

# Comparing Performance of ASR Systems on Native Dutch Children and Teenagers: Google vs. Microsoft

Gert van Dijk - Supervisors: Odette Scharenborg, YuanYuan Zhang

## 1. Introduction

**Research Question:** How do Google and Microsoft's ASR API compare when ran on native Dutch child and teenager speech?

### Child Speech Recognition (CSR):

- Accurate ASR is challenging but necessary.
- Crucial for applications aimed at younger people.
- Children make up a large portion of the userbase.
- Create a baseline

### Google and Microsoft:

- Largest companies with ASR systems.
- Comparing for potential future improvements.
- Relevant to a large number of people.
- Reference for choosing a fitting ASR system.

## 2. Method

### Jasmin-CGN:

Category	Groups
Gender	Male (M), Female (F)
Age	Child, Teenager
Dialect	N1b, N2c, N3b, N4a

### Metrics:

- Word Error Rate (WER):

$$WER = \frac{\text{Substitutions} + \text{Insertions} + \text{Deletions}}{\text{Number of words spoken}} * 100\%$$

- Character Error Rate (CER):

$$CER = \frac{\text{Substitutions} + \text{Insertions} + \text{Deletions}}{\text{Number of characters spoken}} * 100\%$$

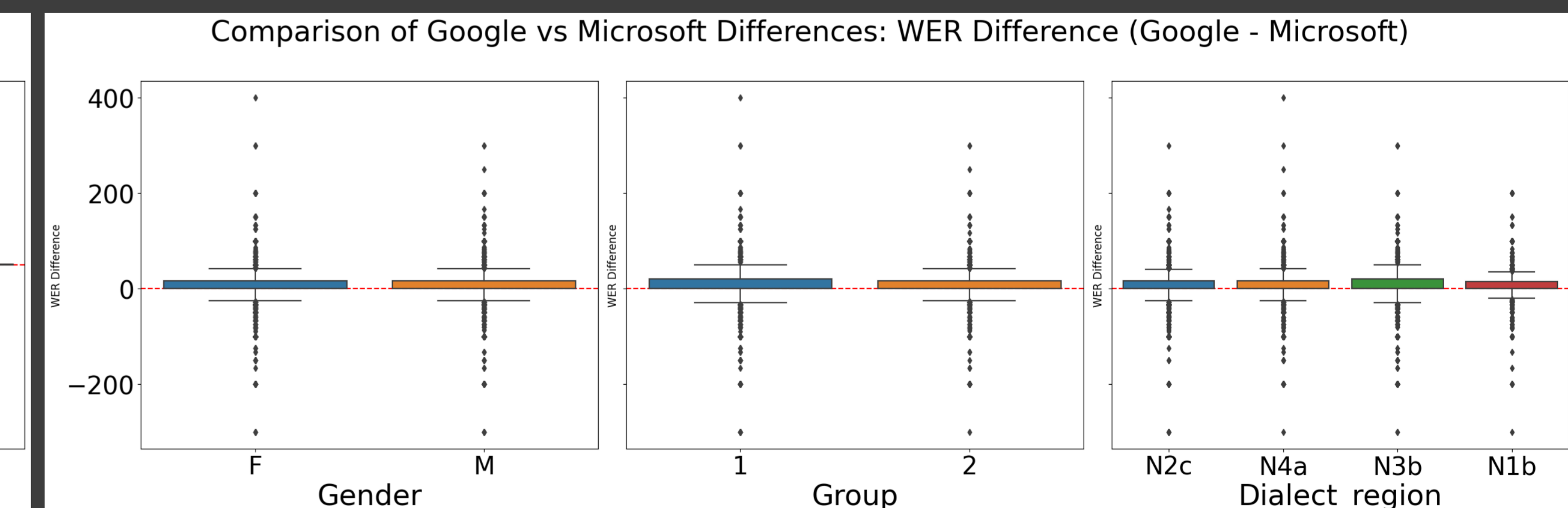
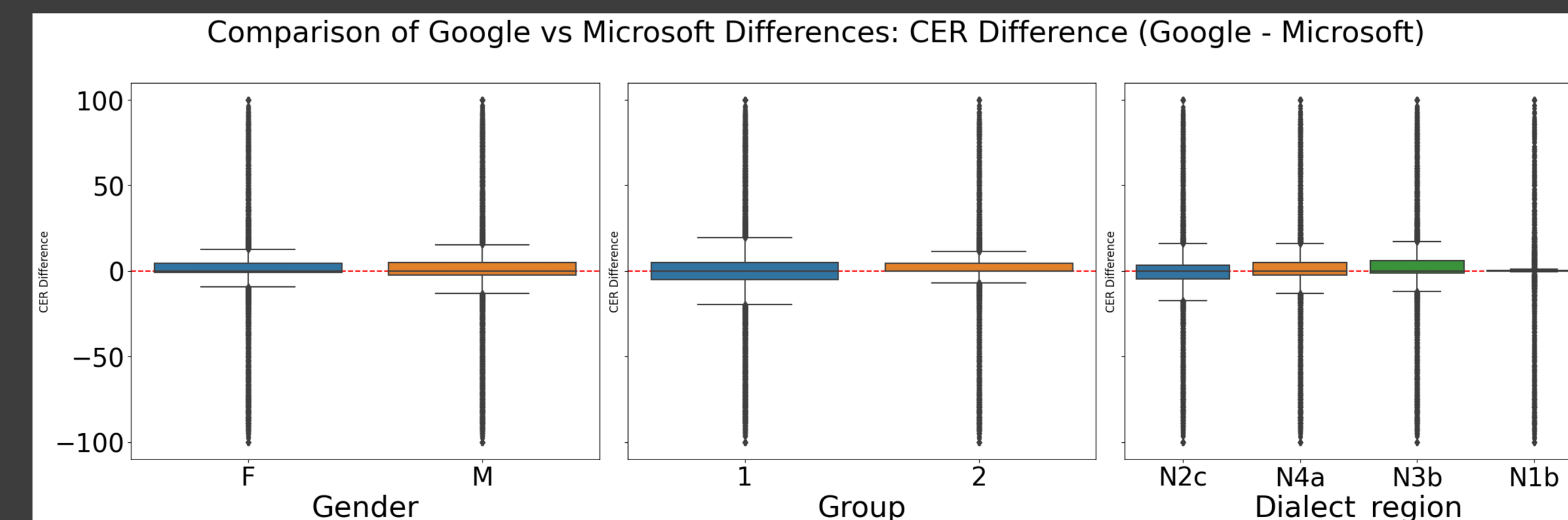
### Experiment:

- Segment files.
- Run all segments on both systems.
- Turn results into dataframe.
- Remove noise files and calculate the errors.

## 3. Results

The results indicate that Microsoft's ASR generally performs better in terms of WER across all demographics, while Google's ASR shows slightly better results in CER.

Metric	Child	Teen	Female	Male	N1b	N2c	N3b	N4a
Google WER	31.55	22.37	26.34	27.55	21.49	29.38	29.38	27.58
Microsoft WER	<b>26.96</b>	<b>16.44</b>	<b>21.07</b>	<b>22.30</b>	<b>16.08</b>	<b>25.07</b>	<b>23.98</b>	<b>22.02</b>
Google CER	<b>20.34</b>	<b>15.71</b>	<b>17.32</b>	<b>18.71</b>	<b>14.65</b>	<b>18.76</b>	19.23	<b>19.22</b>
Microsoft CER	21.89	17.21	19.12	19.96	16.76	21.77	<b>19.14</b>	20.80



## 5. Conclusion and future work

### Conclusions:

- Microsoft in terms of WER.
- Google in terms of CER
- Both bias towards teenagers compared to children.

- Both bias towards N1b compared to all other regions.
- Both slightly bias towards women compared to men.
- Overall Google is slightly less bias in terms of WER and CER

### Shortcomings:

- More metrics such as Phoneme Error Rate (PER)
- More data.