# Verständlichkeit von Fernsehton bei unterschiedlichen Abhörpegeln und veränderbarem Sprache-Hintergrund-Verhältnis

Theresa Liebl[1], Christian Hartmann[1]

[1] *Institut für Rundfunktechnik, 80939 München, E-Mail: liebl@irt.de*

## Einleitung

Die Sprachverständlichkeit des Fernseh-Programms ist für die öffentlich-rechtlichen Rundfunkanstalten ein wichtiges Thema. Seit einigen Jahren werden die Zuschauerbeschwerden dazu immer mehr, was für die Rundfunkanstalten einen dringenden Handlungsbedarf bedeutet. Es werden verschiedenste Ansätze verfolgt, um das Problem zu lösen.

Auf der DAGA 2017 wurde eine Untersuchung des IRT präsentiert, die sich mit der Sprachverständlichkeit des Tonkanalformats 3.0 im Fernsehen beschäftigt hat [1]. Bei dieser Versuchsreihe wurde unter anderem untersucht, ob eine manuelle Steuerung der Lautstärke des Center-Kanals von den Probanden genutzt wird, um eine Verbesserung der Verständlichkeit zu erzielen. Es wurde davon ausgegangen, dass der Center-Kanal einer 3.0-Mischung größtenteils Sprache enthält und dass durch eine Anhebung des Pegels eine verbesserte Verständlichkeit bewirkt werden kann. Die Untersuchung zeigte, dass für alle untersuchten Inhalte eine deutliche Anhebung im Pegel durchgeführt wurde. Diese Anhebung lag im Schnitt bei 2,6 dB.

Bei der Ergebnisauswertung dieser Untersuchung gab es Anzeichen, dass die Abhörlautstärke einen Einfluss auf den gewählten Pegel des Center-Kanals haben könnte. Auch eine in 2014 durchgeführte Untersuchung zur Verständlichkeit von „Voice-Over-Voice“-Passagen im Fernsehen [2] zeigte einen deutlich erkennbaren Einfluss unterschiedlicher Abhörpegel auf das gewünschte Mischungsverhältnis zwischen zwei Sprachspuren in Übersetzungspassagen.

Daher wurde entschieden, den Versuch zur Anpassung der Lautstärke des Center-Kanals einer 3.0-Mischung noch einmal mit verschiedenen fest eingestellten Abhörpegeln zu wiederholen, um die sich abzeichnende Tendenz zu bestätigen.

## Versuchsidee

Aufbauend auf die bereits durchgeführte Untersuchung sollte der Hörversuch den Einfluss der Abhörlautstärke bei der manuellen Einstellung einer gut verständlichen Mischung im Tonkanalformat 3.0 in einer Wohnzimmerumgebung deutlich machen. Dazu sollte bei drei unterschiedlichen Abhörpegeln die gewünschte Lautstärke des Center-Kanals so eingestellt werden, dass eine gute Verständlichkeit der Mischung gewährleistet ist. Als feste Pegel wurden 50, 60 und 70 dB(A) gewählt, da Untersuchungen [2] gezeigt haben, dass 50 dB(A) die ungefähre Abhörlautstärke für TV-Inhalte in einer Wohnzimmerumgebung ist und 70 dB(A) etwa die durchschnittliche Abhörlautstärke beim Mischen im Produktionsstudio.

## Versuchsrahmen

Die Anpassung der Mischungsverhältnisse wurde auf zwei verschiedenen Wiedergabegeräten durchgeführt, welche in der ersten Versuchsreihe [1] bereits verwendet wurden. Ein 3.0-Heimkinosystem von B+W und eine Soundbar von Yamaha. Die gewählte Soundbar war die einzige aus einer Reihe von Testgeräten, welche die manuelle Ansteuerung des Center-Kanals ermöglichte. Bei der 3.0-Anlage konnte die Ansteuerung über den angeschlossenen A/V-Receiver erfolgen.

Durchgeführt wurde der Hörversuch in einem Raum, der akustisch einem guten Wohnzimmer angepasst wurde. Das 3.0-Lautsprechersystem wurden in einem Stereodreieck aufgestellt, die Soundbar wurden in einem Regal unter einem Fernseher platziert, welcher für die optische Darbietung der Testsequenzen genutzt wurde. Die Anordnung der Wiedergabegeräte ist in Abb. 1 zu sehen.

**Abbildung 1:** Hörversuchsumgebung; Aufstellung der Wiedergabegeräte

Die Abhörposition während der Hörversuche befand sich im Sweetspot der 3.0-Aufstellung. Für die Hörversuche wurden die Wiedergabesysteme mit einem akustisch transparenten Stoff verdeckt, um die Versuchsteilnehmer optisch nicht zu beeinflussen.

Gesteuert wurde der Hörversuch durch eine browser-basierte Testsoftware des IRT. Über einen Laptop konnten die Probanden den Befehl zur Wiedergabe der Testsequenzen ausführen und die Lautstärke des Center-Kanals verändern (siehe Abb. 2).

Als Versuchsmaterial wurde real ausgestrahltes TV-Material aus verschiedenen Genres verwendet. Bei der Auswahl wurde speziell auf eine schlechte Sprachverständlichkeit des Ausgangsmaterials und ein breites Spektrum an Inhalten geachtet. Es wurden fünf Testsequenzen zwischen 15 und 30 Sekunden Länge aus den Genres Sport, Dokumentation, Spielfilm und TV-Show ausgewählt.

## Versuchsdurchführung

Jede Testsequenz wurde in den drei gewählten Abhörpegeln 50, 60 und 70 dB(A) sowohl auf der B+W-Anlange, als auch auf der Yamaha Soundbar wiedergegeben. So ergaben sich insgesamt 30 zu bewertende Sequenzen. Die Aufgabe der Probanden war es mit Hilfe eines stufenlosen Reglers (siehe Abb. 2) für jede Testsequenz das gewünschte Mischungsverhältnis zwischen dem Center und dem linken und rechten Kanal einzustellen. Es war dabei auf eine klar verständliche Sprache und eine geringe Höranstrengung zu achten. Die Startposition des Reglers war bei jeder Sequenz auf einer zufälligen Höhe. Die Laustärke des Center-Kanals konnte in einem Bereich zwischen -6 und +6 dB im Vergleich zur Originalmischung verändert werden.

Vor dem eigentlichen Versuch wurde mit jedem Probanden ein kurzes Training durchgeführt, um vorab einen Einblick in das Testverfahren zu geben.

An dem Hörversuch haben 20 Personen zwischen 23 und 60 Jahren mit normalem Hörvermögen teilgenommen.

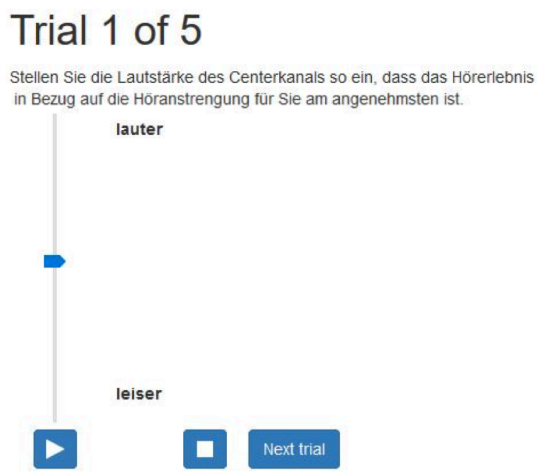


**Abbildung 2:** Oberfläche der Hörversuchssoftware

## Versuchsergebnisse

In Abbildung 3 und 4 sind die Ergebnisse des Hörversuchs in Form von Boxplot-Diagrammen dargestellt. Die Grafiken zeigen die eingestellten Center-Kanal-Pegel für jede Testsequenz auf der Yamaha Soundbar (Abb. 3) und der 3.0-Anlage von B+W (Abb. 4), gemittelt über alle Teilnehmer für die drei Abhörpegel (50 dB(A) in blau, 60 dB(A) in rot und 70 dB(A) in grün). Die rot-gestrichelte Linie in der Mitte, am 0-Punkt auf der y-Achse, gibt die Original-Mischungsverhältnisse an.

Tabelle 1 zeigt die Mittelwerte der eingestellten Mischungsverhältnisse für die drei Abhörpegel auf beiden Wiedergabesystemen, gemittelt über alle Versuchsteilnehmer und Testsequenzen. An den Werten in der Tabelle ist zu sehen, dass auf beiden Wiedergabegeräten ähnliche Ergebnisse erzielt werden konnten und dass sich eine eindeutige Tendenz abzeichnet. Bei einem geringeren Abhörpegel wurde der Pegel des Center-Kanals im Vergleich zu Originalmischung deutlich angehoben.

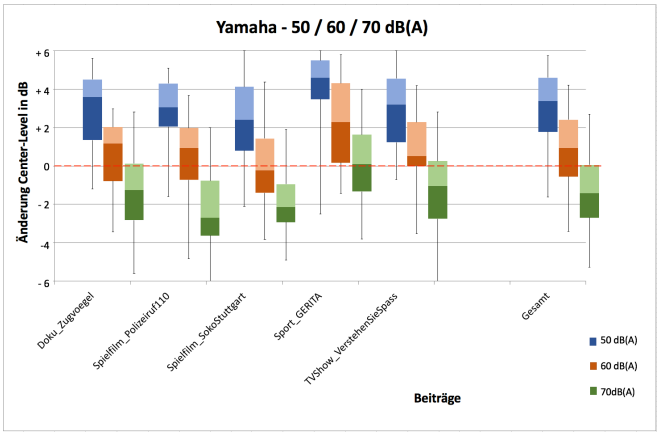


**Abbildung 3:** Ergebnisse des Hörversuchs als Boxplot-Diagramm; Eingestellter Center-Kanal-Pegel auf der Yamaha Soundbar für die drei untersuchten Abhörpegel 50, 60 und 70 dB(A) und alle Testsequenzen

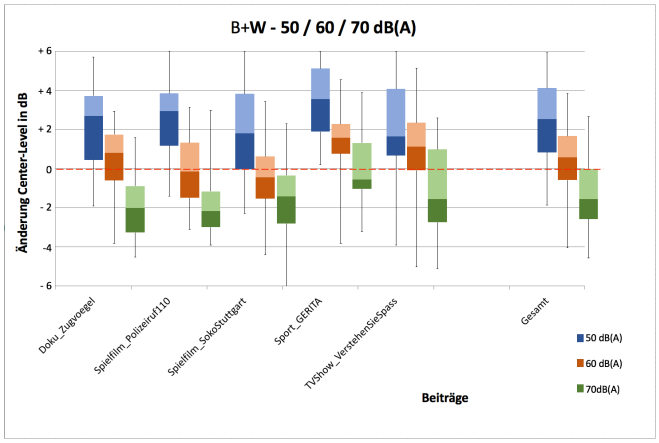


**Abbildung 4:** Ergebnisse des Hörversuchs als Boxplot-Diagramm; Eingestellter Center-Kanal-Pegel auf der 3.0-Anlage von B+W für die drei untersuchten Abhörpegel 50, 60 und 70 dB(A) und alle Testsequenzen

| Wiedergabesystem | Abhörpegel | Mittelwert |
|---|---|---|
| Yamaha Soundbar | 50 dB(A) | +3,0 dB |
| Yamaha Soundbar | 60 dB(A) | +0,8 dB |
| Yamaha Soundbar | 70 dB(A) | -1,3 dB |
| B+W 3.0-Anlage | 50 dB(A) | +2,4 dB |
| B+W 3.0-Anlage | 60 dB(A) | +0,9 dB |
| B+W 3.0-Anlage | 70 dB(A) | -1,3 dB |

**Tabelle 1:** Gesamtmediane für die drei untersuchten Abhörpegel auf den beiden Wiedergabesystemen, gemittelt über alle Versuchsteilnehmer und Testsequenzen

Auch die beiden Boxplot-Diagramme zeigen diese Tendenz deutlich. Bei 50 dB(A) Schalldruckpegel am Abhörort wurde für alle Testsequenzen eine deutliche Anhebung des Center-Kanal-Pegels vorgenommen. Bei 60 dB(A) Abhörpegel verschieben sich die Werte bei allen Sequenzen etwas nach unten, der Center-Pegel bewegt sich hier näher am Null-Punkt, das heißt nahe der Originalmischung. Bei 70 dB(A) liegen die eingestellten Pegel-Werte des Center-Kanals sogar vermehrt unterhalb der Null-Linie. Dies ist bei beiden Wiedergabesystemen in gleicher Weise der Fall.

Zu erkennen ist auch, dass bei der Testsequenz aus dem Bereich Sport der Pegel des Center-Kanals auf beiden Wiedergabesystemen und bei allen Abhörpegeln etwas höher gewählt wurde als bei den anderen Testsequenzen. Dies konnte bereits in der ersten Untersuchung [1] mit variablem Abhörpegel festgestellt werden.

Die Ergebnisse des Hörversuchs zeigen deutlich, dass bei einem geringeren Abhörpegel der Pegel des Center-Kanals bei allen Inhalten deutlich angehoben wurde, um die Höranstrengung zu vermindern und die Sprachverständlichkeit zu verbessern. Eine bei höherem Abhörpegel erstellte Tonmischung kann im Studio also gut verständlich sein, bei einer leiseren Wiedergabe im Wohnzimmer wäre allerdings ein lauterer Center- oder Sprach-Kanal-Pegel nötig, um eine gute Sprachverständlichkeit zu gewährleisten. Um auch zu Hause gut verständliche Tonmischungen sicherzustellen, wäre es daher sinnvoll Mischungen schon im Tonstudio bei einem niedrigen Pegel gegenzuhören und die Verständlichkeit der Sprache zu überprüfen.

Bei der Interpretation der Ergebnisse ist allerdings auch zu beachten, dass bei der Auswahl der Testsequenzen für die Untersuchung speziell auf eine schlecht verständliche Sprache geachtet wurde. Würde man optimal verständliche Tonsignale verwenden, könnten die Ergebnisse dieses Hörversuchs anders ausfallen. Für schlecht verständliches Material und eventuell auch für Zuschauer mit einer Hörminderung würde die Möglichkeit der Pegel-Anhebung des Center-Kanals jedoch einen deutlichen Vorteil bieten und das Fernsehprogramm für bestimmte Zuschauergruppen attraktiver machen.

## Umsetzung der Ergebnisse

Neben dem reinen Mischungsverhältnis haben allerdings auch noch weitere Faktoren einen Einfluss auf die Verständlichkeit von TV-Inhalten [3]. Neben oftmals unsauber aufgenommener Sprache und zu lauten Hintergrundgeräuschen nehmen auch die Abhörbedingungen der Zuschauer einen erheblichen Einfluss. Unterschiedlichste Tonwiedergabegeräte kommen für Rundfunkinhalte zum Einsatz, die räumlichen Bedingungen in den Wohnzimmern beeinflussen die Wiedergabequalität und somit die Verständlichkeit und auch das individuelle Hörvermögen der Zuschauer ist ein ausschlaggebender Faktor.

Um allen unterschiedlichen Anforderungen bei den Zuschauern zu Hause möglichst gut gerecht zu werden und eine optimale Sprachverständlichkeit für alle gewährleisten zu können, wären individuelle Lösungen nötig. Eine Möglichkeit die Verständlichkeit von TV-Produktionen bei den Konsumenten verbessern zu können, wäre die Übertragung eines separaten Sprachkanals, welcher wirklich ausschließlich Sprache enthält.

Eine Umsetzung dieser Lösung ist aktuell allerdings schwierig, da ein separater Sprachkanal von Produktionen meist nicht vorhanden ist und auch nicht einfach gesendet werden kann. Zudem ist die manuelle Steuerung des Center-Kanals einer 3.0- oder auch 5.1-Mischung bisher nur auf wenigen Endgeräten möglich.

**Objektbasierte Audioproduktion**

Eine Lösung für diese Problemstellungen wäre eine Umsetzung mit objektbasierter Audioproduktion. Aktuell finden Fernsehproduktionen rein kanalbasiert statt. Dabei werden mehrere Audiospuren in einer DAW gemischt und in einer festen Kanalbelegung, welche für ein spezifisches Lautsprechersetup ausgelegt ist (z.B. Stereo oder 5.1), gespeichert (siehe Abb. 5). Jeder Audiokanal der fertigen Mischung muss von einem Lautsprecher wiedergegeben werden. Diese festgelegte Mischung wird zum Endnutzer übertragen, welcher am Ende so gut wie keine Möglichkeit mehr hat die Mischung an seine Bedürfnisse anzupassen.

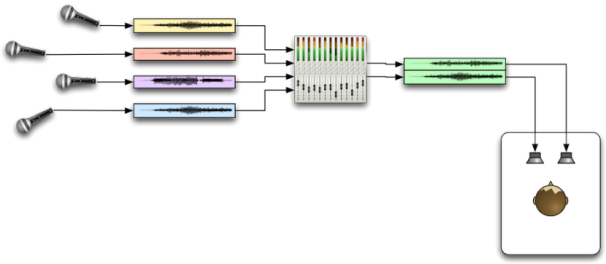

**Abbildung 5:** Schematische Darstellung einer kanalbasierten Audioproduktion

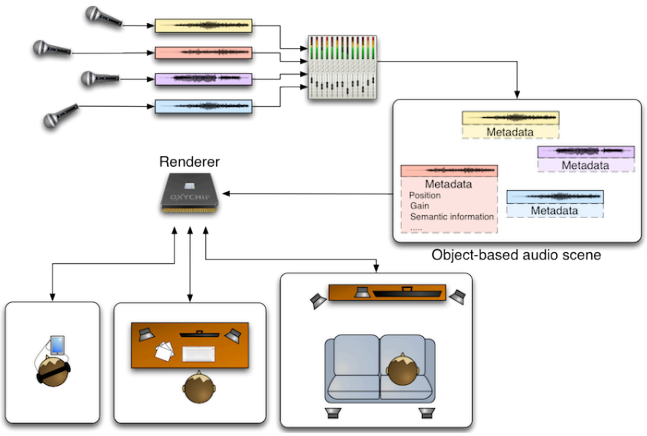


**Abbildung 6:** Schematische Darstellung einer objektbasierten Audioproduktion

Eine objektbasierte Audioproduktion hingegen bietet die Möglichkeit erstellte Mischungen individuell anzupassen.

Hierbei werden die Audiospuren in einer DAW oder einem Editor wie gewohnt gemischt. Gespeichert und übertragen wird die fertige Mischung dann allerdings nicht in einer festen Kanalbelegung, sondern in Form von Einzelspuren, sog. Objekten, mit verknüpften Metadaten als objektbasierte Audioszene. Die Metadaten enthalten die nötigen Informationen für die erstellte Mischung, wie Pegel und Position der einzelnen Objekte. Ein Renderer übernimmt am Ende dann die Zusammensetzung der gespeicherten Szene in die gewünschte Mischung, angepasst auf das jeweils angeschlossene Wiedergabesystem. Dadurch kann gewährleistet werden, dass die empfangene und gehörte Mischung immer dem jeweiligen Format des Endgerätes

entspricht, was unerwünschte Veränderungen in der erstellten Mischung durch Up- oder Downmix-Prozesse verhindert.

Für den hier angedachten Anwendungsfall könnte der Ersteller der Mischung in den Metadaten eines Objektes also festhalten, dass es sich um eine reine Sprachspur handelt, wo das Sprach-Objekt in der Szene platziert sein soll, wie hoch der Pegel dieser Sprachspur in der Mischung sein soll und, dass dem Endnutzer die Möglichkeit gegeben wird diesen Pegel in einem bestimmten Bereich zu verändern. Die Sprachspur wird also nicht fest mit den restlichen Anteilen der Mischung vermischt, sondern als eigenes Objekt in der Gesamtszene übertragen. Informationen über die Möglichkeit der Ansteuerung dieses Objektes können dann von der Produktion bis zum Zuschauer zu Hause mitgegeben werden und der Pegel je nach angeschlossenem Endgerät individuell angepasst werden. Aktuell kann eine solche Anpassung auf Smartphones bereits über eine App erfolgen, bei einer Wiedergabe am PC über einen im Browser integrierten Player. Bei der Wiedergabe über ein Fernsehgerät ist es allerdings vom jeweiligen Gerät abhängig, ob eine Ansteuerung eines einzelnen Objektes möglich ist. Standard-Fernsehgeräte haben dies noch nicht integriert, einige Hersteller treiben die Entwicklung dahingehend aber bereits voran.

Vorteil dieser Art der Produktion und Übertragung wäre also, dass Sprache als eigenes Objekt übertagen wird, Informationen zur Ansteuerung mitgegeben werden können und der Pegel zu Hause individuell angepasst werden kann. Was eine schnelle Umsetzung dieser Technik jedoch erschwert ist, dass eine Anpassung in der Aufnahme und Produktion stattfinden müsste und dass die Unterstützung einer Pegel-Steuerung von Einzelobjekten von den Endgeräten abhängig ist.

## Fazit

Der zusätzlich durchgeführte Hörversuch mit verschiedenen Abhörlautstärken hat eindeutig gezeigt, dass diese einen deutlichen Einfluss auf das gewünschte Mischungsverhältnis zwischen Sprache und Hintergrund hat, wenn eine gute Verständlichkeit von Fernsehinhalten gewährleistet werden soll. Ein zusätzliches Abhören von Mischungen bei geringeren Pegeln im Studio und eventuelles Anpassen der Mischung könnte somit schon eine Verbesserung in der Verständlichkeit für die Zuschauer bewirken. Um jedoch für alle Zuschauer zu Hause eine geringe Höranstrengung gewährleisten zu können, wäre es nötig einen separaten sauberen Sprachkanal zu übertragen, mit der Möglichkeit diesen individuell im Pegel zu verändern. Die Umsetzung dieser Lösung wäre am effektivsten mit einer objektbasierten Audioproduktion und Übertragung möglich.

## Literatur

[1] Liebl, T., Heinz, S., Hartmann, C., „Kann durch das Tonkanalformat 3.0 die Sprachverständlichkeit im Fernsehen verbessert werden?“. DAGA, Kiel, 2017

[2] Liebl, T., Goossens, S., Krump, G., „Verbesserung der Sprachverständlichkeit, speziell bei Voice-Over-Voice-Passagen“, Tonmeistertagung, Köln, 2014

[3] Sprachverständlichkeit im Fernsehen - Empfehlung für Programm und Technik
URL: `https://www.irt.de/de/publikationen/technische-richtlinien.html`