

# Integration einer Sprachsteuerungsfunktion in Mobile Apps

Themenbereiche:	Sprachsteuerung, Audio, Machine Learning, Triggerwort-Erkennung, Mobile Apps
Studierender:	Ruben Nuñez
Dozent:	Dr. Florian Herzog
Experte:	Damien Piguet
Auftraggeber:	Stefan Reinhard
Keywords:	Sprachsteuerung, Mobile Apps, Machine Learning, Triggerwort-Erkennung, Audioverarbeitung, Datensatzerstellung, Modelltraining, App-Integration

## 1 Aufgabenstellung

Die Bachelorarbeit zielt darauf ab, eine Sprachsteuerungsfunktion für mobile Apps zu entwickeln, wobei der Fokus auf der Erkennung von Triggerwörtern in akustischer Sprache liegt. Zentrale Aspekte sind die Entwicklung eines Machine Learning-Modells, seine Implementierung in eine mobile Plattform wie iOS und die Beachtung von Datenschutz und ethischen Richtlinien beim Erstellen und Verwenden eines Datensatzes.

## 2 Ergebnisse

Im Rahmen Bachelorarbeit wurden folgende Ergebnisse erzielt:

- Ein **Machine Learning-Modell**, das in der Lage ist, Triggerwörter in akustischer Sprache zu erkennen.
- Eine **Integration in eine iOS-App**, die das Modell verwendet, um Triggerwörter in Echtzeit zu erkennen.
- Ein **Datensatz**, welcher ethische Richtlinien und Datenschutzbestimmungen einhält welcher für das Training des Modells verwendet wurde.
- Eine **Dokumentation**, die die Entwicklung des Modells, die Erstellung des Datensatzes und die Implementierung der App beschreibt.

Beim Machine Learning Modell handelt es sich um eine Kombination aus einem Convolutional Neural Network (CNN) und einem LSTM-Netzwerk (Long Short-Term Memory). Das CNN wird verwendet, um die akustischen Merkmale der Sprache zu extrahieren, während das LSTM-Netzwerk die zeitlichen Abhängigkeiten zwischen den akustischen Merkmalen erfasst. Das Modell wurde in PyTorch implementiert und mit dem erstellten Datensatz. Das Modell wurde mit dem erstellten Datensatz trainiert und erreichte eine Accuracy von 0.88 und eine F1-Score von 0.80 auf dem Testdatensatz. Das Modell hat den Namen **WakeupTriggerConvLSTM2s** und ist in der Lage, Triggerwörter in Echtzeit zu erkennen. Eine Transformationskomponente mit dem Namen **AudioToSpectrogramTransformJit** zur Umwandlung von Audio in Spektrogramme wurde ebenfalls in PyTorch entwickelt und dient als Feature-Extraktions-Modul.

Ein kritischer Aspekt für die Lösungsakzeptanz war die Gewährleistung der Effizienz des Modells bezogen auf die Real-Time-Fähigkeit. Das Modell erzielt eine Vorhersagegeschwindigkeit von 0.01 Sekunden pro Spektrogramm und erfüllt damit die Anforderungen an eine Echtzeiterkennung, ohne die Geräteleistung zu beeinträchtigen. Die nachfolgende Abbildung zeigt schematisch den Aufbau der Verarbeitungskette in der App.

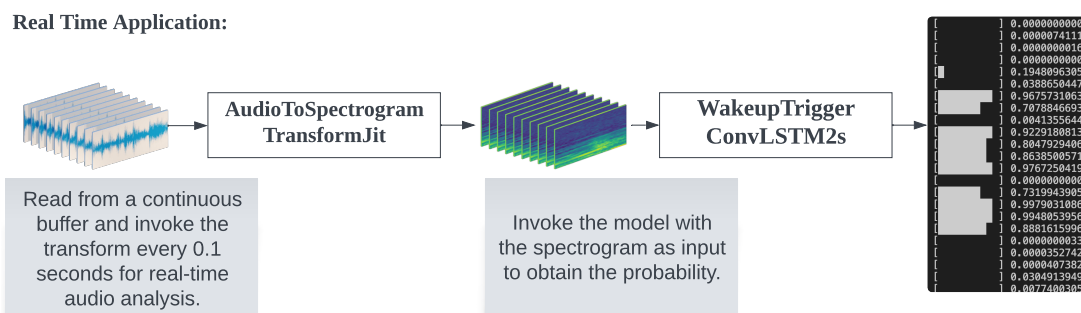


Abbildung 1: Real Time Anwendung

Die realisierte Applikation wurde als iOS App implementiert und verwendet die entwickelten PyTorch Module *AudioToSpectrogramTransformJit* und *WakeupTriggerConvLSTM2s*.

### **3 Lösungskonzept**

### **4 Spezielle Herausforderungen**

### **5 Ausblick**