

Einfluss des Uncanny-Valley-Effekts auf die Wahrnehmung von Authentizität und emotionaler Glaubwürdigkeit bei KI-generierter Stimmen

**wissenschaftliche Arbeit im Rahmen des
Moduls “Ton-Seminar WS 25/26”**

im Studiengang
Audiovisuelle Medien (AM-7)
an der Hochschule der Medien

vorgelegt von:
Robin Wolf
Matrikelnummer: 5016548

Prüfer: Prof. Oliver Curdt

im Februar 2026

Gliederung

1. Einleitung.....	2
1.1 Problemstellung.....	2
1.2 Ziel der Arbeit.....	2
1.3 Thematische Abgrenzung und Aufbau der Arbeit.....	2
2. Theoretische Grundlagen.....	3
2.1 Das Uncanny-Valley-Konzept.....	3
2.2 Wahrnehmung von Stimmen.....	3
2.3 Übertragung des Uncanny Valley auf auditive Stimuli.....	4
3. KI-generierte Stimmen: Stand der Forschung.....	4
3.1 Präferenz, Authentizität, Glaubwürdigkeit.....	4
3.2 Uncanny-Valley-Effekt bei Stimmen.....	5
3.3 Akustische Einflussfaktoren.....	5
4. Methodische Ansätze der bisherigen Forschung.....	6
4.1 Untersuchungsdesigns.....	6
4.2 Erhobene Variablen.....	6
4.3 Methodische Grenzen.....	7
5. Diskussion und Fazit.....	7
Literaturverzeichnis.....	8

1. Einleitung

1.1 Problemstellung

Künstlich erzeugte Stimmen sind inzwischen fester Bestandteil unseres Alltags. Sie begegnen uns in Sprachassistenten wie Siri oder Alexa, in Navigationssystemen oder bei Durchsagen im Zug. Durch rasanten technologischen Fortschritt sind diese Stimmen deutlich natürlicher geworden und nähern sich in ihrer Klangqualität zunehmend der menschlichen Sprechweisen an. Dieser Zugewinn an Realismus scheint jedoch nicht automatisch mit einer höheren Akzeptanz bei den Nutzerinnen und Nutzern einherzugehen (Abdulrahman & Richards, 2022, S. 1; Ross, Corley & Lai, 2024, S. 1116). Mehrere Studien weisen vielmehr darauf hin, dass synthetische Stimmen gerade dann ein schwer greifbares, irritierendes Gefühl auslösen können, wenn sie als „fast menschlich“ wahrgenommen werden (Abdulrahman & Richards, 2022, S. 2; Diel & Lewis, 2024, S. 2). Solche Reaktionen werden häufig im Zusammenhang mit dem sogenannten Uncanny-Valley-Effekt diskutiert (Mori, MacDorman & Kageki, 2012, S. 99). Das ursprünglich von Mori für die visuelle Wahrnehmung humanoider Roboter entwickelte Wahrnehmungsmodell beschreibt einen abrupten Einbruch der Akzeptanz, sobald ein künstliches Wesen eine nahezu perfekte menschliche Ähnlichkeit erreicht (Mori et al., 2012, S. 100). Ob sich dieser Effekt jedoch in vergleichbarer Weise auf die auditive Wahrnehmung übertragen lässt, ist bislang nicht eindeutig geklärt. In der aktuellen Forschung zeigt sich hier ein unstimmiges Bild (Ross et al., 2024, S. 1116). Während einige Untersuchungen einen linearen Zusammenhang zwischen Natürlichkeit und Akzeptanz nahelegen, deuten andere Ergebnisse auf komplexere Wahrnehmungsmuster hin (Diel & Lewis, 2024, S. 2).

1.2 Ziel der Arbeit

Ziel dieser Arbeit ist es, den bisherigen Forschungsstand zur Wahrnehmung KI-generierter Stimmen im Kontext des Uncanny-Valley-Effekts systematisch zu analysieren und einzuordnen (Bose, 2024, S. 136). Im Rahmen einer literaturbasierten Untersuchung wird betrachtet, in welchem Maße realitätsnahe synthetische Stimmen unsere Akzeptanz, wahrnehmbare Authentizität sowie emotionale Glaubwürdigkeit beeinflussen können (Abdulrahman & Richards, 2022, S. 1; Bruder, Breda & Larrouy-Maestri, 2025, S. 2). Der Schwerpunkt liegt dabei auf der auditiven Perspektive, um zu prüfen, inwieweit sich das ursprünglich visuelle Uncanny-Valley-Modell auf die Wahrnehmung von Stimmen übertragen lässt (Ross et al., 2024, S. 1116; Diel & Lewis, 2024, S. 2).

Vor diesem Hintergrund wird im Folgenden der Frage nachgegangen, ob KI-generierte Stimmen Uncanny-Valley-Effekte hervorrufen können und wie sich diese Effekte auf die wahrgenommene Authentizität sowie die emotionale Glaubwürdigkeit auswirken.

1.3 Thematische Abgrenzung und Aufbau der Arbeit

Die Arbeit beschränkt sich auf synthetische Stimmen, die als eine Art „auditives Gesicht“ fungieren, soziale Interaktionen ermöglichen und mit einer klaren Identitätszuschreibung verbunden sind (Belin, Fecteau & Bédard, 2004, S. 129). Andere Bereiche, etwa KI-generierte Musik oder ethische Fragestellungen, werden bewusst ignoriert. Diese Einschränkung soll den Fokus auf die psychologische Wirkung der Stimme als besonders sensiblen und subjektiv erlebten Wahrnehmungskanal legen (Kühne, Fischer & Zhou, 2020, S. 3).

Angesichts dieser hohen emotionalen Sensibilität gegenüber stimmlichen Reizen erscheint diese Abgrenzung als sinnvoll (Yoon, Oh & Kent, 2026, S. 1). Inhaltlich beginnt die Arbeit mit einer Einführung in die theoretischen Grundlagen des Uncanny-Valley-Konzepts sowie der menschlichen Stimmenwahrnehmung. Darauf aufbauend wird der aktuelle Forschungsstand zu KI-generierten Stimmen dargestellt. Anschließend folgt eine Analyse zentraler methodischer Ansätze der bisherigen Studien. Den Abschluss bilden eine Diskussion der Ergebnisse, sowie ein zusammenfassendes Fazit.

2. Theoretische Grundlagen

2.1 Das Uncanny-Valley-Konzept

Das sogenannte Uncanny-Valley-Konzept geht auf den japanischen Robotiker Masahiro Mori zurück, der es erstmals 1970 formulierte. Es beschreibt eine psychologische Reaktion auf künstliche Objekte, mit zunehmend menschlicher werdenden Merkmalen im Erscheinungsbild (Mori et al., 2012, S. 98). Nach Mori steigt mit wachsender Menschenähnlichkeit zunächst die Sympathie. Erreicht ein Objekt jedoch eine beinahe perfekten menschlichen Eindruck, kommt es zu einem abrupten Einbruch in unserer Akzeptanz (Mori et al., 2012, S. 98). Dieser Bereich wird häufig mit dem Begriff „Eeriness“ beschrieben und umfasst Empfindungen wie Unheimlichkeit, Irritation oder eine diffuse Form der Abstoßung (Mori et al., 2012, S. 98). Als eine mögliche Ursache dieses Effekts wird eine sogenannte kategoriale Unsicherheit diskutiert. Gemeint ist damit das Misstrauen gegenüber der Lebendigkeit eines Wesens, das sich weder menschlich noch künstlich einordnen lässt (MacDorman & Chattopadhyay, 2016, S. 191). Dieser Gedanke knüpft an frühere Überlegungen von Jentsch an, der das Unheimliche bereits mit solchen Sinnestäuschungen in Verbindung brachte (MacDorman & Chattopadhyay, 2016, S. 191). In der neuen Forschung wird allerdings zunehmend hinterfragt, ob das Uncanny Valley tatsächlich als klar abgegrenzter „Talbereich“ existiert. Empirische Befunde deuten teilweise eher auf lineare oder kontextabhängige Zusammenhänge zwischen steigender Realitätsnähe und wahrgenommener Gefälligkeit hin (Ross et al., 2024, S. 1116; Diel & Lewis, 2024, S. 5). Vor diesem Hintergrund gewinnen alternative Erklärungsansätze an Bedeutung. So führt die Realismus-Inkonsistenz-Theorie das empfundene Unbehagen auf widersprüchliche Merkmalsausprägungen zurück, die einen neurokognitiven Erwartungsbruch auslösen könnten (MacDorman & Chattopadhyay, 2016, S. 191). Evolutionäre Erklärungen werden ergänzend herangezogen, die das Uncanny Valley als Ausdruck eines biologisch verankerten Selbsterhaltungstriebes interpretieren (Mori et al., 2012, S. 100).

2.2 Wahrnehmung von Stimmen

Die menschliche Stimme dient für den Menschen als „auditives Gesicht“ und stellt einen zentralen Klang der phonetischen Umwelt dar, über den soziale Informationen vermittelt werden (Belin et al., 2004, S. 129). Anhand der Stimme ordnen Hörende ihrem Gegenüber unterbewusst Merkmale wie Geschlecht, Alter, emotionalen Zustand und eine Absicht zu (Belin et al., 2004, S. 129). Diese Einschätzungen beruhen auf Ausdrucksmerkmalen des Gesprochenen, insbesondere auf der Gesamtheit der klangbezogenen Eigenschaften (*Prosodie*), die Rückschlüsse auf emotionale Zustände erlaubt (Belin et al., 2004, S. 130). Die Wahrnehmung von Natürlichkeit scheint dabei eng mit der Vorstellung eines lebendigen Körpers verknüpft zu sein (Kühne et

al., 2020, S. 14). Lebendigkeit in der Sprache erklärt sich in Unregelmäßigkeiten wie Atemgeräuschen, natürlichen Pausen und kontextabhängigen Klangfarbenänderungen (Kühne et al., 2020, S. 14; Lavan, Burston & Garrido, 2019, S. 576). Stimmen werden dadurch nicht lediglich als Schallereignis wahrgenommen, sondern automatisch mit Persönlichkeit, Intention und Identität verknüpft. Entscheidend ist dabei, dass gerade diese scheinbaren „Unsauberkeiten“ eine vertraute und glaubwürdige Wahrnehmung begünstigen (Lavan et al., 2019, S. 567; Kühne et al., 2020, S. 14).

2.3 Übertragung des Uncanny Valley auf auditive Stimuli

Das Uncanny-Valley-Konzept wurde ursprünglich für die visuelle Wahrnehmung menschenähnlicher Roboter entwickelt (Mori et al., 2012, S. 98). Ob sich dieses Modell in vergleichbarer Weise auf rein auditive Reize übertragen lässt, ist bislang nicht abschließend geklärt (Ross et al., 2024, S. 1116). Optische und auditive Wahrnehmung unterscheidet sich dabei grundlegend, da visuelle Prozesse sensibler auf Abweichungen von der Norm reagieren als das akustische System (Diel & Lewis, 2024, S. 8). Hierzu führt die unterschiedliche technische Erzeugung der Reize. Künstliche Gesichter erfordern eine vollständige visuelle Rekonstruktion menschlicher Erscheinungsmerkmale, während künstliche Stimmen in der Regel auf realen Sprachaufnahmen basieren. Dadurch lassen sich bestimmte Abweichungen gezielter steuern und kaschieren (Diel & Lewis, 2024, S. 9). Entsprechend zeigen Studien bei isolierten Stimmen meist eine lineare Zunahme der Akzeptanz mit steigendem Realismus (Kühne et al., 2020, S. 469). Ein ausgeprägtes „Tal“ der Sympathie scheint im auditiven Bereich vor allem dann aufzutreten, wenn starke akustische Verzerrungen oder unnatürliche Artefakte vorhanden sind (Diel & Lewis, 2024, S. 9). Vor diesem theoretischen Hintergrund richtet sich der Blick im folgenden Kapitel auf die nachweislichen Befunde zur Wahrnehmung KI-generierter Stimmen und die Frage, inwieweit sich Uncanny-Valley-Effekte im auditiven Bereich tatsächlich folgerichtig nachweisen lassen.

3. KI-generierte Stimmen: Stand der Forschung

3.1 Präferenz, Authentizität, Glaubwürdigkeit

Mehrere Untersuchungen zeigen, dass Hörende menschliche Stimmen im Vergleich zu künstlich generierten im Allgemeinen präferieren und diese gleichbleibend als natürlicher wahrnehmen (Kühne et al., 2020, S. 12). Besonders in Bezug auf lebendiger Intonation schneiden echte Stimmen besser ab. Sie werden dadurch häufiger als sympathisch und zugleich als glaubwürdig eingeschätzt (Kühne et al., 2020, S. 462, 12). Synthetisch erzeugte Stimmen gelten im direkten Vergleich dagegen mehrfach als weniger interessant (Bruder et al., 2025, S. 2). Trotz deutlichen technischen Fortschritten wird KI-generierte Sprache weiterhin oftmals als emotional uneindeutig oder vergleichsweise flach beschrieben (Yoon et al., 2026, S. 2). Entsprechend gelingt die Erkennung emotionaler Zustände bei menschlichen Stimmen präziser, insbesondere bei komplexen Emotionen wie Angst (Yoon et al., 2026, S. 12). Dennoch zeigen aktuelle Befunde, dass moderne KI-Stimmen rein von ihren technischen Mitteln auch durchaus in der Lage sind, real zu wirken (Bruder et al., 2025, S. 3). Hier zeigt sich, dass zunehmende Menschlichkeit und Akzeptanz nicht zwangsläufig korrelieren. Vielmehr reagieren Hörende sensibel auf das Fehlen natürlicher Unregelmäßigkeiten, wodurch synthetische Stimmen häufig als zu perfekt oder monoton wahrgenommen werden (Kühne et al., 2020, S. 9; Yoon et al., 2026, S. 5).

3.2 Uncanny-Valley-Effekt bei Stimmen

Der bisherige Forschungsstand zum Uncanny-Valley-Effekt im auditiven Bereich ist uneinheitlich (Ross et al., 2024, S. 1116). Mehrere Studien berichten, dass nahezu menschlich klingende, künstliche Stimmen bei Hörenden subjektives Unbehagen auslösen können und als wesentlich unheimlicher wahrgenommen werden, als echte menschliche Stimmen (Abdulrahman & Richards, 2022, S. 1-2). Vergleichbare Effekte zeigen sich auch im Kontext virtueller Charaktere, bei denen maschinelle Stimmen häufig zu negativen Gesamteindrücken führen (Gurung, Grant & Hearth, 2024, S. 1265). Als möglicher Auslöser solcher Reaktionen gilt dabei insbesondere das Fehlen feiner rhythmischer Gestaltung, das die Wahrnehmungen von Unheimlichkeit verstärken kann (Yoon et al., 2026, S. 13). Demgegenüber stellen andere Arbeiten die Existenz eines klar ausgeprägten, auditiven Uncanny-Valley-Effekts grundsätzlich infrage. Deren Studien veranschaulichen eine grundsätzlich lineare Zunahme der Akzeptanz mit steigender, menschlicher Ähnlichkeit der Stimme (Kühne et al., 2020, S. 5; Diel & Lewis, 2024, S. 12). Anstelle eines deutlichen Akzeptanzeinbruchs wird hier bei hohem Realismus häufig eine Ebene beschrieben, bei der sich humane und synthetische Stimmen kaum noch signifikant unterscheiden (Ross et al., 2024, S. 1116; Abdulrahman & Richards, 2022, S. 1). Zur Erklärung dieser widersprüchlichen Befunde wurden verschiedene theoretische Ansätze vorgeschlagen. Die Theorie der kategorialen Unsicherheit führt das empfundene Unbehagen auf Zweifel darüber zurück, ob es sich um menschliches oder künstliches Sprechen handelt (MacDorman & Chattopadhyay, 2016, S. 191). Ergänzend werden evolutionäre Erklärungen diskutiert, die das Gefühl nach Unbehagen als Ausdruck eines biologisch verankerten Schutzmechanismus interpretieren (Mori et al., 2012, S. 100). Neuere Ansätze rücken hingegen Abweichungen von typischen organischen Stimmmerkmalen sowie die Verletzung von Erwartungshaltungen der Konsumenten in den Vordergrund (Diel & Lewis, 2024, S. 2; Abdulrahman & Richards, 2022, S. 3).

3.3 Akustische Einflussfaktoren

Wie generierte Stimmen wahrgenommen werden, ist maßgeblich durch deren konkreten akustischen Gestaltungsmerkmale geprägt. Eine zentrale Rolle spielen dabei Sprechweise und Intonation. Diese Faktoren sind zur Einschätzung von Natürlichkeit essenziell (Ross et al., 2024, S. 1118). Eine geringe Tonhöhenvariation lässt Stimmen häufig mechanisch und monoton wirken, während eine stärkere prosodische Dynamik eher mit Vertrauenswürdigkeit und Lebendigkeit assoziiert wird (Ross et al., 2024, S. 1118; Kühne et al., 2020, S. 10; Yu & Levitan, 2024, S. 1240). Sprachrhythmus und zeitliche Struktur beeinflussen den Höreindruck zudem deutlich (Yoon et al., 2026, S. 9). KI-generierte Stimmen weisen dabei oft gleichmäßige zeitliche Abstände zwischen Silben auf, was einen künstlichen Gesamteindruck verursachen kann (Yoon et al., 2026, S. 10). Bei menschlichen Sprecher:innen hingegen variiert die Silbendauer bei jedem gesprochenen Satz flexibel, um bestimmte Inhalte hervorzuheben oder kommunikative Schwerpunkte zu setzen (Yoon et al., 2026, S. 10). Ein weiterer Unterschied zeigt sich in der Vermittlung emotionaler Signale. Synthetischen Stimmen fehlen häufig die dynamische Anpassung an soziale und situative Zusammenhänge, wodurch komplexe Emotionen durch übertriebene Lautstärkenschwankungen dargestellt werden (Yoon et al., 2026, S. 1). Emotionale Identität wird bei menschlichen Stimmen unter anderem eher durch eine natürliche Variabilität physischer Merkmale des Vokaltrakts erzeugt (Lavan et al., 2019, S. 576). Im Kontrast dazu führt eine fehlende

Kohärenz zwischen Stimmklang und inhaltlicher Situationsbezogenheit zu einer Minderung der wahrgenommene Glaubwürdigkeit bei vielen KI-Stimmen (Kühne et al., 2020, S. 12-13). Weil sich statisches, auditives Signal merkbar von vertrauten organischen Stimmmerkmalen abhebt, kann es zu Verwirrungen kommen (Diel & Lewis, 2024, S. 2). Insgesamt scheint das Vertrauen in eine Stimme also eng an eine natürliche Variabilität und einer dem Kontext angemessene Ausdrucksweise gebunden zu sein (Kühne et al., 2020, S. 2).

Die dargestellten Einflussfaktoren verdeutlichen somit, dass Wahrnehmungsunterschiede nicht allein auf den Grad der Menschenähnlichkeit zurückzuführen sind. Vielmehr spielen konkrete Nuancen akustischer Parameter sowie konsistente Überzeugungskraft eine entscheidende Rolle. Vor diesem Hintergrund stellt sich die Frage, mit welchen methodischen Ansätzen diese Effekte in der bisherigen Forschung untersucht wurden.

4. Methodische Ansätze der bisherigen Forschung

4.1 Untersuchungsdesigns

In der Forschung zur Wahrnehmung synthetischer Stimmen kommen überwiegend onlinebasierte Erhebungsmethoden zum Einsatz (Bruder et al., 2025, S. 2). Diese Form der Datensammlung bietet zwar praktische Vorteile, da sie schnelle Rekrutierung großer Stichproben ermöglicht, bringt jedoch auch systematische Einschränkungen mit sich.

Um Unterschiede in der Stimmwahrnehmung zu erfassen, werden sowohl Between-Subjects- als auch Within-Subject-Designs verwendet (Abdulrahman & Richards, 2022, S. 5; Kühne et al., 2020, S. 4). Während erstere Vergleichseffekte zwischen Gruppen ermöglichen, erlauben letztere eine feinere Analysen individueller Bewertungsunterschiede. Als zentrale Erhebungsinstrument dienen in den meisten Studien Rating-Experimente auf Basis von „Likert-Skalen“. Mit ihnen werden Merkmale wie Natürlichkeit, Attraktivität oder Vertrauenswürdigkeit eingeschätzt (Kühne et al., 2020, S. 5; Yu & Levitan, 2024, S. 1243; Bruder et al., 2025, S. 8). Zur Beurteilung der allgemeinen Sprachqualität wird häufig der „Mean Opinion Score“ herangezogen, der ursprünglich aus der Telekommunikationsforschung stammt und inzwischen auch auf künstliche Stimmen angewendet wird (Bruder et al., 2025, S. 2). In Vergleichsstudien werden dabei menschliche Originalaufnahmen direkt synthetischen Stimmen gegenübergestellt. Um Störvariablen möglichst gering zu halten, basieren die dafür verwendeten Stimuli oft auf den Sprachdaten eines einzelnen Sprechers (Ross et al., 2024, S. 1116). Diese Methode erleichtert zwar die Kontrolle, kann jedoch die Übertragbarkeit der Ergebnisse einschränken. Ergänzend kommen Forced-Choice-Aufgaben zum Einsatz, bei denen Stimmen kategorial eingeordnet werden (Yoon et al., 2026, S. 8).

4.2 Erhobene Variablen

In der Forschung zur Bewertung nachgebildeter Stimmen werden mehrere zentrale Wahrnehmungsdimensionen erfasst. Die Natürlichkeit wird dabei meist über Likert-Skalen erhoben, die von „sehr unnatürlich“ bis „sehr natürlich“ reichen (Kühne et al., 2020, S. 4). Die Vertrauenswürdigkeit einer Stimme wird häufig über Einschätzungen der Glaubwürdigkeit und Verständlichkeit konkretisiert, wobei diese beiden Aspekte nicht immer klar voneinander getrennt werden (Yu & Levitan, 2024, S. 1241).

Sympathie beziehungsweise „Likeability“ erfassen viele Studien mithilfe bipolarer Gegensatzpaare wie „unangenehm - angenehm“ (Abdulrahman & Richards, 2022, S. 9). Ein weiteres zentrales Konstrukt stellt die Unheimlichkeit (*Eeriness*) dar, die über verankerte Skalen wie „beruhigend“ und „unheimlich“ gemessen wird. (Abdulrahman & Richards, 2022, S. 1; Mori et al., 2012, S. 98). Die emotionale Glaubwürdigkeit wird vor allem anhand der Korrektheit der Emotionserkennung bestimmt. Teilnehmende Stimmproben in Forced-Choice-Aufgaben werden dabei spezifischen Gefühlszuständen zugeordnet (Yoon et al., 2026, S. 8; Bruder et al., 2025, S. 4). Ergänzend berücksichtigen Studien Dimensionen, wie die Angemessenheit der klanglichen Gestaltung gesprochener Sprache und die wahrgenommene Wärme der Stimme (Ross et al., 2024, S. 1115).

4.3 Methodische Grenzen

Subjektive Beurteilungen sind aufgrund individueller Unterschiede zwischen Teilnehmenden nur eingeschränkt verallgemeinerbar (Bruder et al., 2025, S. 6). Viele Studien weisen zudem eine hohe Stimulus-Abhängigkeit auf, da häufig nur wenige oder sehr spezifische Sprachproben verwendet werden (Bruder et al., 2025, S. 8). Die Nutzung unterschiedlicher Sprechdaten in Vergleichsstudien können darüber hinaus unkontrollierte Störvariablen verursachen (Ross et al., 2024, S. 1116). Auch kulturelle Hintergründe, Akzeptanzpräferenzen und die Sprachkompetenz der Hörenden beeinflussen die Bewertung von Natürlichkeit und Stimmqualität erheblich (Ross et al., 2024, S. 1118; Yoon et al., 2026, S. 2; Kühne et al., 2020, S. 12). Zudem ist die ökologische Validität laborbasierter Studien begrenzt, da reale Alltagssituationen meist nicht situationsgerecht abgebildet werden können (Ross et al., 2024, S. 1118).

Eine uneinheitliche Einhaltung methodischer Standards erschwert schließlich die Vergleichbarkeit der Ereignisse (Bruder et al., 2025, S. 8). Diese Einschränkungen verdeutlichen, dass Befunde zur Wahrnehmung synthetischer Stimmen stets im Kontext ihrer jeweiliger Untersuchungsbedingungen interpretiert werden müssen.

5. Diskussion und Fazit

In Bezug auf die Forschungsfrage zeigt sich, dass KI-generierte Stimmen Uncanny-Valley-Effekte nicht durchgehend hervorrufen, sondern diese offenbar stark von akustischen Merkmalen und dem jeweiligen Nutzungskontext abhängen (Ross et al., 2024, S. 1118). Während einige Studien eine Zuwachsrate der Akzeptanz mit steigender Natürlichkeit nahelegen, berichten andere von einem spürbaren Unbehagen gegenüber nahezu menschlich klingenden Stimmen (Diel & Lewis, 2024, S. 2; Abdulrahman & Richards, 2022, S. 2). Die wahrgenommene Authentizität und emotionale Glaubwürdigkeit scheinen insbesondere dann beeinträchtigt zu sein, wenn feine prosodische Unregelmäßigkeiten fehlen. In solchen Fällen wirken synthetische Stimmen oft technisch präzise, zugleich aber emotional wenig differenziert (Kühne et al., 2020, S. 12; Yoon et al., 2026, S. 2). Eingeschränkt wird die Aussagekraft der Befunde durch methodische Faktoren wie subjektive Ratings und kulturspezifische Stimulus - Sets (Bruder et al., 2025, S. 6; Ross et al., 2024, S. 1118). Zukünftige Studien sollten daher verstärkt längere Interaktionen und komplexere emotionale Ausdrucksformen berücksichtigen (Abdulrahman & Richards, 2022, S. 14; Yoon et al., 2026, S. 13).

Literaturverzeichnis

- Abdulrahman, A., & Richards, D. (2022).** Is natural necessary? Human voice versus synthetic voice for intelligent virtual agents. *Multimodal Technologies and Interactions*, 6(7), 51 <https://doi.org/10.3390/mti6070051>
- Belin, P., Fecteau, S., & Bédard, C. (2004).** Thinking the voice: Neural correlates of voice perception. In G. Calvert, C. Spence, & B. E. Stein (Hrsg.), *The handbook of multisensory processes* (S. 129-135). MIT Press.
<https://doi.org/10.1016/j.tics.2004.01.008>
- Bose, I. (2024).** Künstliche Stimmen: Identifikation und Befremdung. In: S. Voigt-Zimmermann, A. Wessel, M. Haase, (Hrsg.), *Stimme – Sprechen – Theater* (Schriften zur Sprecherwissenschaft und Phonetik, Bd. 30, S. 125-139) Frank & Timme.
https://doi.org/10.57088/978-3-7329-8900-3_12
- Bruder, C., Breda, P., & Larrouy-Maestri, P. (2025).** Attractive synthetic voices. *Computer in Human Behavior: Artificial Humans*, 6, 100211. <https://doi.org/10.1016/j.chbah.2025.100211>
- Diel, A., & Lewis, M. (2024).** Deviation from typical organic voices best explains a vocal uncanny valley. *Computers in Human Behavior Reports*, 14, 100430. <https://doi.org/10.1016/j.chbr.2024.100430>
- Gurung, N., Grant, J. B., & Hearth, D. (2023).** The uncanny effect of speech: The impact of appearance and speaking on impression formation in human – robot interactions. *International Journal of Social Robotics*, 16, 1265-1280.
<https://doi.org/10.1007/s12369-023-00976-4>
- Kühne, K., Fischer, M. H., & Zhou, Y. (2020).** The human takes it all: Humanlike synthesized voices are perceived as less eerie and more likable: Evidence from a subjective ratings study. *Frontiers in Neurorobotics*, 14, 593732.
<https://doi.org/10.3389/fnbot.2020.593732>
- Lavan, N., Burston, L. F. K. & Garrido, L. (2019).** How many voices did you hear? Natural variability disrupts identity perception from voices. *British Journal of Psychology*, 110(3), 576-593. <https://doi.org/10.1111/bjop.12348>
- MacDorman, K. F., & Chattopadhyay, D. (2016).** Reducing consistency in human realism increases the uncanny valley effect: Increasing category uncertainty does not. *Cognition*, 146, 190-205. <https://doi.org/10.1016/j.cognition.2015.09.019>
- Mori, M., MacDorman, K. F., & Kageki, N. (2012).** The uncanny valley. *IEEE Robotics & Automation Magazine*, 19(2), 98-100.
<https://doi.org/10.1109/MRA.2012.2192811>
- Ross, A., Corley, M., & Lai, C. (2024).** Is there an uncanny valley for speech? Investigating listeners' evaluations of realistic synthesized voices. In *Proceedings of Speech Prosody 2024*, (S. 1115-1119). International Speech Communication Association. <https://doi.org/10.21437/SpeechProsody.2024-225>
- Yoon, D., Oh, G. E., & Kent, R. (2026).** Perceiving emotion in human and AI voices: Sensitivity to acoustic cues in Korean speech. *Lingua*, 330, 104083. <https://doi.org/10.1016/j.lingua.2025.104083>
- Yu, Y., & Levitan, S. I. (2024).** What makes a conversational agent sound trustworthy? Exploring the role of acoustic-prosodic factors. In *Proceedings of Speech Prosody 2024*. International Speech Communication Association.
<https://par.nsf.gov/servlets/purl/10536688>

Ehrenwörtliche Erklärung

Hiermit versichere ich, Robin Wolf, ehrenwörtlich, dass ich die vorliegende Seminararbeit mit dem Titel: „Einfluss des Uncanny-Valley-Effekts auf die Wahrnehmung von Authentizität und emotionaler Glaubwürdigkeit KI-generierter Stimmen“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor - SPO (6 Semester), § 24 Abs. 2 Bachelor - SPO (7 Semester), § 23 Abs. 2 Master - SPO (3 Semester) bzw. § 19 Abs. 2 Master - SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Drosendorf, 12.02.2026

5016548, Robin Wolf,



Hilfsmittelverzeichnis

Welches Hilfsmittel wurde eingesetzt?	Wozu wurde das Hilfsmittel eingesetzt?	In welchem Teil der Arbeit wurde das Hilfsmittel eingesetzt?
DeepL	Englisch-Übersetzung	Ganze Arbeit immer wieder
OpenAI ChatGPT	Erstellung der Erstversion der Gliederung (wurde im Laufe der Arbeit angepasst); Hilfe beim APA-7 Zitierstil (einzelne Regeln erklären, ...) Vorschläge zu Quellen (welche Art? Wie viele? ...)	Gliederung Ganze Arbeit immer wieder, Literaturverzeichnis Ganze Arbeit
Zotero	Erstellung Literaturverzeichnis	Literaturverzeichnis
Google Notebook LM	Heraussuchen von Passagen aus Papers und Büchern, Quellenbezogene Recherche (z. B. "wo finde ich eine Aussage zu ...?")	Ganze Arbeit immer wieder
Google/Semantic Scholar	Finden von zitierwürdigen Quellen	Ganze Arbeit