

ASR of Non-Natives in Noisy Environment

- Human error level: 4–6% WER (word error rate).
- Best neural nets are reportedly there, too.
- Our test of 90-second speeches of high-school students:
 - Average WER: 40–50% KIT, 80–90% Google, UEDIN.

The best recognized segment:

| Manual | Google | UEDIN | KIT |
|---|---|---|--|
| why do you wear those high heels , if you would wear some sneakers ? I know one really good store , that deals with the sale of freetime | why do ∅ where does high heels if you would wear some sneakers I no one really good star that deals with the sale of freedown food to | 're ready where tells us if you would ∅ sneaker us I know won the really good store that deals with the sale of freedown food | why do are those highs heels if you would where some sneakers i no one really good story that deals with the sale of freetime food to our |