

---

## Marcus Erbe, Aycha Riffi, Wolfgang Zielinski

### Einleitung

Ist die menschliche Stimme einzigartig und unverwechselbar? Der Schauspieler und Synchronsprecher Oliver Rohrbeck verkörpert bereits seit 1979 die Figur des Justus Jonas in der Hörspielserie *Die drei ???*. Der damals Dreizehnjährige steht noch heute regelmäßig mit Andreas Fröhlich („Bob Andrews“) und Jens Wawrczeck („Peter Shaw“) vor dem Studiomikrofon, um neue Abenteuer einzusprechen. Seit 2004 feiern die drei auch mit Live-Events große Erfolge. Die Ton-Bild-Schere – die von ihnen verkörperten Figuren sind in all den Jahren nicht mitgealtert – macht dabei den besonderen Charme der Veranstaltungen aus. Interessanter aber ist, dass die Stimmen in den Hörspielen noch heute funktionieren – für Neueinsteiger\*innen ebenso wie für nostalgietrunkene Hörspielfans, die sie seit Jahrzehnten begleiten. Indes scheint bei einer nicht minder populären Serie das Phänomen alternder Stimmen der narrativen Kontinuität im Weg gestanden zu haben. Als es für eine Folge von *The Mandalorian* (USA 2019ff.) den jungen Luke Skywalker wiederzubeleben galt, musste dessen Originaldarsteller Mark Hamill nicht nur äußerlich, sondern auch vokal verjüngt werden. Um den fraglichen *Star Wars*-Charakter im benötigten Alter möglichst glaubwürdig auftreten lassen zu können, bedienten sich die Produzent\*innen einer Kombination aus ‚falschem‘ Körper, facialer Deepfake-Software und dem Sprachsyntheseprogramm Respeecher. Digitale Abbilder eines rund 40 Jahre jüngeren Hamill wurden aus Archivaufnahmen seines Gesichts wie seiner Stimme gewonnen und mit der Mimik und Gestik des Schauspielers Max Lloyd-Jones technisch verflochten. Wenngleich derartige Kombinationen filmhistorisch betrachtet kein grundsätzliches Novum darstellen (z.B. resultieren auch der Körper, das Gesicht und die Stimme Darth Vaders in der zeitlich ersten *Star Wars*-Trilogie 1977-83 aus der Vermischung mehrerer individueller Performanzen), so verweist diese Anekdote auf das Moment der zunehmenden Perfektionierung artifizierender Identitätskonstruktionen mittels digitaler Werkzeuge.

Seit Beginn der Tonaufzeichnung geht es nicht nur darum, individuelle Stimmen zu konservieren und zu reproduzieren, sondern auch zu modifizieren, unseren Erwartungen und Bedürfnissen anzupassen und zu modellieren. Die Möglichkeit, durch Veränderung der persönlichen Stimme andere Identitäten annehmen zu können, passt in die Welt der anonymisierten Online-Kommunikation. DeepDubs könnten zukünftig Synchronisationsarbeit automatisieren, Sprachassistenzen vielleicht in unserem

bevorzugten Dialekt oder mit der Stimme unserer Lieblingsprotagonist\*innen mit uns kommunizieren. Wir stehen noch ganz am Anfang einer durch künstliche Intelligenz und Algorithmen sich grundlegend verändernden stimmlichen Kommunikation.

Die vorliegende Publikation ist aus dem Forschungsprojekt *Kulturelle Implikationen medial konstruierter Stimmen* hervorgegangen, das wir als Teil des Grimme-Forschungskollegs an der Universität zu Köln 2020 durchführten. Ziel dieses Projekts war die Untersuchung medial zirkulierender Stimmwürfe unter der leitenden Fragestellung, welche Sozialvorstellungen – z.B. von Angemessenheit, Autorität und Handlungsmacht – medial konstruierten Vokalitäten zugrunde liegen bzw. sich in ihnen ausdrücken. Seit der Einführung stimmbasierter Navigationsgeräte um das Jahr 2000 ist es zunehmend selbstverständlich geworden, dass digitale Apparate sich anhand menschlich klingender Stimmen mitteilen. Zwar datieren die Anfänge der elektronischen Stimm synthese in die späten 1930er-Jahre. Dennoch war es vor der allgemeinen Verfügbarkeit ‚sprechender‘ Navis, Computer und Smartphones hauptsächlich das Privileg phantastischer Filme, Hörspiele und Fernsehserien, den Mediennutzer\*innen eine Vorstellung von der Beschaffenheit künstlicher Stimmen zu vermitteln. Doch so wenig medial konstruierte Stimmen in fiktionalen Kontexten neutral sein können, so wenig neutral präsentieren sich die Stimmen rezenter Applikationen und Betriebssysteme. Dies zeigt sich etwa in der aktuell geführten Debatte um den inhärenten Sexismus bei Sprachassistenzsystemen wie Alexa, Siri und Cortana. Indes mangelt es an Untersuchungen, die zugleich theoretische, medienpraktische und kulturübergreifende Gesichtspunkte vokaler Designs berücksichtigen, und zwar nicht nur im Hinblick auf IT-Erzeugnisse, sondern die Medienproduktion ganz allgemein.

Bereits in der Planungsphase war uns sehr daran gelegen, Akteur\*innen aus Wissenschaft, Kunst, medienpädagogischer Praxis und Produktentwicklung in einen Dialog über die Spezifika heutiger wie früherer Stimmtechnologien zu bringen. Als projektinterner Höhepunkt erwies sich somit ein am 4. August 2020 abgehaltener digitaler Workshop, an dem sich Vertreter\*innen aus Computerlinguistik, Musikwissenschaft, Theaterwissenschaft, Filmwissenschaft und Voice-Design gemeinsam mit Rundfunkredakteur\*innen, Medienautor\*innen, Schauspieler\*innen und Kommunikationsberater\*innen beteiligten. Diskutiert wurde über die Stimme als Kulturphänomen, ihre Gestaltung im Kontext von Bewegtbildmedien, die Signifikanz ‚körperloser‘ Stimmen im Radio, die Stimme in der Mensch-Computer-Interaktion sowie über technische oder technisierte Stimmen in Musik, Theater und Performancekunst. Einigen unserer damaligen Gäste war es erfreulicherweise möglich, ihre Überlegungen aus dem Workshop schriftlich weiter auszuarbeiten und in diesem Buch niederzulegen. Andere gewährten uns in eigens geführten und hier abgedruckten

Interviews wichtige Einblicke insbesondere in die Arbeitsabläufe und Herausforderungen stimmbasierter Tätigkeitsfelder. Des Weiteren konnten wir – nicht zuletzt im Sinne einer internationalen Perspektivierung des behandelten Gegenstandes – Autor\*innen aus Brasilien, den Niederlanden, Südafrika und den USA hinzugewinnen. Da es mit den Voice Studies inzwischen einen globalen Forschungszweig gibt, der stimmlichen Phänomenen in den Künsten, Medien sowie der Alltagskommunikation nachgeht und dessen Diskurs vornehmlich auf Englisch geführt wird (siehe exemplarisch Neumark et al. 2010; Young 2015; Pettman 2017; Cox 2018; Frühholz & Belin 2018; Eidsheim & Meizel 2019; Meizel 2020), haben wir uns bewusst gegen eine Übersetzung der englischsprachig eingereichten Texte entschieden. Lediglich die Abstracts wurden ins Deutsche übertragen, um möglichst allen Leser\*innen eine schnelle Orientierung über die jeweiligen Inhalte offerieren zu können.

Wenngleich die einzelnen Beiträge in vielerlei Hinsicht Berührungspunkte aufweisen und überwiegend interdependente Aspekte behandeln, ist der Band aus Gründen der Übersichtlichkeit in vier Großabschnitte gegliedert. Im ersten Teil *Stimmforschung heute* widmet sich Katherine Meizel gleich einem ganzen Bündel soziokultureller Faktoren, die es beim Einsatz aktueller Stimmtechnologien mitzudenken gilt. Vor allem geht sie der Frage nach, wie die gesellschaftlichen Machtverhältnisse, in denen Stimmen erklingen, durch Technik nicht nur gespiegelt werden, sondern sich ebenfalls auf technologischem Wege – zum Guten wie zum Schlechten – verändern lassen. Dieser Umstand wird am Beispiel menschlich wirkender Sprachassistentenprodukte und ihrem Verhältnis zu den teilweise dahinter stehenden realen Sprecher\*innen sowie anhand der digitalen Hervorbringung, Bearbeitung und des Verkaufs von Singstimmen expliziert. Zusätzlich bespricht Meizel die Probleme und Chancen sprachlicher und stimmlicher Partizipation während der COVID-19-Krise unter Berücksichtigung neuester Entwicklungen im Bereich der medizinisch-therapeutischen Stimmdiagnostik. Ihre Analyse digital erzeugter, übertragener, archivierter und veränderter Stimmen im Spannungsfeld zwischen vokaler Individualität und kollektiver Technisierung markiert wichtige Themengebiete einer zeitgemäßen Beschäftigung mit stimmlichen Phänomenen. Im Mittelpunkt des Beitrags von Lillian Campesato und Fernando Iazzetta steht die Verbildlichung der Stimme. Dabei wird nicht nur über die Bildhaftigkeit stimmlicher Aktionen, sondern über das imaginative Potenzial von Hörinhalten insgesamt nachgedacht. Ausgehend vom Resonanzphänomen, also dem Vorgang des Mitschwingens, der hier sowohl in einem physikalisch-akustischen als auch kulturtheoretischen Sinne verstanden wird, analysieren die Autor\*innen drei Fallbeispiele aus den Bereichen der bildenden Kunst, der Popmusik und des Geisterglaubens. Unter zusätzlicher Beachtung aktueller Forschungsdiskurse aus Philosophie, Anthropologie, Kognitionswissenschaft und

Sound Studies gelangen sie zu der Einsicht, dass die Momente der Hervorbringung, Repräsentation, Wahrnehmung und Bewertung stimmlicher Laute stets miteinander verknüpft sind und sich auf materieller wie auf affektiver Ebene ständig wechselseitig beeinflussen.

Den Abschnitt *Künstliche Stimmen* eröffnet Marc Böhlen mit einer umfassenden Untersuchung der Sprachsynthese. Aus seinem Text geht hervor, wie historische Versuche der Nachbildung menschlicher Stimmen und damit verbundene Vorstellungen von einer quasi perfekten Imitation menschlicher Eigenschaften entsprechende Vorhaben auch heute noch prägen. Gegenwärtige Manifestationen synthetisch erzeugter Stimmen und ihrer sprachlichen Äußerungen versteht er folglich als alte neue Technologie, die im Unterschied zu ihren frühen Ausprägungen aber nicht mehr nur in isolierten Kontexten vorkommt, sondern den Alltag durchdringt. Mit der zunehmenden Ununterscheidbarkeit von realen und digital generierten Stimmen ergeben sich nicht zuletzt in alltäglichen Nutzungssituationen Möglichkeiten der Täuschung und des Missbrauchs. Dies stellt neue Herausforderungen an den Erwerb und die Vermittlung von Medienkompetenz. Auch Christine Bauer und Johanna Devaney werfen einen kritischen Blick auf die sozialen Effekte ubiquitärer Stimmtechnologien. Beginnend mit der Frage, wie Geschlechtsidentitäten durch den Einsatz synthetischer Stimmen konstruiert werden und wer auf welche Weise daran mitwirkt, stellen sie mannigfaltige Verbindungen zwischen Verkörperungsstrategien und Genderkonzepten bei Sprachassistenzsystemen sowie der digitalen Bearbeitung bzw. Erzeugung von Singstimmen her. Dabei erweist es sich, dass die aktuelle Stimmforschung einer intersektionalen Perspektive bedarf, um die Zusammenhänge zwischen Formen von Vokalität und sich überschneidenden Faktoren wie Gender, Lebensalter, Körperbild, soziale Stellung oder Ethnizität – insbesondere vor dem Hintergrund von Diskriminierungsvorgängen – besser verstehen zu können. Zwei Gespräche mit Praktikerinnen aus der Voice-Branche komplettieren das Themenfeld *Künstliche Stimmen*. Aus der Sicht einer Computerlinguistin, die an der Entwicklung intuitiv handhabbarer Assistenzsysteme arbeitet, berichtet Laura Dreessen von den aktuellen Anforderungen an die stimmlich gebundene Mensch-Maschine-Interaktion. Gesichtspunkte wie die technische Realisierbarkeit nonverbaler Kommunikation und der Datenschutz kommen ebenso zur Sprache wie das gendersensible Design der Assistenzcharaktere und das Manipulationspotenzial KI-basierter Systeme. Stefanie Ray war im Auftrag von Amazon als Autorin für die Persönlichkeit der deutschsprachigen Alexa mitverantwortlich. Sie gewährt seltene Einsichten in den Designprozess und erläutert unter anderem, wie sich die Lokalisierung dieser originär US-amerikanischen Assistenzsoftware vollzog, auf welche Weise ein Sprachroboter eine eigene Identität vorgaukeln kann und wie das natio-

nale Entwickler\*innenteam mit Alexas stimmlich und sprachlich bereits vorgezeichneter Genderkodierung umging. Beide Interviews markieren eine aufschlussreiche Ergänzung (teilweise sogar ein inhaltliches Gegengewicht) zu den Beiträgen, die sich rein wissenschaftlich mit der Gestaltung und den Kommunikationseffekten von Sprachassistenzsystemen befassen.

Der Teil *Künstlerische Stimmen* adressiert vokale Praktiken im Film, in der Popmusik und im Gegenwartstheater. Oksana Bulgakowa beschäftigt sich mit der audiovisuellen Konstruiertheit filmischer Stimmen. Unter besonderer Berücksichtigung des Aspekts einer Trennung von Stimme und Körper – sowohl im Sinne der technischen Aufspaltung als auch narrativ hinsichtlich der Repräsentation gespaltener Persönlichkeiten – vergleicht sie nordamerikanische, russische, französische und deutsche Produktionen. Auf die zeitgenössisch empfundene Artifizialität audiotekhnisch fixierter Stimmen wird ebenso eingegangen wie auf frühe Synchronisationsverfahren und die damit verknüpften Problemfelder der schauspielerischen Identität und Authentizität. Zudem gelingt es Bulgakowa, einzelne Facetten der Filmstimme im Kontext vorgängiger literarischer und musikalischer Traditionen neu zu deuten. Malte Kobels Aufsatz über die vokalen Experimente von Kate Bush demonstriert, dass technisch transformierte Stimmen im Bereich der Popmusik Konzepte von Identität und Authentizität ebenfalls ins Wanken bringen können. Indem sich Bush die in den 1980er Jahren noch junge Sampling-Technologie rasch aneignete, vermochte sie ihre Singstimme zum Ausgangspunkt mannigfaltiger musikalischer Verwandlungen werden zu lassen. In Erweiterung der damals bestehenden Overdub-Verfahren kam es zu akustischen Neuschöpfungen, welche die gewohnten Produktions- und Rezeptionsweisen musikalisch dargebotener Stimmen nachhaltig veränderten. Kobel nimmt dies zum Anlass, gängige Auffassungen stimmlich-körperlicher Individualität kritisch zu hinterfragen, weil Bushs vokale Praktiken im Zusammenspiel von Performativität, Reproduktion und klanglicher Transformation weitaus mehr als nur den Sound ‚ihrer‘ Stimme hörbar machen. Im anschließenden Interview mit Doris Kolesch geht es schwerpunktmäßig um die Rolle der Stimme und des Körpers in Theaterprojekten der letzten Jahre. Neben der Frage, wie der Einsatz digitaler Mittel und speziell die künstliche Intelligenz das Theater und die Performancekunst beeinflussten, wird darüber gesprochen, welche Auswirkungen die COVID-19-Pandemie auf die inszenatorische Praxis bisher gehabt hat und wie sich die Partizipationsmöglichkeiten des Publikums heutzutage gestalten.

Nachdem bislang mehrheitlich neuere Medientechnologien in ihrem Bezug zur Stimme analysiert wurden, schließt das Buch unter der Überschrift *Radio-Stimmen* mit der Betrachtung eines gleichsam beständigen Mediums, jedoch ohne die Effekte digitaler Werkzeuge auf den Rundfunk zu übersehen. Zunächst gehen Dumisani

Moyo und Kundai Moyo detailliert auf die sozialen und politischen Dimensionen des Radios in afrikanischen Gesellschaften ein. Sie erläutern, wie es sich von einem Mittel der Unterdrückung zum Medium der Artikulation von Freiheits- und Unabhängigkeitsbestrebungen entwickelte. In diesem Zuge wird deutlich, dass in weithin ländlich geprägten Regionen mit niedrigem Alphabetisierungsgrad die stimmlich vermittelte Kommunikation einen besonderen Stellenwert genießt und folglich das Radio als eine Art Leitmedium verstanden werden kann. Die persönliche Dimension der direkten Ansprache von Hörer\*innen nebst ihrer Einbeziehung ins Programm sehen die Autor\*innen durch algorithmische Prozesse teilweise gefährdet. Sie plädieren daher für einen ethisch verantwortungsvollen Einsatz künstlicher Intelligenz, und zwar auch, weil die Kombination aus immer besser werdenden Stimmnachbildungen und den Reichweiten des Rundfunks für Desinformationskampagnen gerade in solchen Staaten Verwendung finden könnte, deren Regierungen schon jetzt durch den zunehmenden Import von Überwachungstechnologie ihre Macht zu verstetigen suchen. Bis zu welchem Grad künstliche Stimmen bereits heute das Klangbild deutscher Radiosender prägen, erklärt die Moderatorin und Chefredakteurin Colleen Sanders. Aus dem Interview mit ihr geht hervor, dass Emotionalität und Stimme im Rundfunkalltag ganz eng miteinander verwoben sind und dass Moderator\*innen, die der Hörschaft aufgrund ihres Stimmklangs vertraut sind, vom Publikum auch als real präsente Personen außerhalb des Studios erlebt werden wollen. Welche Erwartungen darüber hinaus an stimmliche Timbres und Sprechweisen herangetragen werden, zeigt sich bei den Themen Werbung, Sexismus und Diversität.

Einige Aspekte medialer Stimmwürfe konnten in der Kürze der Zeit – das Forschungsprojekt war auf ein Jahr begrenzt – nur angerissen werden, andere mussten wir ganz ausklammern. Insbesondere der aktuell angesichts global operierender Streamingdienste an Bedeutung gewinnenden Diskussion um automatisierte Synchronisationstechniken hätten wir uns gern bereits an dieser Stelle gewidmet, werden dies aber sicherlich in anderen Kontexten tun. Das Forschungsprojekt war eine Zusammenarbeit des Musikwissenschaftlichen Instituts der Universität zu Köln mit der Grimme-Akademie und der Grimme Medienbildung. Es wäre – wie auch der vorliegende Band 7 der Schriftenreihe Digitale Gesellschaft NRW – ohne die Förderung des Grimme-Forschungskollegs an der Universität zu Köln nicht möglich gewesen. Dafür danken wir herzlich. Unser besonderer Dank gilt den beteiligten Autor\*innen und Interviewpartnerinnen sowie Judith Kirberger, Katharina Makosch und Elisabeth Turowski für ihre Mitarbeit und Unterstützung.

## Literatur

- Bulgakowa, Oksana (Hrsg.) (2012): Resonanz-Räume. Die Stimme und die Medien. Berlin: Bertz + Fischer.
- Cox, Trevor (2018): Now You're Talking: Human Conversation from the Neanderthals to Artificial Intelligence. London: Vintage.
- Eidsheim, Nina; Meizel, Katherine (Hrsg.) (2019): The Oxford Handbook of Voice Studies. New York: Oxford University Press.
- Frühholz, Sascha; Belin, Pascal (Hrsg.) (2018): The Oxford Handbook of Voice Perception. New York: Oxford University Press.
- Meizel, Katherine (2020): Multivocality: Singing on the Borders of Identity. New York: Oxford University Press.
- Neumark, Norie; Gibson, Ross; van Leeuwen, Theo (Hrsg.) (2010): V01CE: Vocal Aesthetics in Digital Arts and Media. Cambridge: MIT Press.
- Pettman, Dominic (2017): Sonic Intimacy: Voice, Species, Technics (or, How to Listen to the World). Stanford: Stanford University Press.
- Young, Miriama (2015): Singing the Body Electric: The Human Voice and Sound Technology. Farnham: Ashgate.

---

Der vorliegende Aufsatz entstammt der Publikation

Marcus Erbe / Aycha Riffi / Wolfgang Zielinski (Hrsg.)

**Mediale Stimmwürfe**

Perspectives of Media Voice Designs

Schriftenreihe Digitale Gesellschaft NRW, Bd. 7

Kopaed Verlag, 2022

ISBN 978-3-96848-642-0

---