

Beiträge zur Binauralen Sprecherlokalisierung und -trennung für dynamische akustische Szenen

Mehdi Zohourian

Kurzfassung

Adaptives Beamforming ist eine mehrkanalige Signalverarbeitungstechnik, die häufig in Sprachkommunikationstechnologien wie Freisprech-Mobiltelefonen, Mensch-Maschine-Schnittstellen oder Hörgeräten eingesetzt wird.

In jüngster Zeit sind moderne Hörgeräten mit einer drahtlosen binauralen Verbindung ausgestattet, die die Implementierung binauraler Beamformer ermöglicht und somit eine Überlegenheit gegenüber monauralen Techniken bietet. Normalerweise ist die Einfallsrichtung (engl. *direction of arrival*, DOA) von Schallquellen ein Hauptparameter für die Berechnung eines Beamformers. In Algorithmen für Hörgeräte wird jedoch allgemein davon ausgegangen, dass sich die Zielquelle vor dem Benutzer befindet. Eine solche Annahme kann in realen Szenarien verletzt sein, z. B. wenn ein Hörgeräteträger ein Auto steuert. Darüber hinaus führt das Vorhandensein mehrerer simultaner Sprecher und die Bewegung des Kopfes des Hörgeräteträgers zu erheblichen Verzerrungen.

Daher befasst sich diese Arbeit mit Ansätzen zur Lokalisierung und Trennung von simultanen Sprechern unter Verwendung von vier Mikrofonen, die auf einem binauralen Hörgeräten-Paar verteilt sind. Zuerst schlagen wir neuartige binaurale Lokalisierungsalgorithmen unter Verwendung geschlossener Kostenfunktionen vor, welche auf Beamforming und auf statistischer Modellierung basieren, und vergleichen diese. Bei den auf Beamforming basierenden Ansätzen lenkt der Beamformer den Ziel- bzw. Nullstrahl auf alle möglichen Quellenpositionen und wählt die Kandidaten aus, die die Empfangsleistung maximieren bzw. minimieren. Modellbasierte Ansätze hingegen setzen eine statistische Repräsentation der zugrunde liegenden Signale voraus und verwenden eine Maximum-Likelihood (ML) Optimierung. Diese Ansätze wurden für Freifeld-Mikrofonanordnungen umfassend untersucht, werden in dieser Arbeit jedoch auf die binaurale Konfiguration angewendet, wobei der Abschattungseffekt des Kopfes berücksichtigt wird. Die vorgeschlagenen Lokalisierungsalgorithmen sind in einem *generalised side-lobe canceller* (GSC) Verfahren integriert, um den gewünschten Sprecher in Gegenwart von anderen Sprechern und Hintergrundgeräuschen sowie bei Kopfbewegung des Hörgeräteträgers zu extrahieren. Die GSC-Komponenten werden mithilfe der Schmalband-Zielanwesenheitswahrscheinlichkeiten (engl. *target presence probability*, TPP) und der Breitband-DOA-Schätzungen angepasst. Wir bewerten die Leistungsfähigkeit der Lokalisierungsalgorithmen einzeln und auch im Kontext des adaptiven binauralen Beamformings in verschiedenen akustischen Szenarien.

Als Nächstes stellen wir basierend auf dem ML-Ansatz einen neuen adaptiven binauralen Beamformer vor, der TPP in Verbindung mit der Wahrscheinlichkeit der Mehrkanal-Sprachanwesenheit schätzt. Der vorgeschlagene TPP-Schätzansatz wird mit dem GSC-Algorithmus kombiniert und erweitert, um binaurale Signale zu erzeugen, die den räumlichen Höreindruck für den Hörgeräteträger erhalten. Der vorgeschlagene Algorithmus erzielt eine überlegene Quellentrennungsleistung in gestörten und dynamischen Situationen.

Zum Schluss beschäftigen wir uns mit dem Problem der Schätzung des Abstandes zum Sprecher mit binauralen Hörgerätemikrofonen. Ein verbreiteter Parameter zum Schätzen der Entfernung einer Schallquelle ist das Verhältnis von Direktschall- zu Nachhallenergie (engl. *direct-to-reverberant energy ratio*, DRR). Um die Abstand vom Sprecher zu schätzen, nehmen wir Volumen und Nachhallzeit des Raums als bekannt an. Wir schlagen vor, DRR mithilfe eines Ansatzes basierend auf einem Modell der interauraler Kohärenz zu schätzen. Die Ergebnisse mehrerer Experimente bestätigen die Wirksamkeit des vorgeschlagenen Entfernungsschätzungsalgorithmus.