

Verbesserung der Sprachverständlichkeit von fertig gemischtem Stereo-Fernsehton

Felix Matzura², Sebastian Goossens¹, Jens Groh¹, Eva Wilk²

¹ Institut für Rundfunktechnik, 80939 München

² Hochschule für angewandte Wissenschaften, 20099 Hamburg

Einleitung

Eine ARD Arbeitsgruppe beschäftigt sich seit einiger Zeit verstärkt mit der Verbesserung der Sprachverständlichkeit im Fernsehsehton. Die Probleme können bereits im Produktionsprozess entstehen und werden durch altersbedingte Hörverluste beim älter werdenden Publikum noch verstärkt. Bei der Produktion bereits optimale Sprachverständlichkeit zu gewährleisten sollte oberstes Ziel sein, jedoch kann auch die nachträgliche Verbesserung einer fertigen Mischung von Interesse sein, wenn keine Einzelspuren mehr vorliegen oder eine erneute Abmischung aus finanziellen Gründen nicht möglich ist.

Zur Verbesserung der Sprachverständlichkeit von fertig gemischtem Stereo-Fernsehsehton wurden im Rahmen einer Bachelorarbeit am Institut für Rundfunktechnik [1] zwei Ansätze untersucht. Als Signalbearbeitung der fertigen Mischung wurde eine Expansion der Stereo-Mitte bei gleichzeitiger Absenkung der diffusen Schallanteile vorgenommen. Diese Signalbearbeitung wurde auf kurze Beispiele aus dem realen Fernsehprogramm angewendet und in einem Hörversuch durch Probanden bewertet. Ein weiterer Ansatz zur Verbesserung der Sprachverständlichkeit auf der Empfängerseite wurde über den Vergleich verschiedener Abhörsituationen in einem Wohnzimmer verfolgt. Dort wurde der Unterschied zwischen der Wiedergabe über integrierte Fernsehlautsprecher und über in kurzem Abstand vor dem Hörer aufgestellte externe Lautsprecher messtechnisch überprüft und in einem Hörversuch durch Probanden auch mit der Abhörsituation in einem Tonstudio verglichen und bewertet.

Signalbearbeitung fertiger Stereo-Mischungen

In den meisten Mischungen für Film und Fernsehen werden Sprecher oder Dialoge ausschließlich in der Mitte des Stereopanoramas angeordnet. Der erste Ansatz zur Verbesserung der Sprachverständlichkeit besteht darin die Stereo-Mitte mit den enthaltenen Sprachsignalen zu extrahieren um sie gegenüber der restlichen Mischung im Pegel anpassen zu können. Um aus einer zweikanaligen Stereo-Mischung einen Center-Kanal zu extrahieren, müssen die Signalanteile gleicher Phasenlage und Amplitude aus dem linken und rechten Kanal entfernt werden. Das Freistellen sowie auch das Entfernen der Stereo-Mitte aus dem Signal wird im Rahmen dieser Arbeit mit dem *Stereo-Tool* des Audio Editors *GoldWave* durchgeführt. Die Funktionsweise des zugrundeliegenden Algorithmus ist in [2] beschrieben. Neben Sprachsignalen gibt es in der Stereomitte in der Regel weitere Signale, die mit der verwendeten Technik zunächst nicht von Sprachsignalen unterschieden werden können. Hierzu müsste man weitere Algorithmen verwenden, die eine Detektion und Extraktion von Sprache aus einem Signalgemisch ermöglichen. Die Ausgangssignale L' , R' und C' sind in

Abbildung 1 dargestellt. Als weiterer Bearbeitungsschritt folgt eine Expansion der Stereo-Mitte. Die Expansion soll besonders im für die Sprachverständlichkeit wichtigen Frequenzbereich zwischen 1 kHz und 3 kHz angewendet werden. Breitbandige Expansion verursacht deutliche Klangverfärbungen, besonders in den energiereichen unteren Frequenzbändern während eine herkömmliche Multiband Expansion in den energieärmeren oberen Frequenzbändern zu gering wirkt. Dieser Konflikt konnte mit einem am IRT entwickelten Multiband-Expander gelöst werden. Indem die Signalanalyse nicht pro Teilband sondern breitbandig erfolgt, kann die Expansion der gewünschten Teilbänder effektiver wirken.

Nach der Expansion werden die drei Signale L' , R' und C' wieder zu einer zweikanaligen Stereomischung zusammengefügt, wobei die Signale L' und R' im Pegel reduziert werden.

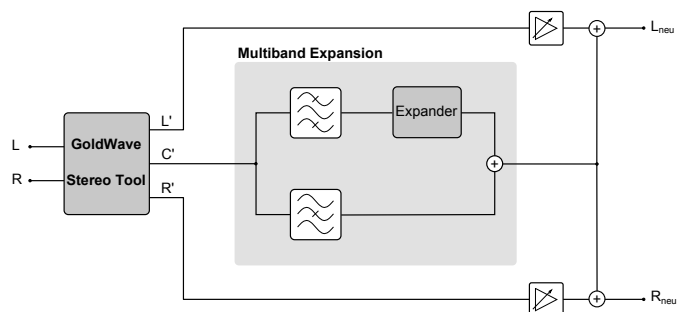


Abbildung 1: Blockschaltbild des Systems zur Sprachverständlichkeitsverbesserung. Im hellgrauen Kasten die Multiband Expansion, die eine gewichtete Expansion in Teilbändern ermöglicht. In der Mischung der Kanäle vor dem Ausgang können L' und R' im Pegel verändert werden.

Untersuchung der Abhörsituation

Während unter optimalen Bedingungen im Studio der Dialog oder die Sprecher noch zu verstehen sind, können kritische Stellen in der Mischung, beispielsweise Dialoge mit Hintergrundmusik, für die Zuschauer unter normalen Abhörbedingungen in ihren Wohnzimmern bereits schwer zu verstehen sein. Diese These wird messtechnisch und in einem Hörversuch überprüft. Zusätzlich wird zur Verbesserung der Sprachverständlichkeit der Einsatz externer Lautsprecher, in Stereo-Standardaufstellung nah vor dem Zuschauer aufgestellt, untersucht.

Für die messtechnische Untersuchung wird der STI-PA verwendet. In einem Raum, dessen Akustik ähnlich zu durchschnittlichen Wohnzimmern ist, werden ein Fernseher und eine Referenzschallquelle in 3 m Abstand sowie ein Paar externer Lautsprecher in 0,9 m Abstand vor dem in Kopfhöhe (sitzend) aufgestellten Messmikrofon positioniert. Die in Tabelle 1 dargestellten Messergebnisse liegen nach der STI-PA-Skala im Bereich guter und ausgezeichneter Sprachverständlichkeit. Jedoch muss berücksichtigt werden, dass das

verwendete Testsignal ein ideales Signal darstellt. Kommt es bei Fernsehton zu Problemen mit der Sprachverständlichkeit, so ist häufig schon das Ausgangssignal verhallt oder durch Störgeräusche beeinflusst. Betrachtet man die Messergebnisse relativ zueinander, deutet sich der negative Einfluss der in den Fernseher integrierten Lautsprecher bereits im Vergleich zur Referenzschallquelle an. Eine Zunahme der Sprachverständlichkeit zeigt sich für die Verwendung externer Lautsprecher gegenüber den integrierten Fernsehlautsprechern.

Tabelle 1: Ergebnisse der STI-PA-Messung verschiedener Abhörsysteme in unterschiedlicher Entfernung zum Hörer

Abhörsystem	Entfernung	STI-PA
Referenzschallquelle TalkBox	3 m	0,74
integr. Fernsehlautsprecher	3 m	0,70
externe Lautsprecher	0,9 m	0,84

Der Unterschied zwischen externen Lautsprechern und in den Fernseher integrierten Lautsprechern wird im Hörversuch mit der Abhörbedingung in einem Tonstudio verglichen. Dazu wurden die kopfbezogenen Impulsantworten der jeweiligen Systeme aufgenommen und anschließend ausgewählte Beispiele aus dem Fernsehprogramm über Kopfhörer auralisiert.

Hörversuch

Am zweiteiligen Hörversuch nehmen 18 Probanden zwischen 21 und 62 Jahren mit einem Durchschnittsalter von 38 Jahren teil. Der Versuch wird mit der Software *STEP* im Modus A/B-Vergleich durchgeführt. Anhand einer Skala von -3 bis +3 (vgl. Abb. 2 und 3) bewerten die Probanden, welches Beispiel aus A und B zu welchem Grad „besser“ ist. Die Probanden werden angeleitet, besonders die Parameter Sprachverständlichkeit und Höranstrengung zu bewerten und weitere klangliche Aspekte von der Bewertung auszuschließen. Die 15-20 sek langen Audiobeispiele enthalten mindestens einen ganzen, sinnzusammenhängenden Satz und werden ohne Bild und in ruhiger Umgebung über Kopfhörer wiedergegeben.

Im ersten Teil des Hörversuchs (vgl. Abb. 2) werden die verschiedenen Abhörsituationen miteinander verglichen. Jeweils acht gleiche Audiobeispiele werden in drei verschiedenen Abhörpaarungen gegenübergestellt.

- Tonstudio - externe Lautsprecher
- Tonstudio - Fernsehlautsprecher
- externe Lautsprecher - Fernsehlautsprecher

Dem Tonstudio wird in beiden Vergleichen eine bessere Sprachverständlichkeit bescheinigt, wobei der Unterschied zum Abhören über die internen Fernsehlautsprecher erwartungsgemäß deutlicher ausgeprägt ist. Verbesserte Sprachverständlichkeit wird auch den externen Lautsprechern gegenüber den integrierten Fernsehlautsprechern bestätigt. Im zweiten Teil des Hörversuchs (vgl. Abb. 3) wird die Signalbearbeitung der fertigen Stereo-Mischungen im Vergleich zu den unbearbeiteten Beispielen nach dem oben beschriebenen Prinzip bewertet.

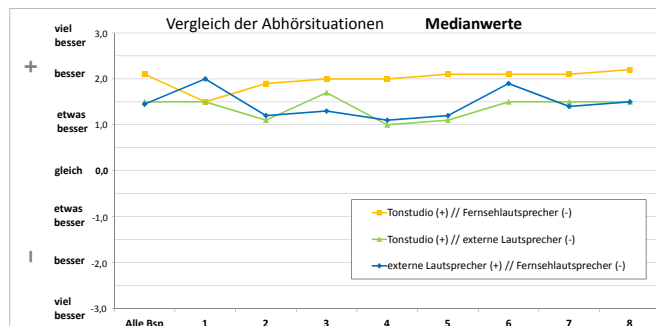


Abbildung 2: Hörversuchsergebnis zum Vergleich verschiedener Abhörsituationen

Eine durchschnittlich leichte Verbesserung der Sprachverständlichkeit zeigt sich für acht der elf bearbeiteten Beispiele. Für zwei Beispiele wird im Mittel keine Veränderung angegeben, während ein bearbeitetes Beispiel im Mittel schlechter bewertet wird als das Original. Die Beispiele, denen eine stärkere Verbesserung bescheinigt wird, zeichnen sich durch einen hohen Anteil unkorrelierter Hintergrundgeräusche oder Atmo aus. Mit dem Absenken des Pegels der Side-Kanäle wird hier folglich die größte Verbesserung des Signal-Stör-Abstands erreicht. Dies zeigt, dass die Mittenextraktion zum großen Teil am Erfolg der Sprachverständlichkeitsverbesserung beteiligt ist.

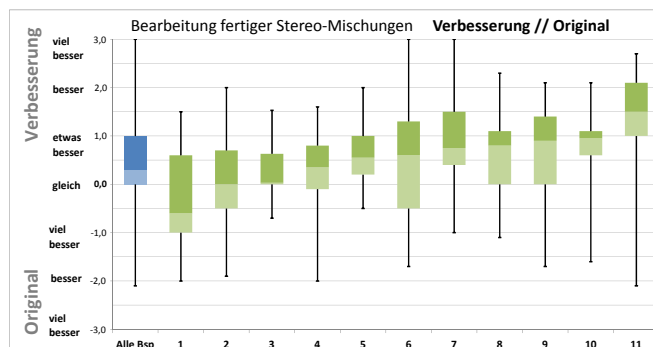


Abbildung 3: Hörversuchsergebnis zur Bearbeitung fertiger Stereo-Mischungen

Fazit und Ausblick

Der Grad der Verbesserung durch das beschriebene Verfahren für fertige Stereo-Mischungen ist nachweisbar, jedoch noch relativ gering. Eine Verbesserung der Extraktion des Center Kanals sollte eine Reduktion von Artefakten und eine verbesserte Kanaltrennung beinhalten. Der Einsatz von Algorithmen zur Spracherkennung kann hierbei ein geeignetes Mittel sein, um ausschließlich die Sprache aus dem Signal zu extrahieren. Die Untersuchungsergebnisse zur Abhör-situation können dazu dienen, ein Bewusstsein für das Problem bei den Toningenieuren zu schaffen, und führen eventuell dazu, dass Mischungen häufiger auch unter nicht-idealen Abhörbedingungen geprüft werden.

Literatur

[1] Matzura, F.: Untersuchungen zur Verbesserung der Sprachverständlichkeit von fertig gemischtem Stereo-Fernseh-ton, HAW, Hamburg 2013

[2] Lee, A. – Virtual Dub: The “center cut” algorithm, URL: <http://www.virtualdub.org/blog/pivot/entry.php?id=102>, 2006