

Abstract

Sleep at the wheel becomes a severe problem for safe driving, each year around 20% of the traffic accidents are drowsy related. Rater-based subjective drowsiness detection is a reliable approach that can be easily applied to in-vehicle systems. For this purpose, MIR scale for rater-based drowsiness detection was developed. However, since it depends on indicators that are mainly derived from human facial expressions, cultural influence need to be considered. In this paper, the reliability of MIR scale across different cultures was validated. Both German and Chinese raters were invited and trained to assess drowsiness levels from various types of drivers. By comparing their rating performance to validate if raters from different cultural backgrounds could have a consistent and stable rating performances on MIR scale. The results indicated that German and Chinese rating results were far from agreed. In general, Chinese raters had a relatively poor rating performance, they recognized less indicators, especially those eye-lid closure indicators, and had a lower level of inter-rater reliability. Also, significant impact regarding drivers' genders and cultural backgrounds can be noticed.

Keywords: MIR scale, drowsiness detection in vehicle, rater-based subjective method, cultural influence

Kurzfassung

Schläfriges Fahren ist ein ernstes Problem für sicheres Fahren, jedes Jahr werden 20% der Verkehrsunfälle durch schläfrig verursacht. Beobachterbasierte Müdigkeits-Detektion ist ein zuverlässiger Ansatz, die leicht im Fahrzeug Systeme angewandt werden können. Zu diesem Zweck, wurde MIR Skala für Beobachterbasierte Müdigkeits-detektion entwickelt. Denn es hängt von Indikatoren ab, die auf die Mimik basieren, dabei müssen kulturellen Einflüsse berücksichtigt werden. In diesem Papier wurde die Zuverlässigkeit der MIR-Skala zwischen verschiedenen Kulturen validiert. Deutsche und chinesische Versuchspersonen wurden eingeladen und trainiert, um die Schläfrigkeit von verschiedenen Arten von Fahrern zu bewerten. Durch den Vergleich ihrer Leistung, wurde überprüft, ob die Testpersonen aus verschiedenen kulturellen Hintergründen eine konsistente und stabile Bewertung in den Aufführungen auf der Mir-Skala haben. Die Ergebnisse zeigen, dass deutsche und chinesische Ratingergebnisse weit davon entfernt waren, zu vereinbaren. Im Allgemeinen, dass die Chinesischen Versuchspersonen eine relativ schlechte Leistung in der Bewertung hatten, sie erkannten weniger Indikatoren, vor allem der Lidschlüsse, und hatten ein niedrigeres Niveau der Inter-rater Zuverlässigkeit. Es gibt keine signifikanten Auswirkungen von Fahrer Geschlecht und kulturelle Hintergründe sind bemerkt.

Keywords: MIR Skala, Müdigkeits detektion im Fahrzeug, Beobachterbasierte Müdigkeits-Detektion, Kultureller Einfluss