DAGA 2022 Stuttgart – 47. Jahrestagung für Akustik, 21. – 24. März 2022

Messung und Modellierung der Höranstrengung von Smart-Speaker-Sprachausgaben unter realistischen akustischen Bedingungen

Rainer Huber[1], Andreas Volgenandt[1], Jan Reimes[2], Magnus Schäfer[2], Jan Rennies[1]

[1] Fraunhofer-Institut für Digitale Medientechnologie (IDMT) Oldenburg

[2] HEAD acoustics GmbH

Abstract

Die Sprachinteraktion mit Smart Speakern nimmt im Alltag eine immer größere Rolle ein. Damit steigt auch der Bedarf an objektiven Messverfahren zur Bewertung der Sprachausgabe der (künstlichen) Stimmen unter realistischen Abhörbedingungen. Für dieses Anwendungsfeld sind typischerweise referenzbasierte Metriken nicht nutzbar. Dieser Beitrag gibt daher einen Einblick in die Entwicklung von referenzlosen instrumentellen Evaluationsmethoden für die sprachliche Interaktion mit Smart Speakern. Eine zentrale Metrik ist ein Modell zur Vorhersage der empfundenen Höranstrengung basierend auf Technologien der automatischen Spracherkennung, das zuvor erfolgreich für natürliche Stimmen in Broadcast-Anwendungen evaluiert wurde. Für das neue Anwendungsfeld Smart Speaker wurden Messungen der empfundenen Höranstrengung (und Sprachqualität) von natürlicher und synthetisierter Sprache unter realistischen akustischen Bedingungen vorgenommen (verschiedene Störgeräusche, Signal-zu-Rausch-Verhältnisse, Abstände zum Smart Speaker sowie Halligkeiten simulierter Räume). Die per Kunstkopf aufgenommenen Testsignale wurden von normalhörenden Proband*innen hinsichtlich bewertet. Neben dem genannten Höranstrengungsmodell wurden die Bewertungen zusätzlich mit dem (ebenfalls referenzlosen) NISQA-Modell verglichen, das ursprünglich zur Bewertung der Natürlichkeit synthetisierter Sprache entwickelt wurde. Es zeigt sich, dass beide Modelle gemäß ihrer ursprünglichen Intention teilweise komplementäre Informationen liefern und somit für den Einsatz bei der Bewertung von Smart Speakern vielversprechende Kandidaten sind.