

Improving ASR performance on Jasmin Flemish Dutch data by performing frequency perturbation

1. Background

- ASR(automatic speech recognition) contains **bias**.
- Flemish is the worst performing in **Jasmin corpus** [1].
- Work has been done on other corpora like accented British English. [2]
- Collecting data is expensive, so **data augmentation** is used.
- Different techniques can be used like **pitch shift** and **SpecAugment**. [3]
- Frequency perturbation** used

2. Research question

Can **data augmentation** improve the **ASR performance** on **Jasmin Flemish Dutch** data?

Can we get a WER better than the original baseline for the three speaker groups of **children, adults** and **older adults**?

3. Method

- Prepare Flemish Dutch data from the **Jasmin-CGN** corpus and **split** it 90/10
- Train **baseline** ASR system using **kaldi**
- Apply **frequency perturbation** to audio files
- Train new ASR system with old data plus new **augmented data**
- Repeat steps 3 and 4 with different **parameters**

4. Results

- Baseline model is on par with previous research [1] with **some differences**
- 4 Different augmented ASR's which used **frequency perturbation**
- Slight **improvements** in all systems
- Degradation** occurring in some systems
- Augment3** has the best combined score but **Augment2** is the most consistent

	Baseline	Augment1	Augment2	Augment3	AugCombined
Combined	47,14	46,01	45,03	44,65	45.59
Read	34,88	33,87	33,15	32.09	32.76
Conversational	72,51	72,13	71,03	72.61	73.74
Male	54,22	52,17	51,23	50.89	51.56
Female	45,5	42,93	41,93	41.60	42.62
Children	51.22	48.87	47.08	46.75	48.12
Teens	37.91	37.15	36.29	36.34	37.41
Non native children	54.16	52.97	52.27	51.82	51.82
Adults	56.21	55.20	53.91	53.86	54.49
Elderly	45.90	45.92	45.10	44.71	45.58

Table 1: Percentage WER score for each ASR system

5. Discussion and Conclusions

- Data augmentation** can increase **ASR performance** for **Jasmin Flemish Dutch** data
 - Children, adults** and **older adults** all improved. But improvements vary.
 - There still is a **big gap** compared to the other Dutch dialects [1]
- Future improvements:**
- Only include **native speakers**
 - Try out different **parameters** for frequency perturbation
 - Perform **different data augmentation techniques**

[1] Siyuan Feng and Olya Kudina and Bence Mark Halpern and Odette Scharenborg, "Quantifying Bias in Automatic Speech Recognition", arXiv 2021.

[2] Najafian, M. (2016). Acoustic model selection for recognition of regional accented speech (Doctoral dissertation, University of Birmingham).

[3] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition," in INTERSPEECH, 2019.