STUDIE ZUR LÖSBARKEIT DES PROBLEMS STARKER PEGELSCHWANKUNGEN IM HOME-ENTERTAINMENT

Georg Schmidt[1]*, Ingo Siegert*[2]

[1]*Fakultät für Elektro- und Informationstechnik, Otto-von-Guericke-Universität Magdeburg*
[2]*Mobile Dialog Systeme, Fakultät für Elektro- und Informationstechnik, Otto-von-Guericke-Universität Magdeburg*
georg.schmidt@st.ovgu.de, ingo.siegert@ovgu.de

Kurzfassung: Das Phänomen erheblicher Pegelunterschiede in der Audiospur ist allgegenwärtig. Diese tritt nicht nur nur beim regulären Fernsehen oder Abspielen von DVDs auf, sondern vermehrt auch bei der Verwendung von Streaming-Diensten. Es ist oft nicht möglich, eine erträgliche Lautstärkeeinstellung zu finden, bei der alle Dialoge verstanden werden können und die Musik- oder Actionszenen nicht zu laut sind sowie Werbeblöcke als zu störend empfunden werden. Dies führt insbesondere für ältere Menschen und Menschen mit Hörschwäche zu Problemen. In diesem Beitrag werden zwei Kompressionsverfahren als Lösung des Problems diskutiert. Ziel ist es, eine Methode zu entwickeln, mit der die (Lautstärke-)Pegel der audiovisuellen Medien konstanter gehalten und Aspekte der Verständlichkeit und der Immersion berücksichtigt werden. Die erste Variante nutzt als Ansatz die dynamischen Kompression (Dynamic Range Compression - DRC). Der zweite Ansatz ist eine Kombination aus dem Clustern der Audiospur über die Berechnung leistungsbasierter Merkmale mittels Mel-Filterbank, gekoppelt mit einer Voice Activity Detection (VAD). Die Kompression wird anschließend anhand der identifizierten Cluster durchgeführt. Die VAD wird genutzt um den Dialoglevel auf einer konstanten Lautstärke zu halten. Beide Verfahren werden mit einer Hörerstudie evaluiert. Es wird durch die Bewertung von ausgewählten Filmausschnitten bestätigt, dass die Lautstärke nach Kompression „insgesamt angenehmer“ ist. Ein eindeutiger Vorteil der zweiten Methode gegenüber der DRC wird durch die Hörerstudie nicht bestätigt. Das finale Ziel ist die Entwicklung eines kostengünstigen Echtzeitsystems, welches Audiospuren im Home-Entertainment hinsichtlich ihrer Lautstärke bewerten und als zusätzliche Steuereinheit für die Lautstärkeregelung fungieren kann.

1 Einleitung

Nahezu Jeder hat schon einmal die Erfahrung gemacht, während eines Films (Sei es eine DVD oder ein Spielfilm im Fernsehen) zur Fernbedienung greifen zu müssen um die eingestellte Lautstärke zu modifizieren. Das Problem ist weit verbreitet und kann von unterschiedlichen Personen unterschiedlich stark wahrgenommen werden. Ein Grund dafür ist aus der Film- und Musikindustrie unter der Bezeichnung „Loudness War“ (Tendenz zu immer höheren Lautstärkepegeln) bekannt [1]. In dieser Arbeit soll es jedoch nicht um den Ursprung dieses Phänomens gehen, sondern darum einen lösungsorientierten Ansatz zu finden. Also eine automatische Lautstärkeeinstellung zu entwickeln oder Pegelschwankungen gering zu halten, bei der die meist zu leisen Dialoge verstanden werden können, die Musik- oder Actionszenen aber nicht ohrenbetäubend laut sind und Werbeblöcke nicht als störend empfunden werden.

Um das Problem adäquat zu adressieren, sollen zuerst die Begriffe Pegel, Lautstärke und Lautheit eingeführt werden. Das Gehör ist in der Lage einen großen Bereich an verschiedenen Schalldrücken wahrzunehmen. Dabei ist die Wahrnehmung von Schalldrücken durch die variable Empfindlichkeit nicht gleichbedeutend mit der empfundenen Lautstärke. Der *Pegel* stellt ein logarithmisches Verhältnismaß dar (Dezibel-Skala), um den gesamten Bereich in überschaubaren Schritten darzustellen [2, S.35]. Entscheidend ist, an welchem Punkt der Signalkette das Signal betrachtet wird. Stammt das Signal von einem Online Streaming-Dienst, liegt es als digitaler Pegel dB_{Fs} (nicht zu verwechseln mit dB_{Spl} - Sound Pressure Level, physikalischer Schalldruck) vor. Die Einheit dB_{Fs} bezieht sich auf „Full Scale“ und damit auf die Wortbreite des Digital-Analog Wandlers (DAC) des Geräts. Der Digitale Pegel kann aus diesem Grund nur $0\mathrm{dB}_{Fs}$ (maximal) und negative Werte annehmen. In diesem Beitrag wird eine digitale Signalverarbeitung diskutiert. Deshalb bezieht sich die Bezeichnung Pegel stets auf digitalen Pegel.
Wird von *Lautstärke* gesprochen, ist damit oft das Phänomen gemeint, dass ein Signal bei gleichem Pegel jedoch unterschiedlicher Frequenz nicht als gleich Laut wahrgenommen wird [3]. Um diesem Punkt zu entsprechen wurde die Einheit Phon eingeführt. Es werden Bewertungskurven definiert, bei welcher die Frequenzen als „gleich Laut“ wahrgenommen werden [4]. Außerdem ist festzustellen, dass das Gehör einzelne Schallintensitäten von akustischen Signalen in schmalen Frequenzbändern zu Gesamtintensitäten zusammenfasst (engl critical bands), welche durch die nicht uniforme Mel-Filterbank beschrieben werden können [5, S.94]. Wenn im Folgenden umgangssprachlich von Lautstärke/Lautstärkeregelung gesprochen wird, ist hier von einer Einstellung des Pegels die Rede, welche in der implementierten Signalverarbeitung Aspekte der Frequenz (bezogen auf Lautstärke) berücksichtigt. Die *Lautheit* definiert ein weiteres subjektives Maß. Die Einheit Sone berechnet sich aus der Einheit Phon und bildet die Wahrnehmung von Lautheit linear ab.
Die Bewertung von Pegel/Lautstärke ist nicht nur von Frequenz, sondern auch stark von individuellen Vorlieben/Geschmack/Alter und vielen weiteren Faktoren abhängig. Aus Studien, wie [6] geht hervor, dass Zuschauer empfindlicher beim Wechsel von leisen auf lauter werdenden Inhalten reagieren als umgekehrt. Außerdem geht daraus hervor, dass die Toleranz für unterschiedliche Pegel von gesprochenem Inhalt deutlich geringer ist als von beispielsweise Umgebungsgeräuschen oder Musik. Aus dieser Überlegung heraus entsteht die Strategie Sprache getrennt zu betrachten und den Pegel auf einen konstanten (Maximal-)Wert zu bringen. Ein weiterer Punkt ist die Anpassung von subjektiv zu starken Unterschieden im Dynamikumfang des Audiostreams (das Action-Scenen-Problem). Während einige Zuschauer lautere Pegel in Action-Scenen als notwendig für die Immersion empfinden, können Andere die Unterschiede als extrem störend empfinden. Eine Möglichkeit den Dynamikumfang eines Audiostreams zu senken ist die Dynamic range compression kurz „DRC “ [7]. Durch die verschiedenen Parameter des Kompressors (Threshold, Make-Up Gain, Knee Width usw.) kann für jeden Zuschauer eine individuelle Regulierung der Pegelschwankungen ermöglicht werden.Es wird mit der vorgestellten Signalverarbeitung die gleichmäßige Anpassung des Pegels angestrebt, unter zusätzlicher Berücksichtigung von Erkenntnissen der Psychoakustik zur Wahrnehmung von Frequenzen (Lautstärke/Lautheit).

2 Methoden

Aufnahme und Beschreibung der Testdaten
Um das Audiosignal nicht durch eine zusätzliche AD-Wandlung zu verändern, wird für die Analysen auf digitale Daten von verschiedenen Streaming-Plattformen, wie Netflix oder YouTube gesetzt. Zur Aufnahme dieser wurde mit der WASAPI loopback Funktion (virtuelles Audiokabel) gearbeitet und eine identische Aufnahmelautstärke gewählt, unter der Clipping ausgeschlossen ist. Als Testdaten werden Ausschnitte aus Actionenfilmen, Fernsehserien, YouTube

Videos und Online-Fernsehen unterschiedlicher Diensteanbieter aufgenommen. Die Daten liegen nach der Aufnahme als 2-Kanal 16-bit WAV mit 44.1kHz Samplingrate vor. Eine erste manuelle Inspektion der Daten zeigt, dass einige Ausschnitte aus der gleichen Folge einer Serie (BB2) sich zwischen Dialoglevel und Actionszenen bis zu -35dB$_{Fs}$ unterscheiden (-38dB$_{Fs}$ im Dialog vs. -3dB$_{Fs}$ in Actionszenen). Beim Vergleich der Actionszenen unterschiedlicher Filme/Serien ergeben sich Unterschiede von bis zu -18dB$_{Fs}$ (Actionszene F5 mit -21dB$_{Fs}$ vs. Actionszene BB1 -3dB$_{Fs}$), die Pegel in Dialogen unterscheiden sich um bis zu -12dB$_{Fs}$ (-23dB$_{Fs}$ im BS vs. -35dB$_{Fs}$ in BB1).
Die erheblichen Pegelunterschiede innerhalb einzelner Medien, sowie im Vergleich ähnlicher Ausschnitte unterschiedlicher Medien, verdeutlicht die Notwendigkeit einer objektiven Bewertungsmethode der Daten. Um subjektive Einflüsse der Autoren auf den Bewertungsprozess auszuschließen, wird die automatische Methode zum Clustern der Daten eingeführt. Die verwendeten Testdaten sind in Tabelle 1 aufgeführt.

Tabelle 1 – Aufgenommene Testdaten.

Medium	**Video/Film/Sender**	**Dauer der Aufzeichnung**	**ID**
Netflix	Breaking Bad	Staffel 1 Folge 1, Staffel 2 Folge 7, Staffel 3 Folge 2, Staffel 4 Folge 5, Staffel 5 Folge 16	BB1-BB5
	Departed, Unter Feinden	1 h 19 min	DP
	Fast and Furious 5	1 h 2 min	F5
	Bridge of Spies	1 h 45 min	BS
	Who am I	1h 17 min	WH
YouTube	Fritz Meinecke	45 min	FM
	Music is Win	56 min	MIW
	WDR Doku	43 min	WDR
Online TV	RTL	35 min	RTL
	Pro7	26 min	PRO
	Vox	33 min	VOX

Signalverarbeitung und Lautheitsbestimmung

Um Erkenntnisse der Psychoakustik in den Algorithmus der Signalverarbeitung mit einzubeziehen, wird mittels der Mel-Filterbank Methode [5, S.95] eine Klassifikation für Audio Signale entsprechend ihres Pegels abgeleitet. Der Ablauf ist in Abbildung 1 dargestellt.

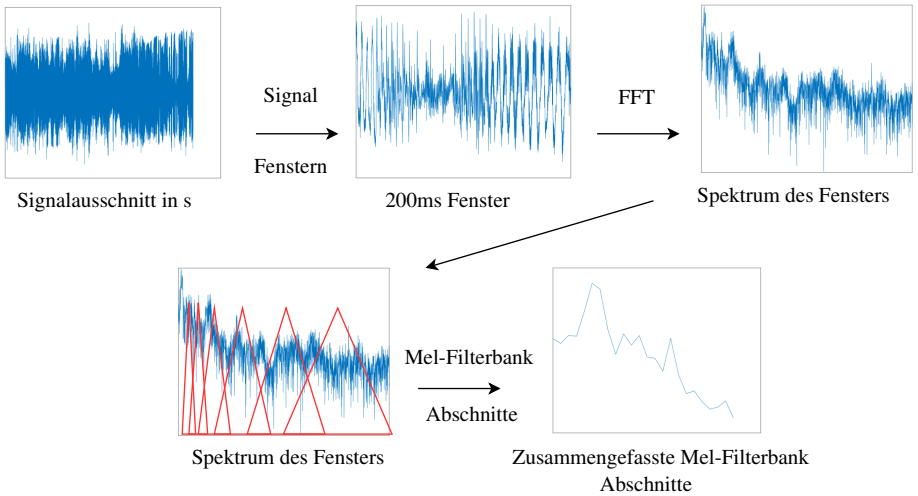


Abbildung 1 – Blockschaltbild zur Funktionsweise der entwickelten Signalverarbeitung.

Im Signal wird jede Spur getrennt betrachtet. Anschließend wird die Spur in 200ms lange Abschnitte unterteilt, welche sich zu 50% überlappen. Nach der Fouriertransformation wird der Pegel in dB_{Fs} über den folgenden Zusammenhang berechnet:

$$Pegel[dB_{Fs}] = 20 \cdot log_{10}\left(\frac{|Wert|}{32768}\right) \tag{1}$$

Anschließend wird die Skalentransformation angewendet um daraus die *critical bands* berechnen zu können:

$$h(f) = 2595 \cdot log_{10}\left(1 + \frac{f}{700Hz}\right) \tag{2}$$

Mit f ist hier die Frequenz in Hz gemeint und mit $h(f)$) die Tonheit in mel. Nach der Unterteilung des Spektrums in, mit zunehmender Frequenz größer werdenden critical bands, wird der Mittelwert über jeden Abschnitt gebildet. Insgesamt wird die Anzahl der critical Bands auf 25 festgelegt [2, S.96]. Der Verlauf der Mittelwerte ist beispielhaft in Abbildung 1 rechts Unten dargestellt Anschließend erfolgt über den Vergleich bisheriger Höhenlinien ein Clustering. Das bedeutet, dass pro 200ms Abschnitt eines Signals die Höhenlinie der zusammengefassten critical bands berechnet und die Summe der critical bands gebildet wird. Die Summe der critical Bands wird dann mit denen vorheriger Fenster verglichen. Bei einer Übereinstimmung zwischen 0,8 bis 0,99 werden die Abschnitte dem gleichen „Lautstärke"-Cluster zugeordnet. Ist die Übereinstimmung geringer, wird ein neuer Cluster erstellt, der in den nachfolgen Abgleichen berücksichtigt wird.
Da in [6] gezeigt wurde, dass Gesprächsanteilen eine besondere Rolle zukommt, werden jene auch in dieser Untersuchung gesondert betrachtet. Für die Erkennung der Sprachanteile wird die interne Voice Activity Detection (VAD) Funktion von MATLAB (inklusive der Optionalen Merkmale Pitch und melSpectrum) genutzt. Zusammenhängende Bereiche, welche von der Funktion als Sprache erkannt worden, werden dann im Post-Processing auf einen konstanten Maximalwert normiert. Dadurch wird in jeder Aufnahme, gesprochener Inhalt immer mit der gleichen Lautstärke wiedergegeben.

Kompression
Unabhängig von der Anpassung der Lautstärke der gesprochenen Inhalte wird eine Stufenweise Kompression durchgeführt. Das bedeutet, dass nach dem Clustering eine Unterscheidung in laute und leise Abschnitte, basierend auf Summe der critical bands, erfolgt. Klassen mit einer höheren Leistung werden stärker komprimiert als Klassen mit geringerer Leistung. Die „leiseste" Klasse wird nicht Komprimiert, während auf die höheren Klassen eine DRC angewendet wird, welche sich in der Auswahl des Schwellwerts (höher für leistungsstärkere Klasse) unterscheidet. Für alle DRC Anwendungen in dieser Arbeit wird der Matlab interne „compressor" mit unterschiedlichen Einstellungen pro Klasse verwendet (Advanced Methode – ADV).
Um einen Vergleich zwischen der entwickelten Signalverarbeitung und einer standardisierten DRC zu erhalten, wird jeder originale Ausschnitt zusätzlich mit einer DRC bearbeitet. Die DRC wird ohne Klassifikation und ohne VAD auf die Audiodatei angewendet (klassische DRC – kDRC).

3 Ergebnisse der Automatischen Vorverarbeitung

In Abbildung 2 ist beispielhaft das Ergebnis der Klassifikation über die Summe der critical bands an der Aufnahme eines Ausschnitts aus dem Sample F5 dargestellt, für einen exemplarischen Abschnitt ist dann in Abbildung 3 das Ergebnis der Lautstärkekompression entsprechend der in Abschnitt 2 beschriebenen Advanced Methode dargestellt.

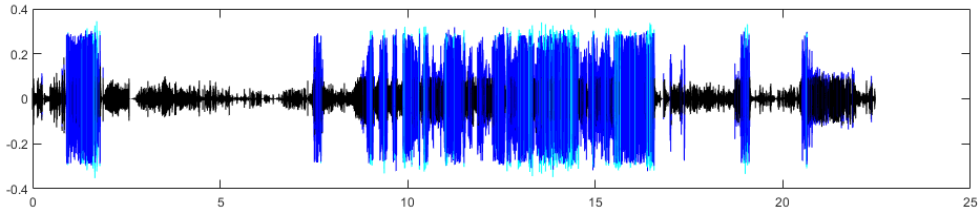
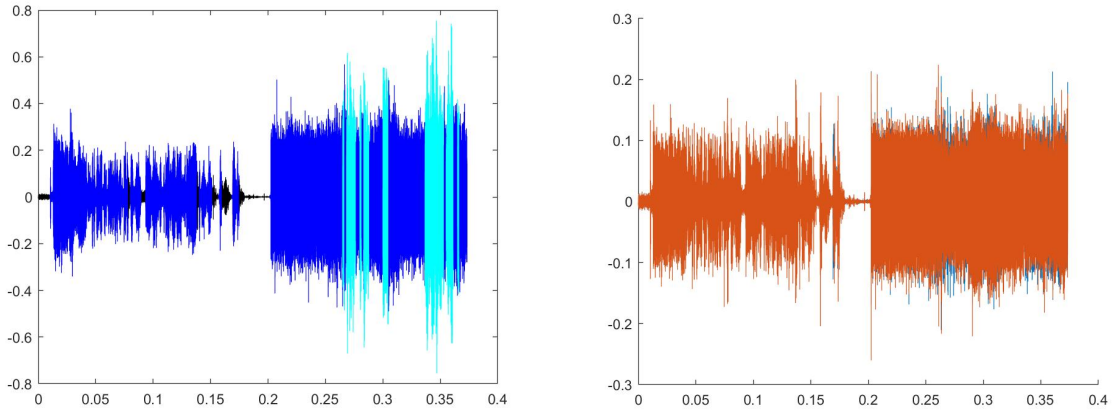


Abbildung 2 – Ausschnitt aus dem Film F5 mit Klassifikation durch Korrelation Summe der critical bands (x-Achse t in min).



(a) Klassifikation mit Summe der critical bands (b) Ausschnitt nach Anwendung der Signalverarbeitung

Abbildung 3 – Klassifikation eines Ausschnitts aus F5 (a) und Ergebnis nach der Anwendung der implementierten Signalverarbeitung (b) (x-Achse t in min)

Als vordefinierte Gewichte für die Klassifizierung dienen die Gewichte aus dem Sample F5. Die Definition und Einteilung in Klassen erfolgt dadurch für alle Ausschnitte für die Hörerbewertung einheitlich. Für die schwarze Klasse wird keine Kompression definiert. Für die blaue Klasse wird ein Threshold mit -25dB$_{Fs}$ und für die hellblaue Klasse ein Threshold mit -30dB$_{Fs}$ gewählt. Für die weiteren Parameter wird auf die Standard-Werte der Matlab Funktion zurückgegriffen. Das Ergebnis der entsprechenden Kompression ist in Abbildung 3 b) dargestellt.

4 Hörerbewertung

Zur Evaluation der entwickelten automatischen Pegelanpassung wurde eine Hörerbewertung durchgeführt. Dafür wurden sechs Ausschnitte (Durchscnittliche Länge 18 Sekunden) aus den Samplen (siehe Tabelle 1) ausgewählt, welche den eingangs beschriebenen Effekt starker Pegelschwankungen beinhalten. Das bedeutet, dass plötzliche Explosionen, Motorengeräusche oder ähnliches auftritt, welche als störend empfunden werden könnten. Außerdem sind in jedem Ausschnitt auch kurze Teile eines Dialogs zu hören, um einen Referenzpunkt zu geben. Ziel der Bewertung soll es sein, mit der Auswertung die folgenden Forschungsfragen beantworten zu können:

1. Wird der unveränderte Ausschnitt, oder der durch Pegelanpassung veränderte Abschnitt als angenehmer wahrgenommen?
2. Falls dies bejaht wird: Wird eine komplexere Methode mit Kompression tendenziell besser bewertet, als eine standardisierte kDRC?

Durchführung der Studie

Die Hörerbewertung wurde mittels SoSci Survey [8], aufgrund der COVID-19 Einschränkungen, realisiert. Nach der Einführung in die Aufgabe, wird ein Auschnitt eines Beispielvideos

(BB2) gezeigt, welches dazu dient, dass die Probanden eine für sie angenehme Lautstärke einstellen können. Die anschließende Bewertung erfolgt durch einen paarweisen Kreuzvergleich, Video mit originaler Tonspur vs. eines der beiden Kompressionsverfahren (ADV und kDRC). Dafür wird jeder Teilnehmer zufällig einer von zwei Gruppen zugeordnet, die dann entweder mit der ADV oder der kDRC Methode starten und dann immer im Wechsel die jeweils andere Version gezeigt bekommen. Insgesamt werden sechs Videopaare bewertet. Die Bewerter sollen hierbei zwei Fragen mittels 6-Item Likert Skala beantworten. Die gestellten Fragen sind:

1. Welches Video war angenehmer (in der Gesamtlautstärke)?
2. Welches Video war verständlicher (im Sprachanteil)?

Zusätzlich werden noch Alter, Geschlecht und die Hardware für die Soundausgabe abgefragt.

5 Ergebnisse der Hörerbewertung (Studie)

Insgesamt haben 61 Teilnehmer an der Hörerbewertung teilgenommen. Das Durchschnittsalter liegt bei 43 Jahren. Die Hälfte der Teilnehmer sind in der Altersgruppe 20-27 Jahre und ein Drittel im Alter von 60-80 Jahren. Das Geschlecht ist nicht ganz gleich-verteilt mit 25 weiblichen und 36 männlichen Teilnehmern. Als Abspiel-Equipment wurden externe Boxen (12 Teilnehmer), Kopfhörer (36 Teilnehmer) und PC interne Lautsprecher (13) verwendet.
Zur ersten Fragestellung (unveränderter Ausschnitt, oder Kompression angenehmer), wird die Gesamtheit der Antworten zur Frage 1 ausgewertet. Dazu ist in Abbildung 4 der Mittelwert über alle Samples aufsummiert über alle Teilnehmer dargestellt.

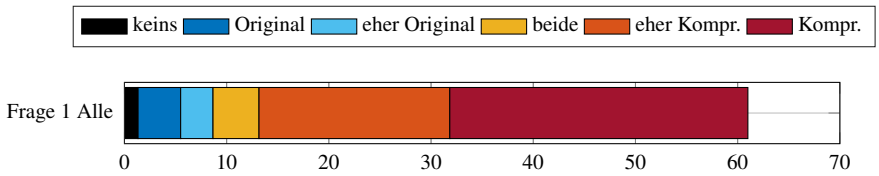


Abbildung 4 – Antworten von 61 Teilnehmern auf die Frage: Welches Video war insgesamt angenehmer in der Lautstärke, gemittelt über alle 6 Samples und je Teilnehmer summiert.

Es ist erkennbar, dass die Kompression bei der Mehrheit der Teilnehmer als „insgesamt angenehmer“ wahrgenommen wird. Das bedeutet, dass Hörer eine komprimierte Variante gegenüber der Tonspur originaler Ausschnitte bevorzugen.
Um die Zweite Frage zu beantworten wird zwischen den beiden Kompressionsmethoden (ADV und kDRC) unterschieden.

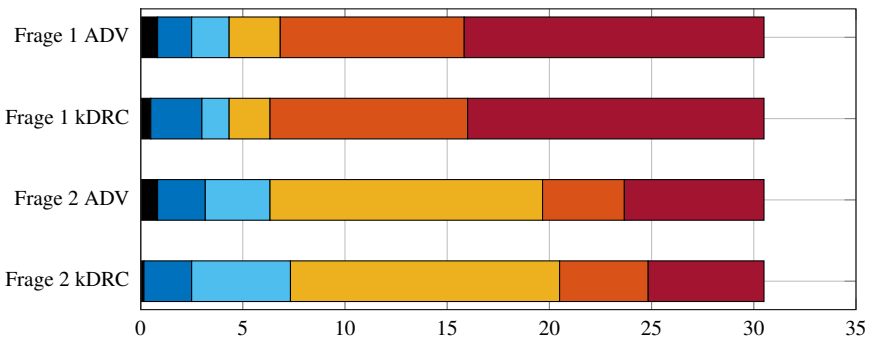


Abbildung 5 – Antworten auf Frage 1 und 2, gemittelt über die jeweiligen 3 Samples je Teilnehmer und anschließend summiert über alle Teilnehmer. Farbkodierung wie in Abbildung 2.

In Abbildung 5 ist kein eindeutiger Trend zu erkennen, dass die entwickelte ADV Methode deutlich besser oder schlechter in der Gesamtlautstärke abgeschnitten hat als die kDRC.

Zur Auswertung der Frage 2 (Welches Video war verständlicher im Sprachanteil?) schneidet die ADV Methode leicht besser ab, als die kDRC. Generell haben in jeder Gruppe ca. 13/30 Teilnehmern für „beide“ gestimmt. Das bedeutet, dass sowohl Kompression, als auch das Original als verständlich im Sprachanteil wahrgenommen wurden. Betrachtet man jedoch die Gruppe der über 65-jährigen, so fällt auf, dass von diesen die Mehrheit der Teilnehmer ADV als verständlicher als das Original bewerten und ein Drittel zwischen kDRC und Original entweder das Original oder keins von Beiden als verständlicher bewerten. Aus der Literatur ist bekannt, dass mit zunehmenden Alter die Verständlichkeit von Dialogen abnimmt [9, 10].

6 Diskussion

In diesem Beitrag wurde eine Methode vorgestellt, die die unterschiedlichen Pegel (unter der Berücksichtigung frequenzabhängiger Wahrnehmung von Lautstärke) innerhalb einer Audioaufnahme angleichen soll und dabei insbesondere zu laute Anteile abmildern soll (mit je nach Cluster unterschiedlichen Schwellwerten), gleichzeitig aber Sprachanteile auf einem gemeinsamem Level belassen soll. Dazu wurde eine entsprechende Signalverarbeitungskette entwickelt und mit unter gleichen Bedingungen aufgenommenen Samples getestet. Anschließend wurde der Effekt in einer Hörerbewertung überprüft. Hierzu wurden Orginalsequenzen und angepasste Sequenzen in einem Paarvergleich von Testhörern evaluiert. Weiterhin wurde auch eine klassische Dynamic range compression (kDRC) evaluiert, um zu überprüfen, ob diese nicht schon ausreichende Effekte zeigt.
In Bezug auf die Signalvorverarbeitung konnte festgestellt werden, dass die Methode „Summe der critical bands“ zum Clustering von Audiosignalen aufgrund ihrer Leistung geeignet ist. Es ist damit eine Möglichkeit geschaffen basierend auf Erkenntnissen der Psychoakustik eine automatisierte Klassifikation in Lautstärkeklassen durchzuführen. Es bleibt Bestandteil zukünftiger Forschung andere Klassifikatoren zu finden, die zusätzliche Inhalte wie Musik von Effekten unterscheiden können.
Durch die Hörerbewertung (besonders mit Auswertung von Frage 1) wird bestätigt, dass originale Ausschnitte als „zu laut“ wahrgenommen werden und im Vergleich dazu eine komprimierte Version des selben Ausschnittes als angenehmer in der Gesamtlautstärke wahrgenommen wird. Die Bewertung ergibt jedoch keinen eindeutigen Vorteil der entwickelten ADV Methode gegenüber einer kDRC bezüglich der Gesamtlautstärke. Es kann ein leichter Vorteil der ADV Methode in der Bewertung nach der Verständlichkeit der Sprachanteile festgestellt werden. Eine mögliche Erklärung kann darin liegen, dass aufgrund der Online-Studie keine kontrollierten Bedingungen hergestellt werden konnten und insbesondere nicht die Eignung/Qualität des Abspiel-Equipments beeinflusst werden konnte. Weiterhin liegt der Altersdurchschnitt der Teilnehmer bei 43 Jahren, mit einen großen Anteil von Jüngeren Probanden (20-27 Jahre) und nur wenigen Teilnehmern über 65 (14 Teilnehmer). Es ist jedoch aus der Literatur bekannt, dass die Verständlichkeit von Dialogen besonders mit zunehmendem Alter (65+) [9, 10] schwieriger wird. Eine gesonderte Analyse dieser Teilnehmergruppe zeigt eine stärkere Präferenz der hier entwickelten ADV Methode im Hinblick auf die Dialogverständlichkeit (Frage 2). Über die Hälfte der >65 jährigen empfindet die ADV als verständlicher im Vergleich zum Original. Während ein drittel beim Vergleich zwischen kDRC und Original entweder das Original oder keins von Beiden als verständlicher bewerten.
Auch wenn kein deutlicher Vorteil der ADV Methode in der Bewertung der Gesamtlautstärke festgestellt werden konnte, ist die Entwicklung der Methode als erfolgreich zu bewerten, da diese im Gegensatz zur kDRC eine zusätzliche Personalisierung, durch Einbeziehung einer VAD und der individuellen Parametrisierung erlaubt. Dadurch ist die hier vorgeschlagene Methode besonders für ältere Personen (mit Hörschwäche) geeignet. Es ist bisher keine lauffähige

Lösung bekannt, die VAD in Echtzeit auf einem low-Budget Prototypen ermöglicht. Diese Problematik bleibt Bestandteil zukünftiger Forschung. Die Implementierung eines standardisierten kDRC ist bereits mit dem aktuellen Stand in Echtzeit möglich. Die Implementierung der hier vorgestellten ADV Methode auf ein Low-Budget System (ESP32-LyraT V4.3.) ist als einer der nächsten Schritte geplant. Eine erste Implementierungen mittels einer Buffer-basierten automatischen Lautstärkeregelung ist unter folgendem Link abrufbar[1].

Literatur

[1] NEWELL, P., K. HOLLAND, J. NEWELL, und B. NESKOV: *Cinema sound and the loudness issue: its origins and implications*. In *Reproduced Sound 2016: Sound with Pictures: Time is of the Essence*, Bd. 38, S. 104–120. Institute of Acoustics, 2016. URL `https://eprints.soton.ac.uk/408248/`.

[2] FRIESECKE, A.: *Die Audio-Enzyklopädie*. DE GRUYTER, 2014. doi:10.1515/9783110340181.

[3] GOCKEL, H. E., R. FAROOQ, L. MUHAMMED, C. J. PLACK, und R. P. CARLYON: *Differences between psychoacoustic and frequency following response measures of distortion tone level and masking*. *The Journal of the Acoustical Society of America*, 132(4), S. 2524–2535, 2012. doi:10.1121/1.4751541.

[4] HEEREN, W., J. APPELL, und J. VERHEY: *Relation between loudness in categorical units and loudness in phons and sones*. *The Journal of the Acoustical Society of America 133*, 133(4), 2013. doi:https://doi.org/10.1121/1.4795217.

[5] PFISTER, B.: *Sprachverarbeitung Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. Springer Vieweg, Berlin, 2017.

[6] RIEDMILLER, J. C., S. LYMAN, und C. ROBINSON: *Intelligent program loudness measurement and control: What satisfies listeners?* 2003.

[7] GIANNOULIS, D., M. MASSBERG, und J. D. REISS: *Digital dynamic range compressor design—a tutorial and analysis*. *J. Audio Eng. Soc*, 60(6), S. 399–408, 2012. URL `http://www.aes.org/e-lib/browse.cfm?elib=16354`.

[8] LEINER, D. J.: *Sosci survey (version 3.1.06)*. 2019. Available at https://www.soscisurvey.de.

[9] TAMBS, K.: *Moderate effects of hearing loss on mental health and subjective well-being: Results from the nord-trøndelag hearing loss study*. *Psychosomatic Medicine*, 66(5), S. 776–782, 2004. doi:10.1097/01.psy.0000133328.03596.fb.

[10] FOZARD, J. L. und S. GORDON-SALANT: *Changes in vision and hearing with aging*. In J. E. BIRREN und K. W. SCHAI (Hrsg.), *Handbook of the psychology of aging*, S. 241—-266. Academic Press, San Diego, USA, 2001.

[1] https://github.com/geoschmi/Esp32-LyraT-V4.3-passthru-example-with-ALC-combination