

Augmenting Dutch Adult Dysarthric Speech into Child-like Speech using Traditional Digital Signal Processing Techniques

1. Introduction

- **Automatic Speech Recognition (ASR)** is an AI-driven technology used to transcribe speech to text
- **Developmental Language Disorder (DLD)** is a neurodevelopmental language disorder
- **Dysarthria** is a motor speech disorder caused by physical muscle weakness, most commonly caused by stroke

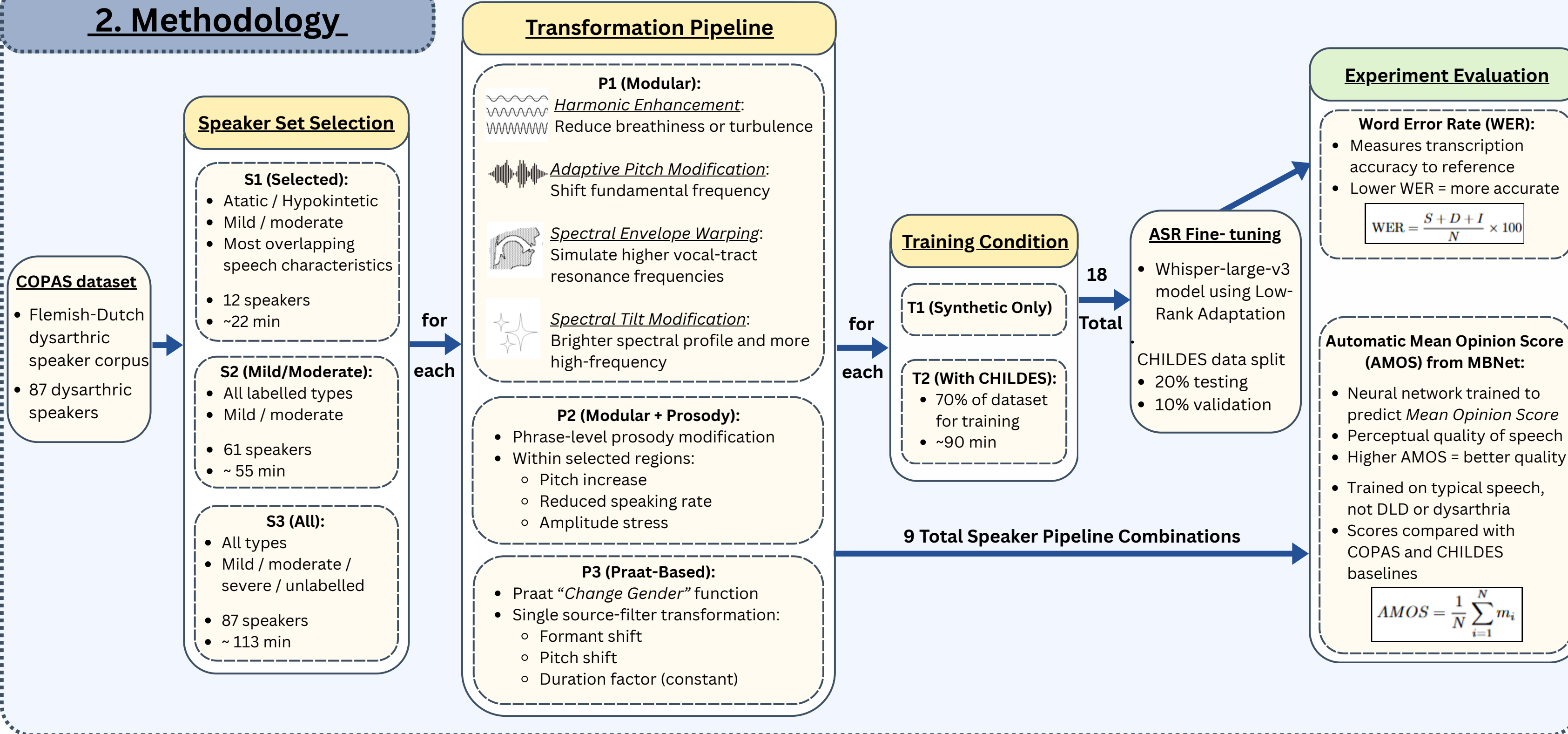
To what extent does converting Dutch adult dysarthric speech into Dutch child speech via **traditional acoustic transformations**, improve ASR performance for Dutch children with DLD?

3. Main WER Results

System	WER (%)
Whisper-large-v3 (Zero-shot)	46.48
CHILDES-only Fine-tuning	31.57
S1-P1-T2	34.82
S1-P2-T2 (Best)	29.00
S1-P3-T2	34.69
S3-P2-T2	38.34

- **Speaker Subset Selection**
 - The targeted S1 dataset consistently outperformed the larger S2 and S3
- **Transformation Pipeline**
 - P3 performed best on average, closely followed by P2, then P1
- **Training Condition**
 - On average combining synthetic speech with real CHILDES data (T2) lowered WER percentage
 - 2 configurations saw increases in WER from T1 to T2

2. Methodology



4. Key Findings

- **Overall improvement over zero-shot baseline**
 - **S1-P2-T2** was the only configuration to outperform both baselines
- **Domain Similarity > Dataset Size**
 - Carefully selected dysarthric speakers resulted in better ASR than larger and more diverse speaker sets
- **Prosody Matters**
 - Adding child-like prosodic characteristics improved ASR beyond pitch and spectral modifications alone
- **Speech Quality != ASR Utility for atypical speech**
 - Highest AMOS scores only had slight negative correlation with WER results
 - Configurations with AMOS scores closer to CHILDES baseline resulted in better ASR performance for children with DLD

5. Limitations

- Some experimental conditions were trained on relatively small datasets (smallest ~22 min), limits the stability and generalizability of the results
- Domain mismatch between Flemish-Dutch and Netherlands-Dutch could have introduced linguistic differences when comparing target and source

6. Conclusion and Future Work

- Synthetic child-like speech **improved** ASR performance for Dutch children with DLD
- Datasets **closer to target domain** were more useful than increasing dataset size
- Validate findings on larger datasets and additional languages
- Develop more advanced prosody modeling techniques for children