# Speech intelligibility in cars: the effect of speaking style, noise and listener age

National University SOKENDAI The Graduate University for Advanced Studies

NII Inter-University Research Institute Corporation / Research Organization of Information and Systems National Institute of Informatics

*Cassia Valentini Botinhao[1], Junichi Yamagishi[1,2,3]*

*[1]The Centre for Speech Technology Research, University of Edinburgh, UK*
*[2]National Institute of Informatics, Japan, [3]SOKENDAI University, Japan*

THE UNIVERSITY of EDINBURGH

## 1. Introduction

- Speech interfaces in cars can increase road safety, particularly for older drivers.
- Older adults, however, have problems decoding speech in noise, particularly in fluctuating noise, even when no other task is engaged and no HI.

In this work we are interested in finding:
- which driving conditions are more challenging for older adults?
- can voices with speaking styles improve older adults results in noise?

## 2. Speech recordings in studio

1 Female / 1 Male English speaker
- 4 speaking styles (plain / confident /clear / Lombard)
- 100 sentences (50 Harvard / 50 SatNav)

1 Male voice talent English speaker
- 2 speaking styles (plain / Lombard)
- 100 sentences (100 Harvard)

## 3. Speech recordings in car

Binaural audio recordings using a head and torso manequin placed in the front seat of a hybrid car.

Speech material was played using the car audio system (4 loudspeakers) at reasonable volume settings.

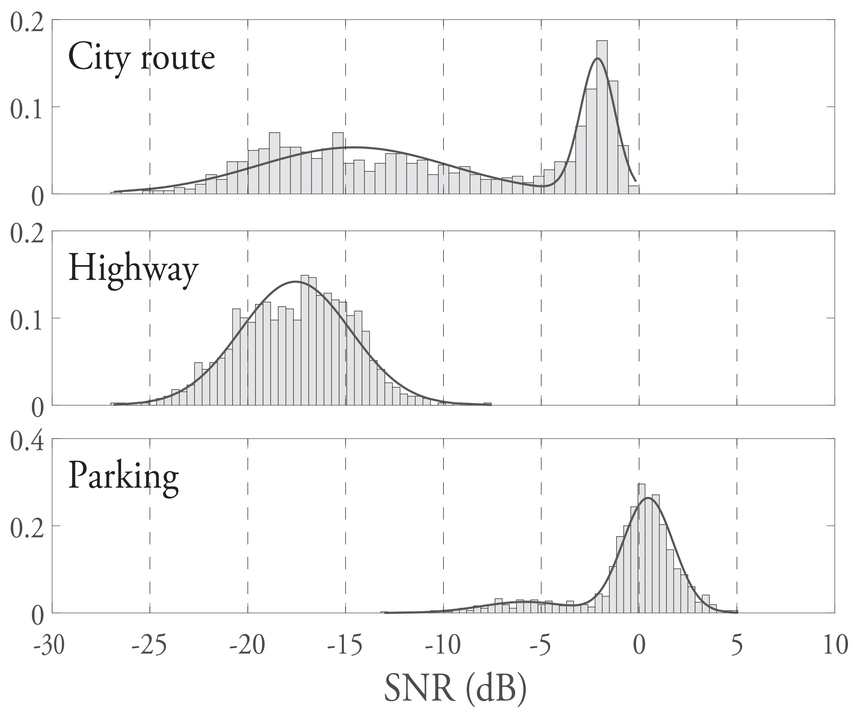
Routes:
City / Highway / Parked

Driving conditions:
- WC - windows closed only
- CS - competing speaker WC
- RAIN - rain WC
- BGM - background music WC
- HZ - hazard light on WC
- WO - windows open

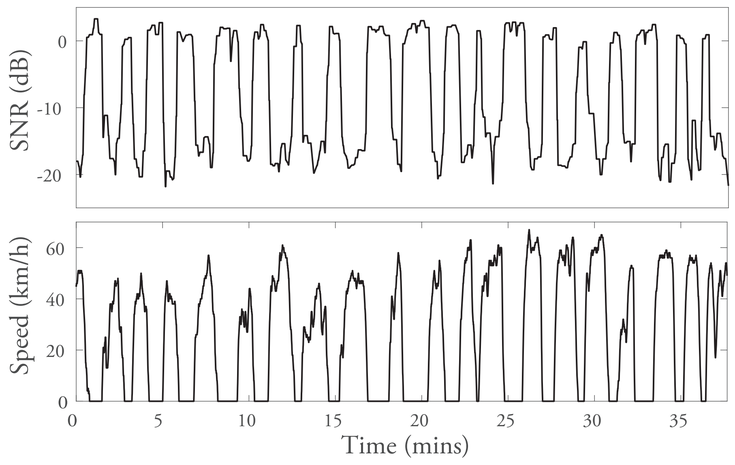
## 4. Data selection

- Recordings were segmented at a sentence level
- The SNR of each sentence was estimated using clean speech as reference and an estimate of the car impulse response
- Gaussians were fitted to the SNR distribution in each route / driving condition.
- Sentences whose SNR values were further than one stadard deviation away from the mean of the fitted Gaussian were excluded from the listening test.

SNR distribution in different routes WC condition




SNR and speed along the city route WC condition



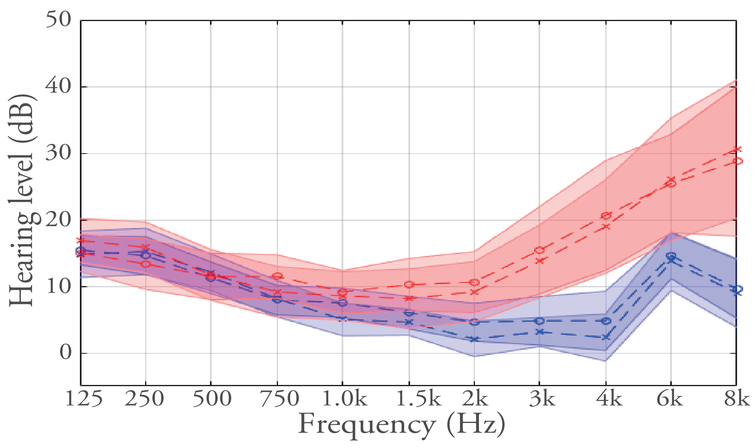

## 5. Listening test

Intelligibility task:
- Participants were asked to transcribe 100 different sentences, of different voices and noise conditions.
- Every sentence could only be heard once.
- Audio samples were played using headphones and responses were entered using a keyboard.

Two groups of native English participants:
- Young (24 participants: 22.7years; 19-29)
- Old (24 participants: 61.1years; 52-76)

All participants reported not being aware of any severe hearing problem.

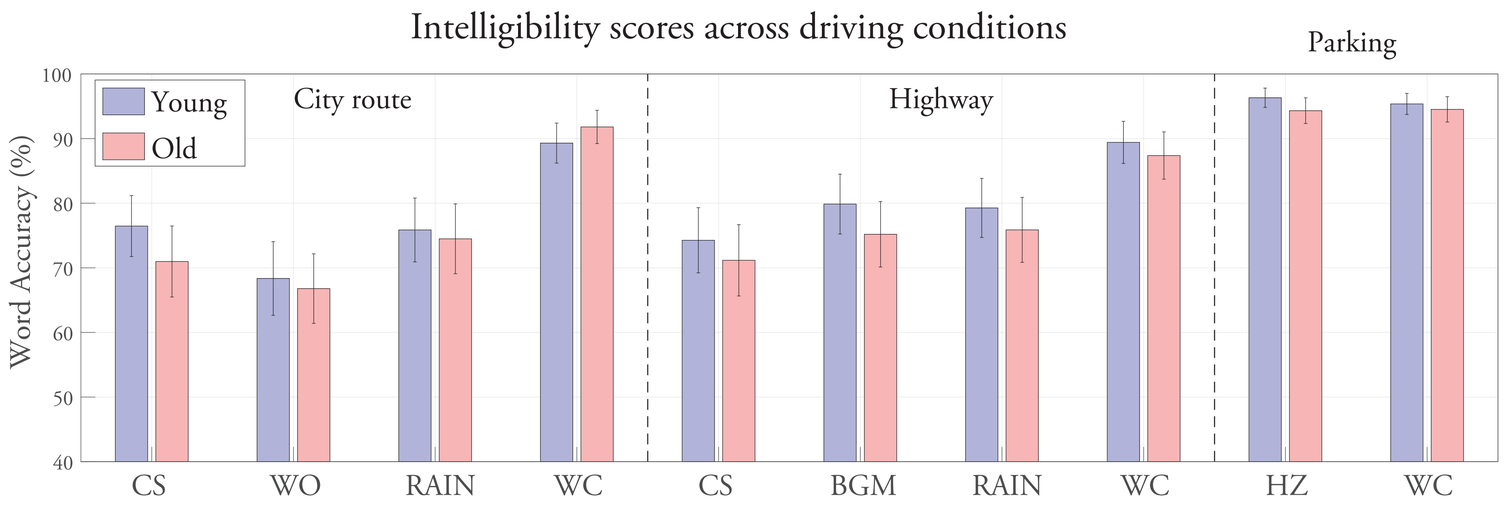
Average hearing curves (o: right ear; x: left ear)




## 6. Results and discussions

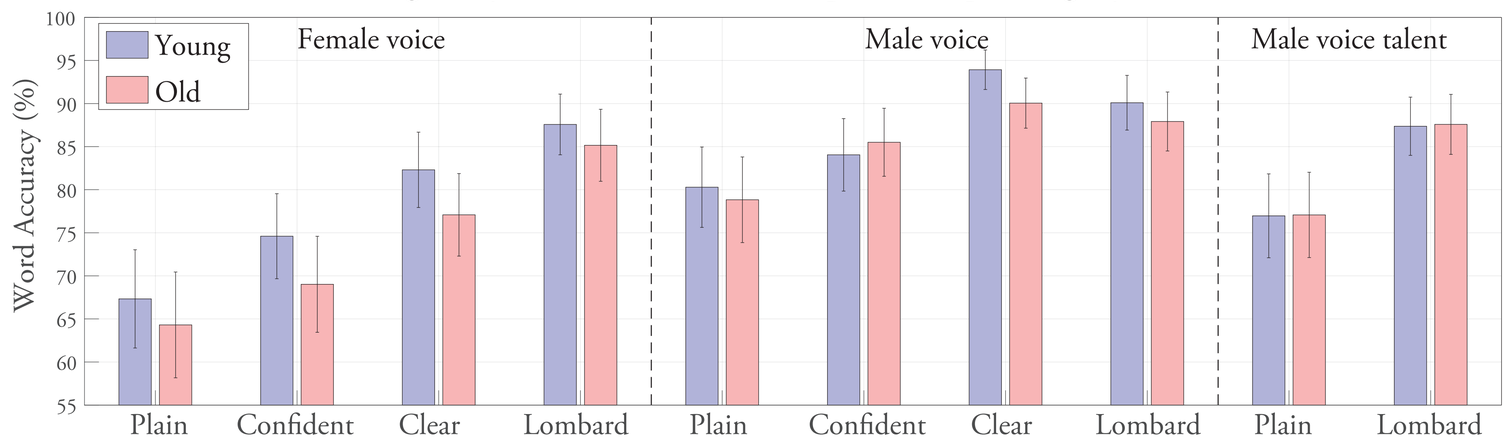
Results are presented in terms of Word Accuracy (%) calculated at a sentence level and averaged accross listeners within each listeners group.

Intelligibility scores across driving conditions




- The least intelligible condition for both groups of listeners was the city route WO.
- Older adults performed comparatively worst in the BGM and CS conditions.

Intelligibility scores across voices (speaker / speaking style)




- All speaking styles improved intelligibility compared to plain speech. This was true for all speakers (even the non professionals) and for both listeners groups.
- Older adults performed comparatively better with the Lombard speech.
- The female speaker was the least intelligible overall (when results are averaged across noise conditions), but for the WC, we observed that the female plain voice was as intelligible as the plain male voices.

**Acknowledgements:** The work presented in this paper was partially supported by the TOYOTA motor cooperation.