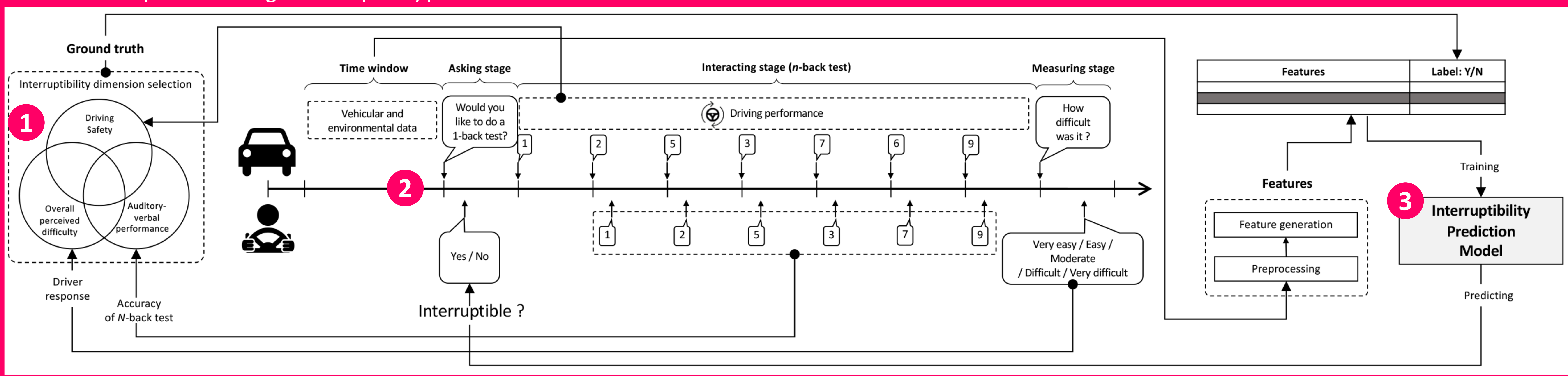
Predicting Opportune Moments for In- Proactive Speech Services

Auk Kim, Woohyeok Choi, Jungmi Park, Kyeyoon Kim, Uichin Lee

The visual description of training an interruptibility prediction model



1 Driver Interruptibility Definition and Metrics

Driving safety measures 'how safely a user drives a vehicle', and indicates interruptible if driving performance was not degraded when dual-tasking of driving and secondary tasks compared to when performing driving task alone (baseline).

Auditory-verbal performance measures '*how well a user performs an auditory-verbal task*', and indicates interruptible if a driver correctly answered all the items in a given n-back test.

Overall perceived difficulty measures 'how difficult it is to perform a dual task', and indicates interruptible if a difficulty rating in the measuring stage is lower.

2 Secondary Task

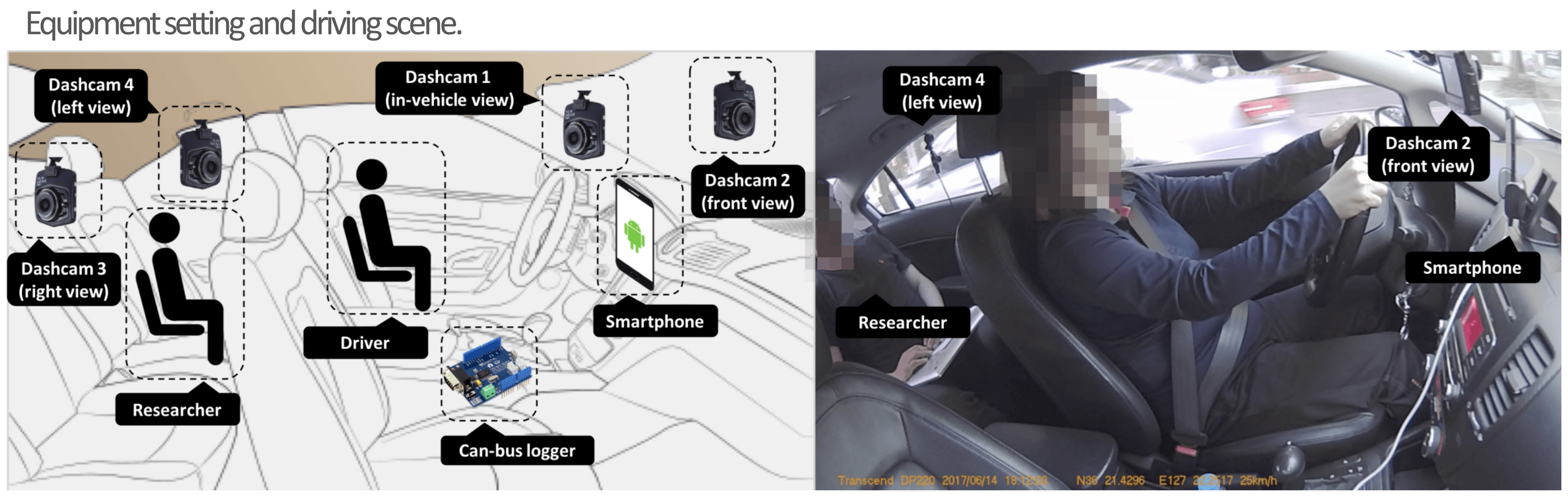
Secondary task consists of three stages: asking, interacting, measuring.

To systematically induce varying levels of cognitive demand, we employed three varying level of *n*-back: 0-back (a very mild task demand), 1-back (a moderate level), or 2-back (a high level of task demand).

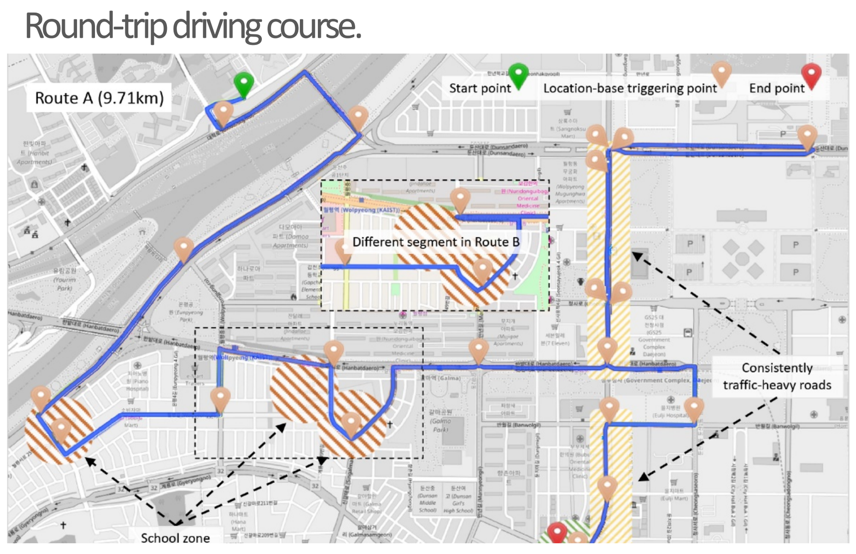
On-road Data Collection and Results

29 subjects drove driving course for twice (baseline and secondary-task driving session). During the secondary-task session, the drivers performed an average of 47.86 (SD = 6.83) secondary tasks

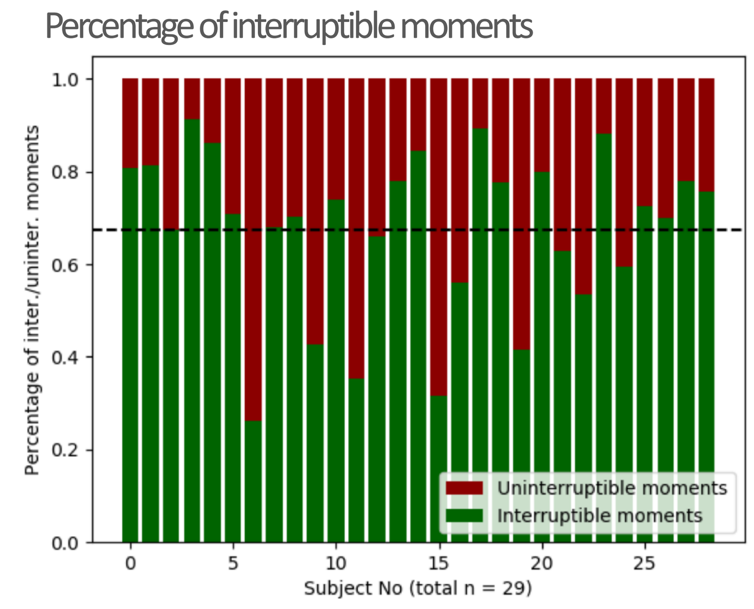
Equipment setting and driving scene.



Round-trip driving course.



Percentage of interruptible moments



3 Interruptibility Prediction

Interruptibility Labeling and Feature Generation

Interruptibility was labeled as a binary outcome - interruptible when all three dimensions indicated interruptible (n = 939); otherwise uninterruptible (n=449).

To generate features, we used the vehicle and environmental data that were collected before the start of a secondary task execution.

Paper Link QR Code

Selection of Best-performing ML Algorithm and Window Size

- We first examined the general models using an aggregated dataset of all drivers.
- The random forest model achieved the best performance among the four ML algorithms.

Performance (F-measure) of general models against machine learning (ML) algorithm and window sizes.

		Window size (in seconds)				
		1	2	3	4	5
ML algorithm	Decision Tree	0.57	0.59	0.59	0.56	0.58
	SVM	0.18	0.15	0.14	0.13	0.13
	Naïve Bayes	0.73	0.70	0.65	0.61	0.61
	Random Forest	0.73	0.74	0.70	0.71	0.69

Driver Variance

- The interruptibility could be varied by individual differences.
- For the user-specific models, each model was individually trained and tested with specific user data.
- The average performance (F-measure) of the user-specific models models was similar to the performance of the general model.

Performance of user-specific models. For the user- specific models, the value of F-measure shows the average of value among the models.

	Model type	
	User specific	General
Average F-measures	0.71	0.74

Interactive Computing Lab
KAIST Knowledge Service Engineering Dept.
KAIST
HYUNDAI