Zürcher Hochschule
für Angewandte Wissenschaften

School of
Engineering

InIT Institut für angewandte
Informationstechnologie

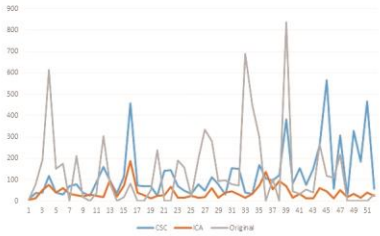
Shazam für Hintergrundmusik: Fingerprinting in Broadcast Audio

Diplomierende
Benjamin Hohl
José Rota

Dozierende
Sigisbert Wyrsch
Thilo Stadelmann

Das Schweizer Fernsehen produziert diverse Fernsehsendungen, in denen Hintergrundmusik zum Einsatz kommt. Das Ziel dieser Bachelorarbeit ist es, diese Musik automatisch identifizieren zu können, damit die Musik bei der SUISA lizenziert werden kann.
Zwar gibt es zur Identifizierung von Musik Programme wie beispielsweise Shazam, das Musik, die in der Umgebung läuft, über das im Smartphone verbaute Mikrophon aufnehmen und als Ergebnis den Interpret und den Titel anzeigen. Allerdings wird dazu ein Umfeld mit wenig bis keinen Störquellen benötigt. In Fernsehsendungen gibt es aber laufend Störquellen in Form von Vordergrundgeräuschen: etwa die Moderation einer Quizshow oder beispielsweise ein vorbeifahrendes Auto in einer Reportage. Daher sind Verfahren nötig, um diese Störquellen soweit zu reduzieren, dass eine Identifizierung der Musik möglich wird.
Im Rahmen dieser Bachelorarbeit werden diese Verfahren zunächst besprochen und evaluiert. Die am besten geeigneten werden danach implementiert. Um die Störungen zu reduzieren, führen in der Theorie je nach Art der Störung unterschiedliche Algorithmen zu einem besseren Ergebnis. Spricht beispielsweise ein Moderator, wäre es sinnvoll, den Centered Speech Cancellation Algorithmus einzusetzen. Bei anderen Störquellen kann das Independent Component Analysis Verfahren zu einem besseren Ergebnis führen. Die beiden Algorithmen werden durch Experimente verglichen. Die danach übrig gebliebene Hintergrundmusik wird anschliessend mit einer Sammlung von Musikstücken verglichen, um sie zu identifizieren. Dazu wird das Framework dejavu verwendet.
Der einheitliche Aufbau unserer Experimente ermöglicht es, die Bereinigungs-Algorithmen zu vergleichen. Dazu werden die Fernsehsendungen zuerst in gleich lange Sequenzen aufgeteilt und anschliessend die Experimente durchgeführt. Das Resultat: Bei fünf Sekunden langen Sequenzen, die noch nicht bereinigt wurden, wird 64% der Musik richtig identifiziert, bei Bereinigung mithilfe von Independent Component Analysis sind es 74%. Wenn Centered Speech Cancellation eingesetzt wird, sind es sogar 81%.
Aufgrund dieser Ergebnisse sollte Centered Speech Cancellation als Bereinigungs-Algorithmus eingesetzt werden, bevor die Musikidentifizierung durchgeführt wird.

Anzahl der korrekt erkannten 5-Sekunden-Sequenzen.



Gegenüberstellung der Confidence-Werte zwischen Original (unüberarbeitetes Ausgangsmaterial) und nach Anwendung von ICA oder Centered-Speech-Cancellation.