Die Wahrnehmung audiovisueller Synchronität in elektronischen Medien

(The perception of audio-visual synchrony in electronic media)

Berta Luise Heide, Hans-Joachim Maempel **

* TU Berlin, Fachgebiet Audiokommunikation,
berta.heide@gmail.com, hans-joachim.maempel@tu-berlin.de

Kurzfassung

Vor dem Hintergrund einer zunehmenden Medienkonvergenz und audiovisuellen Integration gewinnen Erkenntnisse zur Wechselwirkung von Hören und Sehen an Bedeutung. Ein wesentliches intermodales Qualitätskriterium audiovisueller Medieninhalte ist die Synchronität von Ton und Bild. Vorliegende Untersuchungen zur Synchronitätswahrnehmung erscheinen im Hinblick auf Aktualität, Stimulusauswahl oder Methodik unzureichend. Mit drei experimentellen Verfahren wurde daher die Differenz von physikalischer und perzeptiver Synchronität für verschiedene Typen von Inhalten (Sprache, impulshaftes Geräusch und Musik) untersucht. Als Maße des Synchronitätsempfindens wurden jeweils die Asynchronitätsschwellen, der Punkt subjektiver Synchronität und der Punkt optimaler Bewertung bestimmt. Die generelle Asymmetrie zwischen physikalischer und psychologischer Synchronität wurde bestätigt. Es zeigen sich Unterschiede in der Synchronitätswahrnehmung verschiedener Inhaltstypen und eine ähnliche Synchronitätswahrnehmung von Experten und Laien. Die gefundenenen Asynchronitätsschwellen liegen jedoch wesentlich näher zusammen als die bislang publizierten. Zudem liegt die Schwelle wahrgenommenen Tonvorlaufs im Bereich der physikalischen Tonverzögerung. Nach den vorliegenden Ergebnissen können strenggenommen nur Tonverzögerungen zwischen 15 ms und 75 ms als nicht erkennbar asynchron gelten. Daher wird empfohlen, bei der technischen Übertragung audiovisueller Inhalte einen Toleranzbereich von 0-2 Frames Tonverzögerung einzuhalten und jeglichen Tonvorlauf zu vermeiden.

1. Einleitung

Bei der Produktion, Übertragung und Wiedergabe audiovisueller Medieninhalte kann ein Zeitversatz zwischen optischem und dazugehörigem akustischem Reiz entstehen, der die Bildung eines einheitlichen audiovisuellen Wahrnehmungsinhalts verhindert. Ursachen für solche Asynchronitäten sind z.B. die unterschiedlichen Ausbreitungsgeschwindigkeiten von Licht und Schall, gestörte Synchronisationsprozesse oder ungleiche Transmissionslaufzeiten von Ton- und Bildsignal. Mit der Einführung digitaler Formate im Film- und Broadcastbereich haben sich aufgrund der nicht völlig latenzfreien digitalen Signalverarbeitung die möglichen Quellen für den zeitlichen Versatz von Bild und Ton vermehrt. So können AD-Wandlung, Ton- und Bildbearbeitung, Kodierung, digitale Distribution und Latenz von Wie-

ISBN 978-3-9812830-1-3

dergabegeräten (z.B. Soundkarten oder LCD-Bildschirmen) einen opto-akustischen Zeitversatz verursachen.

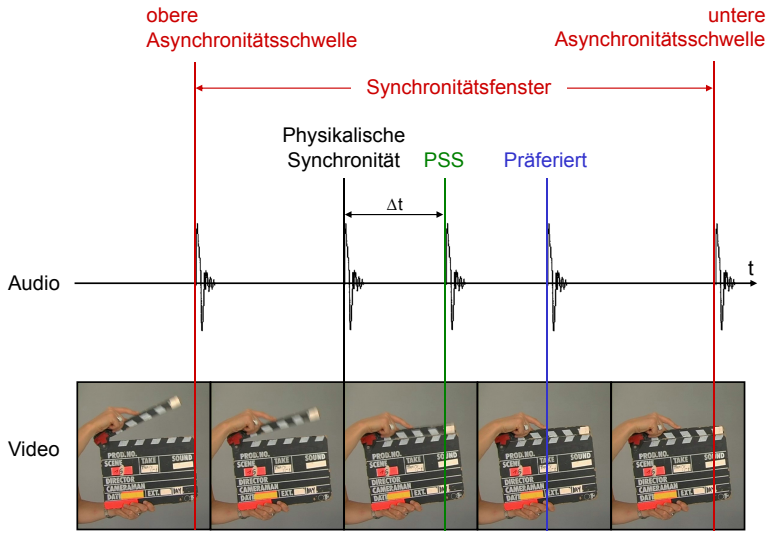

Abb. 1: Etablierte Maße wahrgenommener audiovisueller Synchronität

Mit steigender Asynchronität von Bild und Ton verschlechtert sich jedoch zunehmend die empfundene Qualität und das Verständnis der dargebotenen Inhalte [1]. Zudem entspricht die psychologisch empfundene Synchronität nicht der physikalischen. Daher sollte bekannt sein, wie groß die wahrnehmbare bzw. akzeptable Asynchronität ist. Maße der Synchronitätswahrnehmung sind der Punkt subjektiver Synchronität (PSS), zwei Asynchronitätsschwellen entsprechend dem eben merklichen Unterschied (EMU) sowohl eines (bislang) negativen als auch eines positiven opto-akustischen Zeitversatzes, sowie der ästhetisch präferierte Zeitversatz (Abb. 1). Die Ausprägungen (Δt) dieser Merkmale werden im folgenden bestimmt. Es ist davon auszugehen, dass sie vor allem von der Impulshaftigkeit und der Komplexität des Reizes [2] sowie von der ihm zugeschriebenen Bedeutung beeinflusst werden. Daher wurden in der vorliegenden Untersuchung die Werte für drei verschiedene Audio-Inhalte ermittelt. Die Werte werden als Tonverzögerungen bezogen auf das Videobild angegeben ($\Delta t = t_{Audio} - t_{Video}$). Demnach zeigen negative Werte einen Tonvorlauf an.

2. Stand der Forschung

Auditive und visuelle Wahrnehmung wirken zusammen, indem sie sich ergänzen oder konkurrieren. Eine wichtige Rolle bei dieser Integration auditiver und visueller Information spielen die *colliculi superiores*. Dort wurden in hoher Dichte Neuronen nachgewiesen, die vor allem bei gleichzeitiger Reizung mehrerer Sinne aktiv sind [3] (S. 39-82). Das auditive und das visuelle Sinnessystem interagieren, d.h. die Reizung eines Sinns beeinflusst nicht nur den Wahrnehmungsinhalt desselben, sondern auch den eines anderen Sinns und/oder den gemeinsamen modalitätsunspezifischen Wahrnehmungsinhalt. Solche Interaktionseffekte wurden für zahlreiche Wahrnehmungsaspekte empirisch nachgewiesen, z.B. für die Lokalisation von Ereignissen [4], die Intensitätsbestimmung [3] (S. 15-19) [5], die Erkennung von Sprachlauten [6] oder die Tempobestimmung [7].

Die perzeptive Synchronität von Bild und Ton wurde trotz der rapiden Zunahme der Produktion und Distribution audiovisueller Medieninhalte gerade in jüngster Zeit relativ wenig

ISBN 978-3-9812830-1-3

untersucht. 1875 ermittelte Exner anhand eines Schlages gegen eine Glasglocke und eines elektrischen Funkens die Zeitdifferenz der Empfindungen von Hör- und Sehsinn [8]. Trotz physikalischer Synchronität wurde der akustische Reiz früher als der optische wahrgenommen. Für eine empfindliche Versuchsperson lagen der PSS bei einer Tonverzögerung von 24 ms, die obere und untere Asynchronitätsschwelle bei -15 und 63 ms. Angesichts des historischen Instrumentariums zur Stimulusproduktion und der Untersuchung nur weniger Individuen kann nicht von einer hohen Validität und Reliabilität der Messergebnisse ausgegangen werden.

Jüngere Experimente untersuchten entweder den Einfluss physikalischer Asynchronität auf andere Wahrnehmungsaspekte, z.B. auf das ästhetische Musikurteil zu Videoclips [9], oder ließen in der Regel musikalische Reize außer acht. Dixon & Spitz ermittelten unter Verwendung eines Grenzverfahrens für einen dargestellten Hammerschlag die Asynchronitätsschwellen -75 und 188 ms, für Sprache hingegen -131 und 258 ms [10]. Allerdings entstehen durch die Veränderung der Geschwindigkeit von Magnetbändern Tonhöhenschwankungen, die die Synchronitätswahrnehmung beeinflusst haben könnten. Steinmetz berichtet über Asynchronitätsschwellen von ±80 ms bei Sprache [11]. Rudloff ermittelte unter Verwendung eines Grenzverfahrens einen PSS von 30 ms sowie durch „direkte Testung“ (S. 75) einen präferierten Zeitversatz von etwa 80 ms bis 100 ms für Sprache und Geräusch [12]. Allerdings findet sich keine Angabe dazu, ob und wie die physikalische Synchronität über die gesamte, auch analoge Speichermedien (S-VHS-Videorecorder) umfassende Übertragungsstrecke sichergestellt wurde. Die International Telecommunication Union nennt auf der Basis vergleichender Qualitätsurteile Asynchronitätsschwellen von -45 ms und 125 ms und empfiehlt ein Toleranzfenster zwischen -90 ms und 185 ms [13]. Miner & Caudell ermittelten durch eine Staircase-Methode mit ja/nein-Paradigma ±203 ms für Sprache und ±191 bzw. ±177 ms für verschiedene Geräusche [14]. Lewald und Guski ließen Versuchsteilnehmer u.a. den Grad der Synchronität einfacher künstlicher Stimuli (Ton-Licht-Impulse) bewerten [15]. Allerdings wurde auch die räumliche Position des akustischen Reizes variiert. Die präferierte Tonverzögerung lag bei 65 ms. Conrey und Pisoni bestimmten durch Abfrage von Synchronurteilen mittels Konstanzmethode anhand einer Stichprobe von 30 etwa 20jährigen Personen sowohl für Sprache als auch für einen künstlichen Stimulus (kurzer 2kHz-Sinuston und roter Kreis) einen PSS von jeweils 47 ms [16]. Als Asynchronitätsschwellen ergaben sich -131 und 225 ms für Sprache und -153 und 247 ms für den künstlichen Reiz. Vatakis & Spence untersuchten neben Sprache und Geräusch auch zwei musikalische Inhalte [2]. Die nach dem Konstanzverfahren mit Zeitreihenfolgeurteilen gewonnenen unteren Asynchronitätsschwellen geben die Autoren mit 126 (Geräusch), 154 (Sprache) und 257 bzw. 258 ms (Musik) an. Die PSS differierten je nach spezifischem Inhalt stark und lagen bei -84 (Klavier), -36 (Sprache), 63 (Geräusch) und 65 ms (Gitarre). In einer weiteren Untersuchung ermittelten Vatakis & Spence nach derselben Testmethode untere Asynchronitätsschwellen von 95 ms für gesprochene Silben und ca. 125 ms für kurze Gitarren- und Klavierbeispiele [17]. Die PSS lagen bei -23 ms für die Sprach-, -15 ms für die Gitarren- und 54 ms für die Klavier-Stimuli. Muñoz, Recuero, San Martín, & Fuenzalida bestimmten aus Synchronurteilen Asynchronitätsschwellen von -141 und 158 ms für Sprache und fanden außerdem, dass die Synchronität von Schritten ab drei laufenden Personen nicht mehr erkannt werden kann [18]. Die European Broadcasting Union empfiehlt auf der Grundlage von methodisch nicht näher dargestellten Experimenten mit Sprache als Medieninhalt einen Toleranzbereich von -40 bis 60 ms am Sendeübergabepunkt [19].

Die große Variationsbreite der genannten Werte mag unter anderem auf die verschiedenen Erhebungsmethoden zurückzuführen sein. Obwohl für die Ermittlung von sensorisch be-

ISBN 978-3-9812830-1-3

dingten Unterschiedschwellen kriterienfreie Verfahren (z. B. Forced-Choice-Paradigmen) angezeigt sind, wurden solche von keiner der vorliegenden Untersuchungen angewandt. Zudem unterscheiden sich die dort bemühten kriterienbehafteten Verfahren in der Aufgabe für die Versuchspersonen. So wurden in einigen Untersuchungen Synchronurteile, in anderen Zeitreihenfolgeurteile erhoben. Da die Versuchspersonen im letzteren Falle nicht die Existenz, sondern die Richtung eines opto-akustischen Zeitversatzes erkennen sollen, sind die ermittelten Unterschiedsschwellen keine Asynchronitätsschwellen, sondern vielmehr Asynchronitätsrichtungsschwellen. Van Eijk, Kohlrausch, Juola, & van de Par zeigten, dass beide Maße nicht identisch sind [20], vgl. hierzu auch [21].

In der Forschungsliteratur werden implizit oder explizit auch Hypothesen zur Abhängigkeit der perzeptiven Synchronitätsmaße vom Inhaltstyp und von der Expertise der Versuchspersonen aufgestellt. Als Experten können einerseits aufgrund rhythmischen Trainings Musiker gelten, andererseits Personen, die von Berufs wegen mit der Produktion oder Postproduktion von Ton und Bild befasst sind, typischerweise Cutter und Tonmeister. Während Miner & Caudell über eine geringere Toleranz von Experten gegenüber opto-akustischen Zeitversätzen berichten [14], fanden Vatakis & Spence keine signifikanten Differenzen der Asynchronitätsschwellen und PSS von musikalisch erfahrenen und unerfahrenen Versuchsteilnehmern [2]. Nach Steinmetz halten sich jedoch Produktionsexperten bezüglich der Erkennung von Asynchronität für hochsensitiv [11], und die International Telecommunication Union geht allgemein von einer höheren Empfindlichkeit von Expertenhörern aus [22]. Nach [12] unterscheiden sich die Synchronitätspräferenzen von Medienproduktionsexperten und -laien allerdings nicht. Empirisch begründet sind die Hypothesen, dass der Inhalt des medialen Stimulus sowohl die Asynchronitätsschwellen als auch den PSS beeinflusst [10] [23] [14] [16] [2] [17].

3. Forschungsfragen

Die vorliegenden Untersuchungen zur Synchronitätswahrnehmung wenden unterschiedliche verschiedener Methoden und Paradigmen an und kommen zu uneinheitlichen Ergebnissen. Zudem sind – angesichts einer möglichen medienbedingten Veränderung der Wahrnehmung – die wenigsten von ihnen aktuell, und Musik wurde kaum als Medieninhalt berücksichtigt. Daher wurden in der vorliegenden Untersuchung alle relevanten Maße perzeptiver Synchronität erneut bestimmt, wobei die physikalische Synchronität des Referenzreizes gewährleistet war, eine große und nach Expertise geschichtete Stichprobe gezogen wurde, mehrere prototypische Medieninhalte Berücksichtigung fanden, und verschiedene Testmethoden und -paradigmen (darunter ein kriterienfreies Verfahren) zur Anwendung kamen.

Anhand der so gewonnenen Daten wurden außerdem die oben genannten Hypothesen geprüft, dass Experten und Laien Asynchronität unterschiedlich häufig von Synchronität unterscheiden können und dass die PSS mit dem dargebotenen Inhalt variieren. Ergänzend wird untersucht, ob die Diskriminationsrate und Bewertung von Asynchronität vom Medieninhalt abhängt und ob die Mittelwerte der PSS von Laien und Experten signifikant differieren [24].

4. Methode

Über das Zusammenwirken von Hören und Sehen geben typischerweise Experimente Aufschluss, in denen Versuchsteilnehmern entweder Reize in verschiedenem optoakustischen Präsentationsmodus (z.B. mit/ohne Bild) oder Reize mit optoakustisch divergierenden

ISBN 978-3-9812830-1-3

Schlüsselmerkmalen (cues) dargeboten werden. Im vorliegenden Fall kommt das letztere (*conflicting stimulus*) Paradigma zum Einsatz, indem die Zeitpunkte von optischem und akustischem Ereignis divergieren.

4.1. Versuchsdesign

Es wurden drei Typen von Medieninhalten getestet (Abb. 2): Sprache (Aufsager), impulshaftes Geräusch (Filmklappe) und Musik (Klavier). Der Zeitversatz zwischen Ton und Bild wurde in Schritten von 40 ms zwischen -160 ms und 200 ms variiert, so dass für jeden Inhaltstyp neben der physikalisch synchronen Fassung neun asynchrone Medienbeispiele vorlagen.

Abb. 2: Verwendete Filmbeispiele: Sprache (Aufsager), impulshaftes Geräusch (Filmklappe) und Musik (Klavier)

Die Experimente wurden sowohl mit Experten als auch mit Nichtexperten durchgeführt. Damit ergeben sich drei unabhängige Variablen in einem vollständigen Design: Grad der physikalischen Asynchronität und Inhaltstyp als Messwiederholungsfaktoren und Expertise der Versuchsteilnehmer als Gruppierungsfaktor.

4.2. Erhobene Merkmale

Abhängige Variablen waren die untere und obere Asynchronitätsschwelle, der PSS und der präferierte Zeitversatz (vgl. Abb. 1). Die Maße wurden mittels dreier Testverfahren bestimmt.

Die Asynchronitätsschwellen und der PSS wurden aus den Ergebnissen eines Diskriminationsversuchs nach dem Konstanzverfahren mit einem *two-interval-forced-choice*-Paradigma geschätzt, um eine kriterienfreie Schwellenbestimmung sicherzustellen. Ein physikalisch synchroner Reiz, der als Referenz diente, und ein physikalisch asynchroner Reiz wurden in zufälliger Reihenfolge nacheinander dargeboten. Die Versuchsperson sollte entscheiden, welches Darbietungsintervall den asynchronen Stimulus enthielt. Die Aufgabe wurde mit sechs Zeitversätzen von -120 ms bis 120 ms in zufälliger Darbietungsreihenfolge durchgeführt.

Der PSS wurde außerdem durch einen Herstellungsversuch bestimmt. Die Versuchsteilnehmer wurden gebeten, für jeden Inhalt den optoakustischen Zeitversatz innerhalb eines Bereichs von -120 bis 120 ms in Schritten von 40 ms so einzustellen, dass die Darbietung als synchron empfunden wurde.

Der präferierte Zeitversatz wurde in einem Beurteilungsversuch ermittelt. Die Versuchsteilnehmer sollten die Synchronität aller zehn in randomisierter Reihenfolge einzeln dargebotenen Stimuli mit Schulnoten von 1 (sehr gut) bis 6 (sehr schlecht) bewerten.

ISBN 978-3-9812830-1-3

4.3. Stichprobe

Es nahmen 105 Versuchspersonen (35 weiblich, 70 männlich) im Alter von 15 bis 68 Jahre an der Untersuchung teil. Der Altersdurchschnitt betrug 33 Jahre. Im Hinblick auf eine gegenüber vorliegenden Untersuchungen verbesserte Populationsvalidität wurden Schüler (5%), Studenten (35%), Berufstätige (55%) und Rentner (5%) rekrutiert. Gemäß dem Faktor *Expertise* wurde eine Stichprobe aus 51 Experten- und eine Stichprobe aus 54 Nichtexpertenhörern gezogen. Die Expertengruppe wies Praxiserfahrung im Berufsfeld Film- und Fernsehschnitt und -vertonung auf und bestand aus elf Tonmeistern und 40 Filmcuttern.

4.4. Technischer Versuchsaufbau

Die Testinhalte wurden mit einer Kamera vom Typ Panasonic AGDVX100AE aufgenommen und Ton und Bild in einem AVID Media Composer 2.7 gegeneinander verschoben. Die Wiedergabe erfolgte über ProTools TDM 7.3. Das Bild wurde auf einem CRT-Bildschirm mit einer Diagonale von 20 Zoll angezeigt, der 50 cm von der Versuchsperson entfernt aufgestellt war. Der aufnahmeseitig monofon codierte Ton wurde über zwei Lautsprecher vom Typ Fostex 6301B wiedergegeben, die sich links und rechts unmittelbar neben dem Bildschirm befanden. Die physikalische Synchronität der Übertragungskette wurde durch Verwendung des Prüfgeräts Pharoah Syncheck II sichergestellt [25].

4.5. Durchführung

Die drei Experimente wurden mit den Versuchsteilnehmern im Einzelversuch durchgeführt. Die Aufgaben wurden schriftlich gestellt und mündlich erklärt. Nach Abschluss der Tests wurden soziodemographische und berufliche Angaben zu den Versuchsteilnehmern erhoben. Die Experimente wurden im elektronischen Studio der TU Berlin durchgeführt und dauerten pro Person etwa 30 Minuten.

4.6. Datenauswertung

Die Daten aller Experimente wurden zunächst exploriert und deskriptiv ausgewertet. Für den Diskriminationsversuch wurden pro Zeitversatz und Inhaltstyp die absoluten Fehlerhäufigkeiten aller Versuchspersonen summiert. Die fehlende Häufigkeit für 0 ms wurde durch die nach der Ratewahrscheinlichkeit von 0,5 zu erwartende Fehlerhäufigkeit ersetzt. Die relativen Häufigkeiten ergeben eine empirische Verteilung für jeden Inhaltstyp (N=105) sowie nach Datenzusammenfassung (N=315) für die Gesamtheit der Inhaltstypen. Die Verteilungen wurden durch parametrische Modellfunktionen approximiert, um die Schwellen genauer schätzen zu können.

Gemäß der Forschungsfragen und -hypothesen wurden die Unterschiede der Fehlerhäufigkeiten zwischen Experten und Laien sowie zwischen den Inhaltstypen auf Signifikanz getestet. Der Herstellungs- und der Beurteilungsversuch lieferten intervallskalierte Daten. Die Differenzen der Mittelwerte zwischen den Inhaltstypen und/oder zwischen den Expertisegruppen wurden ebenfalls auf Signifikanz getestet.

5. Ergebnisse

Die Verteilung der Fehlerhäufigkeiten für die Gesamtheit der Inhaltstypen (Abb. 3) zeigt, dass die beiden Richtungen des Zeitversatzes unterschiedlich gut erkannt werden bzw. physikalische und perzeptive Synchronität gegeneinander verschoben sind. Verwechseln bei

ISBN 978-3-9812830-1-3

einem Zeitversatz von -80 ms (Tonvorlauf) nur 7% der Versuchsteilnehmer synchronen und asynchronen Stimulus, so beträgt bei 80 ms (Tonnachlauf) die Fehlerhäufigkeit 47%.

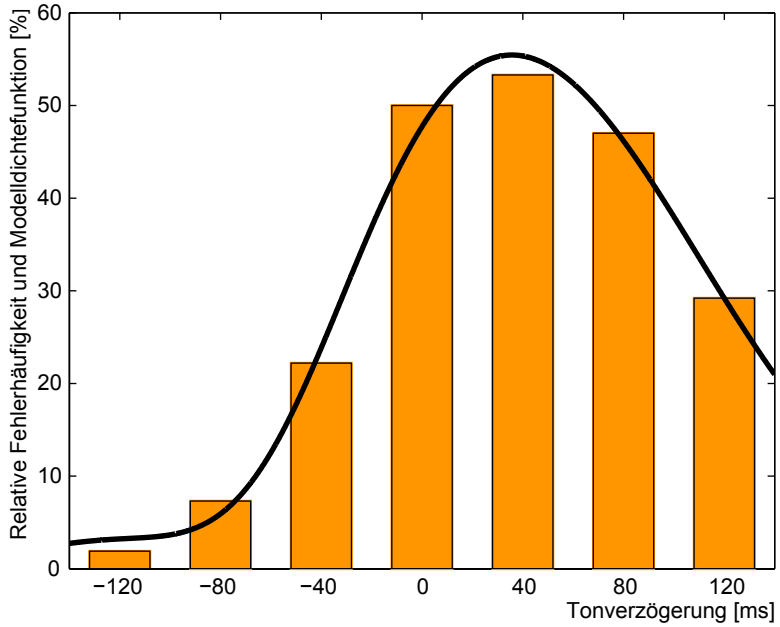

Abb. 3: Relative empirische Häufigkeiten der Verwechslung von asynchronem und synchronem Stimulus für die Gesamtheit der Inhaltstypen (N=315 je Tonverzögerungsstufe) und Modelldichtefunktion (R^2_{adj}=0,96).

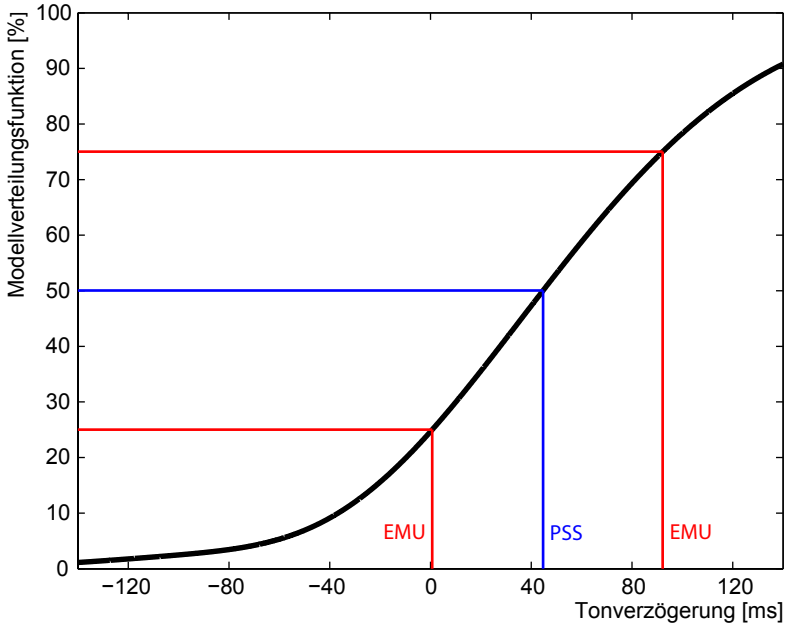

Abb. 4: Modellverteilungsfunktion mit Bestimmung des Punkts subjektiver Synchronität und der eben merklichen Unterschiede (Asynchronitätsschwellen) nach dem 50%-, 25%- bzw. 75%-Kriterium.

Sollen die vorliegenden diskreten empirischen Verteilungen durch kontinuierliche parametrische Modelle beschrieben werden, ist von einer Normalverteilung auszugehen, und An-

ISBN 978-3-9812830-1-3

passungstests (Kolmogorov-Smirnov) bestätigen, dass die empirischen Verteilungen für jeden Inhaltstyp hinreichend normalverteilt sind. Die höchste Varianzaufklärung bieten für Experten (R^2_{adj}=0,97) und Laien (R^2_{adj}=0,97) sowie für Sprache (R^2_{adj}=0,92) und impulshaftes Geräusch (R^2_{adj}=0,98) aufgrund weitgehender Symmetrie jeweils ein einfaches Normalverteilungsmodell, für Musik (R^2_{adj}=1,00) und die Gesamtheit der Inhaltstypen (R^2_{adj}=0,96) aufgrund geringfügiger Linkssteilheit je ein gemischtes Verteilungsmodell bestehend aus zwei summierten Normalverteilungen. Die Modelldichtefunktion für beide Personengruppen und die Gesamtheit der Inhaltstypen ist in Abb. 3 dargestellt.

Abb. 4 zeigt die Verteilungsfunktion der Modellverteilung. Sie stellt die kollektive psychometrische Funktion aller Versuchspersonen für die Gesamtheit der Inhaltstypen dar.

Die eben merklichen Unterschiede werden üblicherweise als diejenigen Differenzen zwischen Vergleichsreiz und Testreiz definiert, für die die psychometrische Funktion die Werte 25% bzw. 75% annimmt [26] (S. 221) [27] (S. 251). Danach liegt die obere Asynchronitätsschwelle bei 0,7 ms und die untere bei 92,2 ms. Der Medianwert der Modellverteilung zeigt den PSS an und beträgt 44,7 ms. Die nach Inhaltstypen und Personengruppen aufgeschlüsselten Asynchronitätsschwellen und PSS sind in Tab. 1 angegeben.

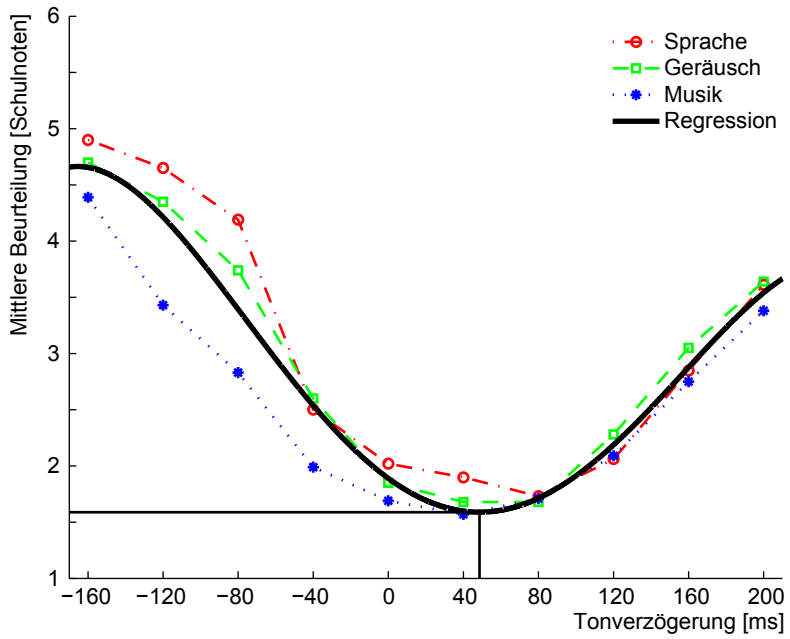

Abb. 5: Mittlere Berteilung der Synchronität verschiedener Zeitversätze in Schulnoten von 1 (sehr gut) bis 6 (sehr schlecht) durch die Gesamtstichprobe (N=105) und Modellfunktion für die Gesamtheit der Inhaltstypen (R^2_{adj}=0,98).

Für Zeitversätze, die nicht nahe am PSS liegen, zeigt sich die Diskriminationsleistung inhaltsabhängig. Für -120, -80, -40 und 120 ms treten zwischen den drei Inhaltstypen signifikante oder hochsignifikante Unterschiede in den Fehlerhäufigkeiten auf (Cochran's Q, N=105). Weiterhin ist die Diskriminationsleistung abhängig vom Faktor Expertise: Die absolute Fehlerhäufigkeit der Expertenhörer war für die Gesamtheit der Inhaltstypen geringer als die der Laienhörer, für das impulshafte Geräusch war der Effekt signifikant (Mann-Whitney-U, Z=-1,997, p=0,046, N=105), für Musik hochsignifikant (Z=-2,802, p=0,005, N=105).

Der Herstellungsversuch lieferte *individuelle* PSS (Tab. 1). Deren Varianz ist geringfügig (η^2=0,032) aber signifikant durch den Inhaltstyp erklärbar (Varianzanalyse für Messwieder-

ISBN 978-3-9812830-1-3

holungen, Greenhouse-Geisser, korrigierte *df*=1,781, *F*=3,470, *p*=0,038). Einzelvergleiche (t-Tests) zeigen, dass nur die Mittelwertdifferenz zwischen Sprache und Musik statistisch bedeutsam ist (*p*=0,044, bonferroni-korrigiert). Die Mittelwerte der PSS von Experten und Laien unterscheiden sich lediglich um 0,8 ms (Sprache), 5 ms (impulshaftes Geräusch) und 3,7 ms (Musik) und damit statistisch nicht bedeutsam (Mann-Whitney-U). Die Mittelwerte der Benotungen aus dem Beurteilungsversuch sind in Abb. 5 für jede Tonverzögerungsstufe und jeden Inhaltstyp dargestellt. Danach wurden Tonverzögerungen von 40 ms bei jedem Inhaltstyp besser bewertet als die physikalisch synchronen Versionen. Ab 120 ms verschlechtert sich die Bewertung wieder deutlich.

Testbedingung		Obere Asynchronitätsschwelle [ms]	Untere Asynchronitätsschwelle [ms]	PSS (Konstanzverfahren) [ms]	PSS (Herstellungsverfahren) [ms]	Präferierter Zeitversatz [ms]
Experten und Nichtexperten	Sprache	14,6	106,9	60,7	40,4	56,6
	Impulshaftes Geräusch	1,7	75,8	38,7	32,4	47,5
	Musik	-11,4	97,7	38,8	24,8	36,9
Gesamtheit der Inhaltstypen	Experten	3,0	77,7	40,3	32,4	44,3
	Nichtexperten	-3,7	95,7	46,0	32,6	50,9
Alle Gruppen und Inhaltstypen		0,7	92,2	44,7	32,5	48,6

Tab. 1: Experimentell ermittelte Maße wahrgenommener audiovisueller Synchronität im Überblick, ausgedrückt als Tonverzögerung bezogen auf das Videobild Δt.

Die mittleren Beurteilungen wurden für jeden Inhaltstyp durch eine polynomiale Regressionsfunktion vierter Ordnung modelliert (R^2_{adj}>0,97). Das Argument deren Minimums ist das Beurteilungsoptimum. Es liegt für Sprache bei der größten, für Musik bei der geringsten Tonverzögerung (vgl. Tab. 1). Als inhaltsübergreifendes Beurteilungsoptimum ergibt sich eine Tonverzögerung von 48,6 ms (vgl. Abb. 5 und Tab. 1). Die mittleren Beurteilungen von Experten und Laien unterscheiden sich für keinen Inhaltstyp signifikant (t-Test).

6. Diskussion

In der vorliegenden Untersuchung wurden Maße wahrgenommener audiovisueller Synchronität auf der Grundlage eines Forced-Choice-Verfahrens, sichergestellter physikalischer Synchronität der Referenzreize, drei medial relevanter Inhaltstypen und einer relativ großen Stichprobe, die nicht nur aus dem studentischen Milieu gezogen wurde, ermittelt. Das so bestimmte Synchronitätsfenster ist wesentlich kleiner, als in allen vorliegenden Publikationen angegeben. Weiterhin ist bemerkenswert, dass die obere Asynchronitätsschwelle nicht im Bereich des Tonvorlaufs, sondern im Bereich der Tonverzögerung liegt. Dies bedeutet, dass physikalische Synchronität zu dem Zeitbereich subjektiven Tonvorlaufs gehört, in dem Asynchronität überwiegend korrekt erkannt wird. Nur für den Inhaltstyp Musik liegt die obere Asynchronitätsschwelle im Bereich des Tonvorlaufs. Ein Vergleich der Inhaltstypen zeigt, dass der PSS für Sprache bei einer etwa 50% höheren Tonverzögerung liegt als die PSS für impulshaftes Geräusch und Musik. Das Synchronitätsfenster ist für Musik am größten und für impulshaftes Geräusch am kleinsten. Die Maße variieren somit leicht inhaltsabhängig.

ISBN 978-3-9812830-1-3

Das Toleranzkriterium für die Asynchronität technisch übertragener Inhalte ist im Sinne einer Qualitätserhaltung sinnvollerweise bei den Asynchronitätsschwellen anzusetzen. Orientiert man sich an den inhaltsübergreifend ermittelten Schwellen, so wäre danach ein Toleranzbereich von ca. 0 ms bis 90 ms zu empfehlen. Soll für keinen der einzelnen Inhaltstypen eine Asynchronitätsschwelle überschritten werden, läge der Bereich sogar nur zwischen 15 ms und 75 ms. Als vereinfachte und handhabbare Empfehlung kann demnach ein Bereich von 0 bis 80 ms (entsprechend 0 bis 2 Frames Tonverzögerung bei 25 Frames pro Sekunde) als wahrnehmungsadäquat gelten. Ein solcher Toleranzbereich würde noch die physikalische Synchronität ein-, aber jeglichen Tonvorlauf ausschließen. In jedem Falle lässt sich feststellen, dass die durch die International Telecommunication Union [13] und die European Broadcasting Union [19] empfohlenen Toleranzbereiche zu groß gewählt sind.

Bezüglich der ermittelten PSS fällt ein Einfluss der Erhebungsmethode auf das Ergebnis auf. Der Herstellungsversuch liefert um ca. 6 ms bis 20 ms geringere Tonverzögerungswerte als der Diskriminationsversuch. Eine Erklärung hierfür könnte sein, dass die Versuchspersonen im Herstellungsversuch aufgrund der linkssteilen psychometrischen Funktion beim Springen zwischen den Tonverzögerungsstufen die obere Asynchronitätsschwelle deutlicher erkennen und deswegen häufiger als Ausgangspunkt für das Herantasten an den PSS wählen. Die Ursache für die geringeren PSS-Werte wäre in diesem Falle ein Ankereffekt.

Nachdem Herstellungsverfahren grundsätzlich keine Trennung von sensorischer und psychologischer Urteilskomponente ermöglichen, Forced-Choice-Verfahren hingegen kriterienfrei sind, müssen die im Diskrminationsversuch ermittelten Schwellen als gültiger angesehen werden als die im Herstellungsversuch gemessenen.

Die im Beurteilungsversuch inhaltsübergreifend präferierte Tonverzögerung weicht vom im Diskriminationsversuch ermittelten PSS-Wert um nicht einmal 4 ms ab. Die von Rudloff [12] festgestellte gravierende und schwer erklärliche Differenz zwischen PSS und präferiertem Zeitversatz von ca. 60 ms konnte also nicht bestätigt werden.

Die inferenzstatistische Analyse zeigt, dass Experten bei der Diskrimination von Asynchronität außer bei Sprache signifikant weniger Fehler machen als Laien, und bestätigt insoweit formal die Untersuchungshypothese einer höheren Sensitivität von Experten. Die bessere Erkennungsleistung der Experten schlägt sich in einem kleineren Synchronitätsfenster und einem geringfüig geringeren PSS nieder (Tab. 1). Jedoch wiesen Experten und Laien im Herstellungsverfahren keine signifikant differenten PSS auf und benoteten im Beurteilungsversuch Asynchronität auch nicht signifikant unterschiedlich. Insoweit kann die Untersuchungshypothese inhaltlich nicht in vollem Umfang bestätigt werden, so dass wir insgesamt allenfalls von marginalen Unterschieden in der Synchronwahrnehmung beider Gruppen ausgehen.

Hingegen bewirkte die Variation von Inhalten im Diskriminationsversuch auf vielen Tonverzögerungsstufen eine mindestens signifikant veränderte Diskriminationshäufigkeit von Asynchroniät, im Herstellungsversuch signifikant differierende PSS und im Beurteilungsversuch hochsignifikant unterschiedliche Benotungen. Damit wird in dieser Untersuchung nicht nur die formale Hypothese einer Inhaltsabhängigkeit der PSS belegt, sondern auch die Annahme einer Inhaltsabhängigkeit der Asynchronitätsschwellen und des präferierten Zeitversatzes empirisch begründet. Gleichwohl sind die in Herstellungs- und Beurteilungsversuch ermittelten Effektgrößen eher als gering einzustufen.

Erklärungen für die Asymmetrie von physikalischer und psychologischer Synchronität verweisen üblicherweise auf die verschiedenen Ausbreitungsgeschwindigkeiten von Licht und Schall in Verbindung mit der Annahme einer mittleren Entfernung relevanter Objekte (z.B.

ISBN 978-3-9812830-1-3

Tiere) in der Menschheitsentwicklung, was zu einer schnelleren auditorischen Reizverarbeitung geführt haben könnte, aus der sich ein sogenannter Gleichzeitigkeitshorizont von etwa 15 m ergibt. Während die Erfahrung einer geringeren Schallausbreitungsgeschwindigkeit als Erklärung für die Richtung der Asymmetrie hinreichen mag, ist eine Erklärung der Größenordnung weitaus schwieriger: Nach Lewkowicz ist das Synchronitätsfenster von Kleinkindern mehr als viermal so groß wie das von Erwachsenen [28]. Insofern die Wahrnehmung von Synchronität also nicht angeboren, sondern erlernt ist, ist die Annahme einer inneren Referenz in Form eines nach evolutionären Notwendigkeiten dimensionierten Gleichzeitigkeitshorizonts schwer haltbar. Nachdem der mittlere Abstand zwischen Kleinkindern und den von ihnen wahrgenommenen relevanten Schallquellen in der Regel wesentlich kleiner als 15 m sein wird, erscheint sogar unwahrscheinlich, dass die Größenordnung der Asymmetrie erfahrungsbasiert ist. Eine neurowissenschaftliche Erklärung der Größenordnung der Asymmetrie liefert der Umstand, dass die Weiterleitung eines optischen Reizes zu den *colliculi superiores* mehr Zeit beansprucht als die Weiterleitung eines akustischen. Neurophysiologische Messungen an Katzen ergaben eine mittlere Laufzeitdifferenz zwischen visueller und auditorischer Weiterleitungsdauer von 63,5 ms [29] (S. 3217).

7. Ausblick

Es ist wahrscheinlich, dass die Synchronitätswahrnehmung von weiteren Faktoren beeinflusst wird, deren Beitrag es systematisch zu untersuchen gilt: Auf akustischer Seite könnten z.B. Lautheit, Spektrum sowie Räumlichkeits- und Entfernungscues, auf optischer Seite z.B. Aufnahmeformat, Brennweite, Bildschirmgröße und Betrachtungsabstand mit dem optoakustischen Zeitversatz als Prädiktor für perzeptive Asynchronität in Wechselwirkung stehen. Nachdem derzeit für LCD-Fernseher mit Bildschirmdiagonalen von 37 Zoll und mehr der stärkste Absatzzuwachs verzeichnet wird [30], ist die Kenntnis des Einflusses der genannten Faktoren auch im Hinblick auf Geräteentwicklung und Rezeptionspraxis von Relevanz. Des Weiteren wäre die Untersuchung des Einflusses der optischen und akustischen Reizkomplexität, der jeweiligen technischen Übertragungsqualität sowie des ästhetischen Eindrucks und des semantischen Gehalts des Perzepts auf die Synchronitätswahrnehmung wünschenswert.

8. Literatur

[1] Spence, Charles, Roland Baddeley, Massimiliano Zampini, Robert James & David I. Shore. „Multisensory temporal order judgments: when two locations are better than one“. In: *Perception and Psychophysics* 65 (2) (2003). S. 318-328.

[2] Vatakis, Argiro & Charles Spence. „Audiovisual synchrony perception for music, speech, and object actions“. In: *Brain Research* 1111 (1) (2006). S. 134-142.

[3] Stein, Barry E. & M. Alex Meredith. *The merging of the senses.* Cambridge/MA: MIT Press, 1993.

[4] Howard, I. P. & W. B. Templeton. *Human spatial orientation.* London: Wiley, 1966.

[5] Stevens, Joseph C. & Lawrence E. Marks. „Cross-Modality Matching of Brightness and Loudness“. In: *Proceedings of the National Academy of Sciences of the United States of America*. 54 (2) (1965). S. 407-411.

[6] McGurk, Harry & John W. MacDonald. „Hearing lips and seeing voices“. In: *Nature* 264 (1976). S. 746-748.

ISBN 978-3-9812830-1-3

[7] Gebhard, Jack W. & G. Hamilton Mowbray. „On discriminationg the rate of visual flicker and auditory flutter“. In: *American Journal of Experimental Psychology* 72 (1959). S. 521-528.

[8] Exner, Sigmund. „Experimentelle Untersuchung der einfachsten psychischen Processe. III. Abhandlung - Der persönlichen Gleichung zweiter Theil“. In: *Archiv für die gesammte Physiologie des Menschen und der Thiere 11,* 1875. S. 403-432.

[9] Schlemmer-James, Mirjam. *Schnittmuster. Affektive Reaktionen auf variierte Bildschnitte bei Musikvideos.* Hamburg: LIT Verl, 2006.

[10] Dixon, Norman F. & Lydia Spitz. „The detection of auditory visual desynchrony“. In: *Perception* 9 (1980). S. 719-721.

[11] Steinmetz, Ralf. „Human Perception of Jitter and Media Synchronization“. In: IEEE *Journal on Selected Areas in Communications* 14 (1) (1996). S. 61-72.

[12] Rudloff, Ingo. *Untersuchungen zur wahrgenommenen Synchronität von Bild und Ton bei Film und Fernsehen*. Dipl.-Arb. Bochum: Ruhr-Univ., Psych. Inst., 1997.

[13] International Telecommunication Union (Hg.). *Relative timing of sound and vision for broadcasting.* Rec. ITU-R BT.1359-1. Genf: ITU, 1998.

[14] Miner, Nadine & Thomas Caudell. „Computational Requirements and Synchronization Issues for Virtual Acoustic Displays“. In: *Presence* 7 (4) (1998). S. 396-409.

[15] Lewald, Jörg & Rainer Guski. „Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli“. In: *Cognitive Brain Research* 16 (3) (2003). S. 468-478.

[16] Conrey, Brianna & David B. Pisoni. „Auditory-visual speech perception and synchrony detection for speech and nonspeech signals“. In: *Journal of the Acoustical Society of America* 119 (6) (2006). S. 4065-4073.

[17] Vatakis, Argiro & Charles Spence. „Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task“. In: *Neuroscience Letters* 393 (2006). S. 40-44.

[18] Muñoz, Roberto, Manuel Recuero, Gonzalo San Martín & Francisco Fuenzalida. „Asynchrony perception in audiovisual productions“. In: *19th International Congress on Acoustics, Madrid, 2-7 September 2007*.

[19] European Broadcasting Union. *The relative timing of the sound and vision components of a television signal*. Rec. EBU-R37-2007. Genf: EBU, 2007.

[20] Eijk, Rob L.J. van, Armin G. Kohlrausch, James F. Juola & Steven van de Par. „Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type.“ In: *Perception & Psychophysics* 70 (6) (2008). S. 955-968.

[21] Hisecke, Oliver. „Audiovisuelle Wahrnehmung: Ergebnisse zweier Experimente zur zeitlichen Integration von Hören und Sehen“. In: Institut für neue Musik und Musikerziehung (Hg.) *Hören und Sehen - Musik audiovisuell: Wahrnehmung im Wandel; Produktion - Rezeption - Analyse*. Mainz et al.: Schott, 2005. S. 53-57.

[22] International Telecommunication Union (Hg.). *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems.* Rec. ITU-R BS.1116-1. Genf : ITU, 1997.

[23] Hollier, Mike P. & Andrew N. Rimell. „An Experimental Investigation into Multi-Modal Synchronisation Sensitivity for Perceptual Model Development“. In: *105th AES Convention, San Francisco, 1998.* Preprint No. 4790.

ISBN 978-3-9812830-1-3

[24] Heide, Berta Luise. *Experimentelle Untersuchungen zur perzeptiven Synchronität von Bild und Ton.* Mag.-Arb. Berlin: Techn. Univ., Fachgeb. Audiokomm., 2009.

[25] Day, Andy. „Pharoah Syncheck II“. In: *Resolution* (4) (2007). S. 34.

[26] Hellbrück, Jürgen & Wolfgang Ellermeier. *Hören. Physiologie, Psychologie und Pathologie.* 2., überarb. u. erw. Aufl. Göttingen: Hogrefe, 2004.

[27] Gelfand, Stanley A. *Hearing. An Introduction to psychological and physiological acoustics.* 4., überarb. u. erw. Aufl. New York: Dekker, 2004.

[28] Lewkowicz, David J. „Perception of auditory-visual temporal synchrony in human infants“. In: *Journal of Experimental Psychology: Human Perception & Performance* 22 (5) (1996). S. 1094-1106.

[29] Meredith, M. Alex, James W. Nemitz & Barry E. Stein. „Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors“. In: *Journal of Neuroscience* 7 (1987). S. 3215-3229.

[30] GfK-Nürnberg e. V. *Innovative Produkte sorgen für Wachstum; Weltweite GfK-Daten zum Markt für Unterhaltungselektronik, IT und Telekommunikation.* Pressemitteilung, 2009.

ISBN 978-3-9812830-1-3