

A Methodological Approach to Evaluate Driver Distraction During In-vehicle Interactions with Voice-based Large Language Models

Nadja Schömig¹, Christina Kremer¹, Dennis Befelein¹, Thomas Brand¹, Denise Sogemeier², Frederik Naujoks², Alexandra Neukum¹ & Andreas Keinath²

Motivation

- Trend: Rapid integration of Large Language Models (LLMs) into automotive environments
 - Potential for enhancing information access and user experience
 - Voice-based interaction decreases visual distraction ([1])
 - Risk: What about Cognitive Distraction?
- Existing protocols and acceptance criteria for assessing visual-manual distractions need adaptations (e.g.[2], [3])

Study design

- Static driving simulator at WIVW
- N = 24 participants
- Continuous car follow task
- Tactile DRT
- 3 minutes phases of interaction with LLM
- Reference task: 3-minutes hands-free phone call with the experimenter



LLM use cases

- Instructions: “You are planning a holiday trip in February - ask Jeannie for tips”/“You want to live healthier- ask Jeannie for support”
- Driver is requested to maintain the conversation for a duration of 3 minutes
- Variation of the following factors:

		Feedback modality	
		Verbal only	Verbal + visual
Feedback length	short (max 120-140 char)	UC1	UC3
	long (max. 400-420 char)	UC2	UC4



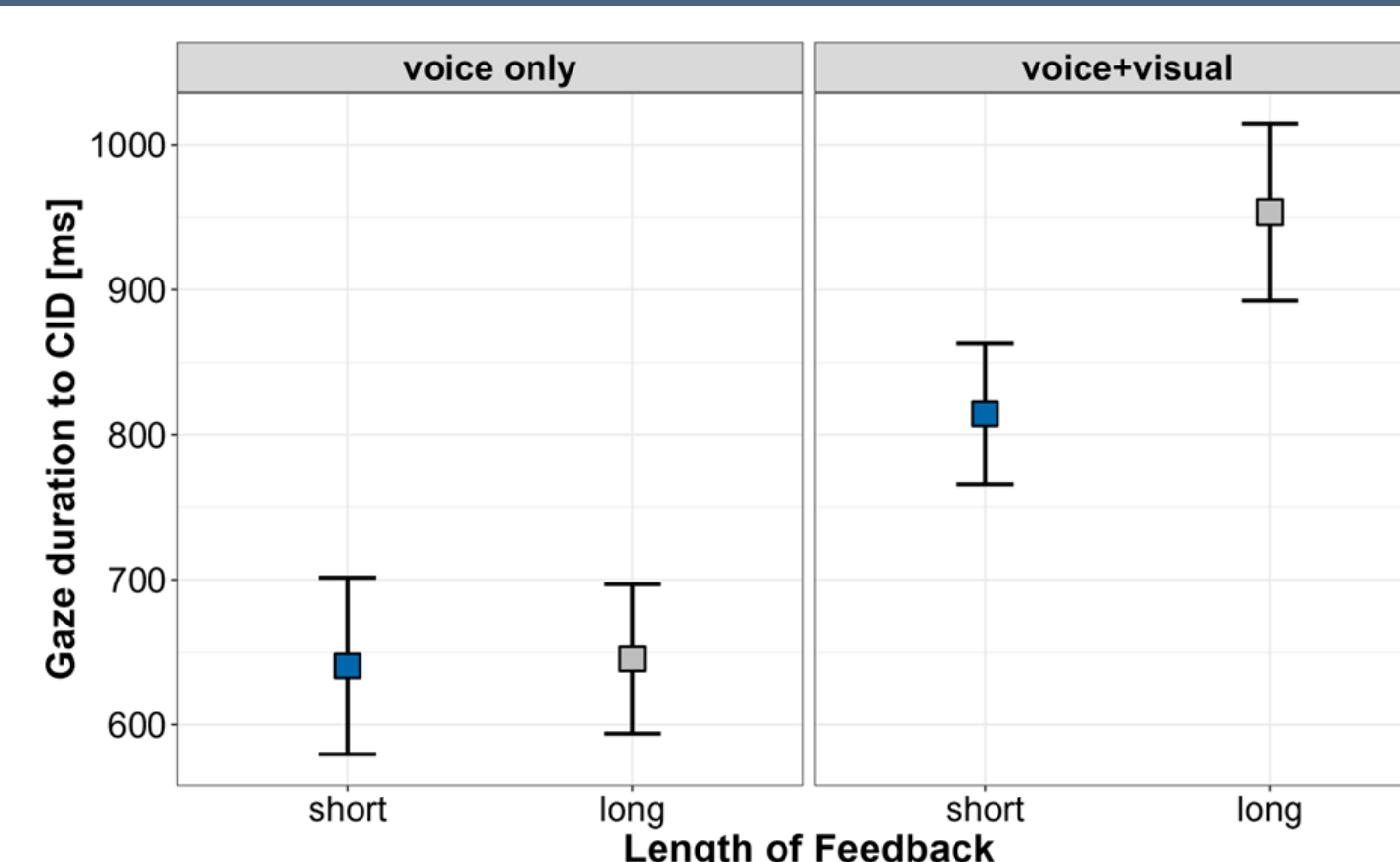
Results

Fulfilment of acceptance criteria

		Use Cases			
		Speech only; short	Speech only; long	Speech-visual; short	Speech-visual; long
Glance criteria	Max. 15% glances ≥ 2s	✓	✓	✓	✓
	Mean glance duration ≤ 2s	✓	✓	✓	✓
Driving performance criteria	TEORT ≤ 12s	✗	✓	✗	✗
	SDLP ≤ Ref	✓	✓	✓	✓
Driving performance criteria	SD_Distance ≤ Ref	✓	✓	✓	✓
	DRT RTs Ref	✓	✓	✓	✓
Driving performance criteria	DRT miss Rate ≤ Ref	✓	✓	✓	✓
		✓	✓	✓	✓

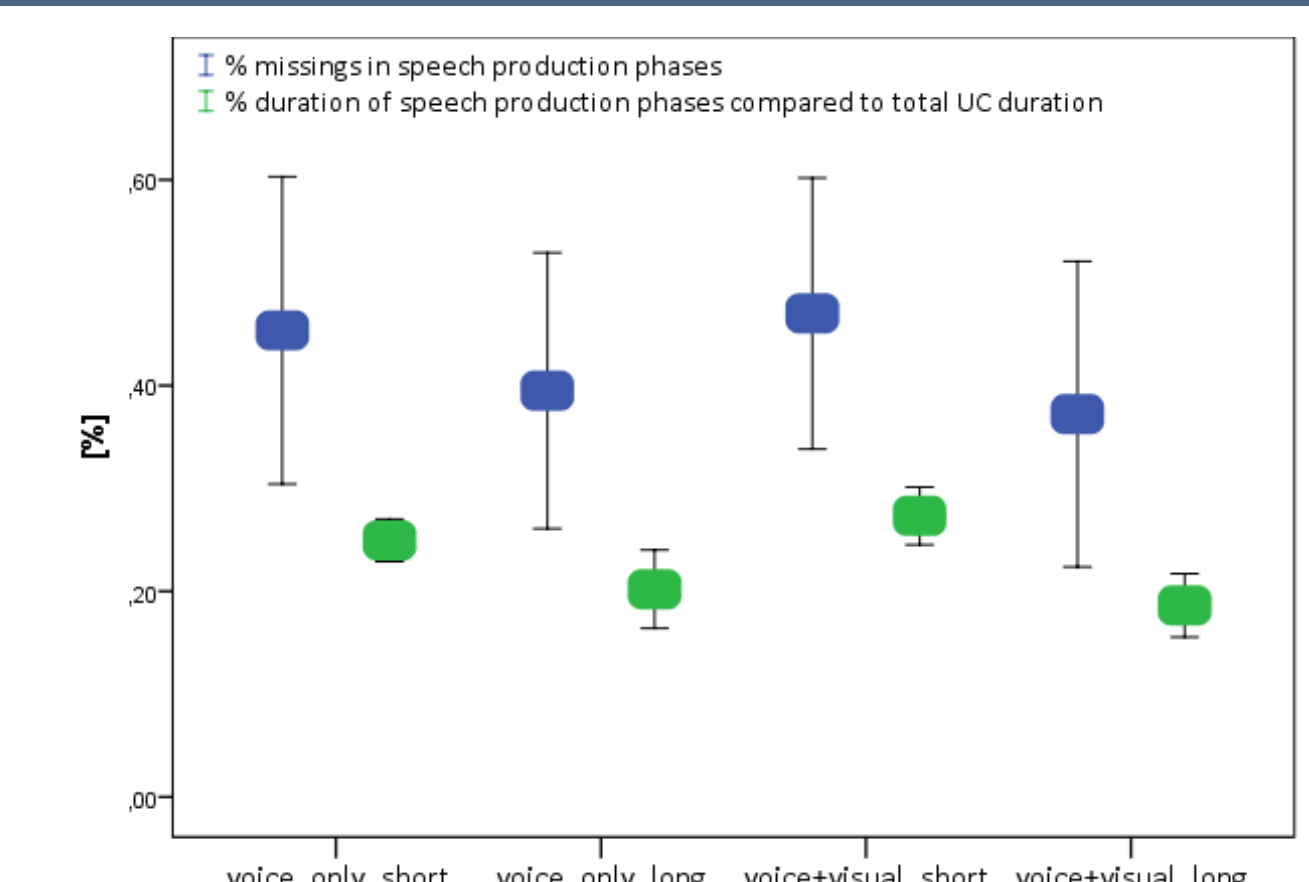
- Glance criteria regarding absolute single glance durations (85% of study sample) fulfilled
- Total-Eyes-Off-Road (TEORT) criterion not fulfilled for 3 of 4 use cases!
- Driving and DRT performance criteria (not worse than reference task) fulfilled

Feedback modality and duration



- Glance duration and number higher for voice+visual compared to voice only
- Repeated glances to CID in voice only UC are due to visual presentation of spoken utterances
- Longer DRT RT for shorter feedback → higher interaction frequency with the system

Distribution of DRT misses



- Speech production phase lasts about 20-25% of total UC duration
- But: This phase accounted for 40-45% of all misses in DRT
- Indicator that speech production during LLM interaction is particularly cognitively demanding

Conclusion

- All examined LLM UCs met the defined distraction evaluation criteria except TEORT and therefore appear to be safe in automotive contexts → TEORT criterion is unsuitable for long LLM interactions
 - Displaying text in addition to voice output increases glance duration and number of glances, but not to a critical extent
 - Different feedback lengths showed similarly low visual distraction effects indicating that drivers did not actually read the text
 - Drivers also glanced at the CID during purely verbal output to verify spoken utterances visually
 - Shorter LLM responses tend to increase distraction —likely due to higher interaction frequency within the 3 minutes interaction phase, leading to increased speech production and visual control needs
- The presented methodological approach reasonably evaluates distraction effects of in-vehicle LLM interactions by assessing both visual and cognitive distraction effects

References

- [1] Monk, C., Sall, R., Lester, B. D., & Higgins, J. S. (2023). Visual and cognitive demands of manual and voice-based driving mode implementations on smartphones. *Accident Analysis & Prevention*, 187, 107033. <https://doi.org/10.1016/j.aap.2023.107033>
- [2] National Highway Traffic Safety Administration (2013). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA).
- [3] Alliance of Automobile Manufacturers (2006). Statement of Principles, Criteria, and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems. Washington, D.C.