testergebnisse den gleichen Ergebnisraum einnehmen, sind sie standardisiert und können miteinander verglichen werden, beziehungsweise lässt sich ein Durchschnittswert berechnen.

Ein Ergebnis lieferten die Retests der Variablen V02begin, V02ende, V04laenge, V303, V305, V306 sowie V601 (Anhang 1, Tabelle 217, S. 129). Dabei lagen vier Variablen bei Werten um eins, nämlich V02begin, V02ende, V303 und V601. Zu einer leichten Abweichung kam es bei V04laenge. Dies ergab sich deshalb, da V02begin und V02ende nur nahe eins liegen und nicht vollkommen konstant waren. Das dennoch sehr gute Retest-Resultat von $r_{v04laenge} = .997$ lässt sich auf die Zuhilfenahme von Adobe Premiere (alternativ wäre auch jedes andere Audioschnittprogramm, bei dem eine Waveform dargestellt wird, denkbar) zurückführen. Neben dem akustischen Signal wird eine Wellenform visualisiert, durch die sich ein zusätzlicher optischer Anhaltspunkt ergibt, wann zum Beispiel eine Moderation endet. Für ein geschultes Auge ist der Unterschied zwischen einer Moderation und einem Lied deutlich erkennbar. Nichtsdestotrotz gab es an dieser Stelle leichte Verschiebungen bei der Erfassung, da Verpackungselemente mit den darin enthaltenen Effekten (Hall, Echo, Swooshes, Swipes und so weiter) das genaue Ende eines Jingles und des nachfolgenden Songs verschleierten. Dabei ging es aber meistens nur um eine Sekunde.

Ebenfalls hohe Übereinstimmungen gab es bei der Ermittlung der Geschwindigkeiten (BPM) am Anfang und Ende der Songs. Die ermittelten

Tempi von V305 und V306 stimmten beim Intracoder-Test zu .999 überein. Die Abweichung zum Idealwert war minimal und konnte sich daraus ergeben, dass der verwendete BPM-Counter das Ergebnis rundete. Um das Resultat zu verfälschen, reichten unter Umständen ein paar Zehntelsekunden gegen Ende der Geschwindigkeitsermittlung aus, um einen nach unten oder oben gerundeten Wert zu erhalten. Dennoch ist ein Wert von .999 ein äußerst belastbares Ergebnis.

Für die Variablen V102, V211 und V602 konnte keine Pearson-Korrelation berechnet werden, da es sich um Konstanten handelte. Auch für Variable V405 war es nicht möglich eine Berechnung durchzuführen, weil es keine Fallzahlen gab.

Aus den berechneten Pearson-Korrelationen ergab sich ein Durchschnittswert von gerundet .99. Dieser hohe Wert lässt sich daraus erklären, dass als Grundlage relativ große Zahlengrundlagen verwendet wurden, zum Beispiel Sekunden oder Beats per Minute, bei denen kleine Abweichungen von eins oder zwei rechnerisch so gut wie nicht ins Gewicht fielen. Würde man die Konstanten und Variablen ohne Fallzahlen noch mit ins Ergebnis mit einbeziehen, würde der Wert weiter gegen eins gehen, was aber aufgrund des hohen Ausgangsergebnisses nur noch eine rechnerische Verbesserung des Resultats mit sich brächte.

Gesamtergebnis Retest-Reliabilität. Für die Berechnung des Gesamtergebnisses werden nun lediglich das berechnete Kappa-Ergebnis als auch das der Pearson-Korrelation verwendet (Anhang 1, Tabelle 218, S. 129). Aufgrund der vorherigen nötigen Trennung zwischen kategorialen und metrischen Datensätzen, ist nun ein Vergleich möglich, da die Ergebnisse standardisiert worden sind. Die mittlere Übereinstimmung liegt bei gerundet .94. Das Codierverhalten ist als stabil einzuordnen und erreicht ein sehr gutes Resultat. Die Codierung war trotz des zeitlichen Abstands von etwa eineinhalb Jahren sehr gut zu reproduzieren, wodurch die quantitative Untersuchung, als auch der Codierer, als reliabel einzustufen sind.

9.5.2 Qualitative Methode

Ähnlich wie quantitative Forschungsarbeiten unterliegen auch qualitative Untersuchungen gewissen Gütekriterien. Dabei wird nach Kuckartz zwischen interner und externer Studiengüte unterschieden. Für ihn sind interne Gütefaktoren zum Beispiel Verlässlichkeit, Glaubwürdigkeit, Auditierbarkeit, Regelgeleitetheit sowie intersubjektive Nachvollziehbarkeit. Letztere wird zum einen dadurch gewährleistet, dass alle wichtigen Er-

kenntnisse sowie die Projektdateien im digitalen Anhang offengelegt werden. Hinzu kommen Faktoren, wie die erhobenen Daten fixiert worden sind, also beispielsweise durch Audio- oder Videoaufnahmen, der Zeitdauer zwischen Interview und Transkription, den Transkriptionsregeln oder der Anonymisierung der Daten. Als externe Studiengüte sieht er Übertragbarkeit und Verallgemeinbarkeit (Kuckartz, 2018, S. 203).

Güte. Um also ein möglichst exaktes Ergebnis nach Kuckartz Vorgaben zu erhalten, wurden die Interviews als Audiofile komplett mitgeschnitten und abgespeichert. Der Vorteil liegt unter anderem in der Genauigkeit der Transkription, die jene eines Protokolls weit übersteigt. Außerdem sind nur so wörtliche Zitate bei der Auswertung möglich (Kuckartz, 2018, S. 165). Da es sich bei den Interviewpartnern um routinierte Journalisten handelt, kann ein „unangenehmes Gefühl bei den befragten Personen, dass alles aufgezeichnet wird und deshalb Gefahr besteht der Verunsicherung und der Verzerrung des Interviews" (Kuckartz, 2018, S. 165) ausgeschlossen werden. Abweichungen können eher dadurch entstehen, dass die Gesprächspartner wegen ihres Berufsbildes ganz gezielt Gegenstrategien verwenden, um kritische Fragen in positive Aussagen umzuformulieren. Noch am selben und am darauffolgenden Tag sind die Interviews vollständig transkribiert worden, da dies die genaueste Datengrundlage schafft. So konnte gewährleistet werden, dass auch einzelne Worte, die möglicherweise bei Online-Videokonferenzen aufgrund von Signalausfällen vorkamen, wenn auch äußerst selten, am Ende richtig vermerkt und ergänzt werden konnten. Eine Anonymisierung der Daten war nicht notwendig, da Interviewführung mit Namensnennung zum Berufsbild der Gesprächspartner gehören und folglich nichts Ungewöhnliches sind. Um für weitere Transparenz zu sorgen, befinden sich sowohl der Leitfaden als auch die komplett transkribierten Interviews im Anhang, beziehungsweise digitalen Anhang, an diese Forschungsarbeit. Hinzu kommen die unbearbeiteten Audiomitschnitte und die letzte Speicherung des MAXQDA-Projekts mit den Codiereinheiten (Kuckartz, 2018, S. 222).

Weitere relevante Punkte für die Studiengüte einer qualitativen Inhaltsanalyse sind unter anderem, ob die inhaltsanalytische Methode der Fragestellung gerecht wird, ob die Analyse computergestützt durchgeführt wird und konsistente Kategoriensysteme eingesetzt werden oder ob Originalzitate zum Einsatz kommen. Besonders hervorzuheben ist die Begründbarkeit der Schlussfolgerungen aus dem Datensatz, und was in welcher Form dokumentiert und archiviert wird. Eines der wichtigsten Kriterien ist die Intercoder-Reliabilität. Kuckartz spricht bei qualitativen Untersuchun-

gen allerdings nicht von Reliabilität, sondern von Intercoder-Übereinstimmung (Kuckartz, 2018, S. 204ff).

Auch wenn in der vorliegenden Untersuchung nur ein Codierer zum Einsatz gekommen ist, musste sich dieser dennoch selbst testen, um festzustellen, ob sich sein Codierverhalten nicht im Laufe der Zeit geändert hat. Man spricht in diesem Fall von Intracoder-Übereinstimmung. Dabei wird ähnlich vorgegangen wie bei der quantitativen Inhaltsanalyse, indem es zu einem zweiten Codierungsdurchlauf kommt, bei dem geprüft wird, ob dieser zu denselben Ergebnissen führt. Mayring bezeichnet diese Vorgehensweise als Re-Test (2015, S. 123). Bei diesem zweiten Durchlauf wird allerdings nur ein Drittel der Interviews einbezogen. Kuckartz ist der Meinung, dass die Codiereinheiten dazu vorher festgelegt werden müssen, da es sonst für den zweiten Codierer fast nicht möglich ist, in einem Text dieselben Codiereinheiten zu treffen:

> „Wie wahrscheinlich ist es, dass bspw. bei einem 3000 Wörtern umfassenden Text das gleiche Wort oder der gleiche Satz zufällig von zwei unabhängig Codierenden mit dem gleichen Code codiert werden. Die Wahrscheinlichkeit dürfte mit wachsendem Textumfang asymptotisch gegen null konvergieren. Daraus folgt, dass die Berechnung von Kappa (oder einem anderen zufallsbereinigten Koeffizienten) eigentlich nur Sinn macht, wenn man Codiereinheiten vorab festlegt“ (2018, S. 216).

Auch in dieser Arbeit wurden folglich die Codiereinheiten aus dem ersten Codiervorgang, die sich an den Fragen des Leitfadens orientierten, erneut beim Intracoder-Test verwendet. Für die Durchführung und Berechnung ist die in MAXQDA integrierte Funktion der „Intercoder-Übereinstimmung“ gewählt worden. Diese liefert für Kappa (k = .855) beziehungsweise 85,94 Prozent (Anhang 1, Tabellen 219 und 220, S. 130). Aufgrund des Ergebnisses von etwa 86 Prozent und dem Kappa Wert von gerundet .86 lässt sich schließen, dass die Codierung und auch das Ergebnis, als sehr reliabel einzustufen sind. Das relativ hohe Übereinstimmungsresultat ist dadurch zu erklären, dass die Codiereinheiten sehr eng mit den Fragen des Leitfadens verbunden waren und die Antworten der Interviewpartner sehr gezielt gegeben wurden.

Methodenkritik. Dennoch gibt es zwei Kritikpunkte an der qualitativen Studie, die an dieser Stelle genannt werden sollen. Zwischen der ersten Codierung und dem Intracoder-Übereinstimmungstest lagen lediglich etwa zwei Monate. Insofern kann davon ausgegangen werden, dass die Zuordnungen der Codiereinheiten beim Test noch etwas im Gedächtnis des Codierers waren. Allerdings kann dieses Argument teilweise widerlegt

werden, da die Zuordnungen an sich sehr logisch und nachvollziehbar sind, weil sich die Codiereinheiten größtenteils nach dem Leitfaden richten. Außerdem wäre es nicht zuträglich, im Sinne der Veröffentlichung der Studie, mehrere Monate mit dem Selbsttest zu warten, nur um am Ende ein Resultat zu bekommen, das vermutlich auf ähnlich hohem Niveau liegen würde.

Ein weiteres Manko könnte sein, dass der zeitliche Abstand zwischen den Hörfunkmitschnitten der Jahre 2008 und 2014, sowie den Experteninterviews relativ groß war. Es besteht also durchaus die Möglichkeit, was auch bei einzelnen Antworten der Interviewpartner deutlich wurde, dass nicht mehr genau rekapituliert werden konnte, welche Intension hinter einer Entscheidung oder einem Ergebnis steckte. Allerdings hat der zeitliche Abstand auch einen Vorteil: Höchstwahrscheinlich sind die Antworten aus den Experteninterviews ehrlicher, da keine aktuellen Betriebsgeheimnisse und taktische Senderausrichtungen angesprochen werden.

Außerdem kann kritisiert werden, dass es sich bei der vorliegenden Untersuchung nicht um eine vollständige Triangulation handelt, sondern um ein Mixed-Methods-Design bei dem der quantitative Teil den wesentlich größeren Umfang hat, als der qualitative. Um die Arbeit besser zu verzahnen und nicht zwei Teilstudien zu erhalten, hätte es vermutlich eine übergeordnete, erkenntnisleitende Frage benötigt sowie der Erkenntnis zu Forschungsbeginn, dass es einer qualitativen Studie bedarf.