Phonetic Analysis and Tempo Adjustments for Improved Speech Recognition in Dysarthric Speakers

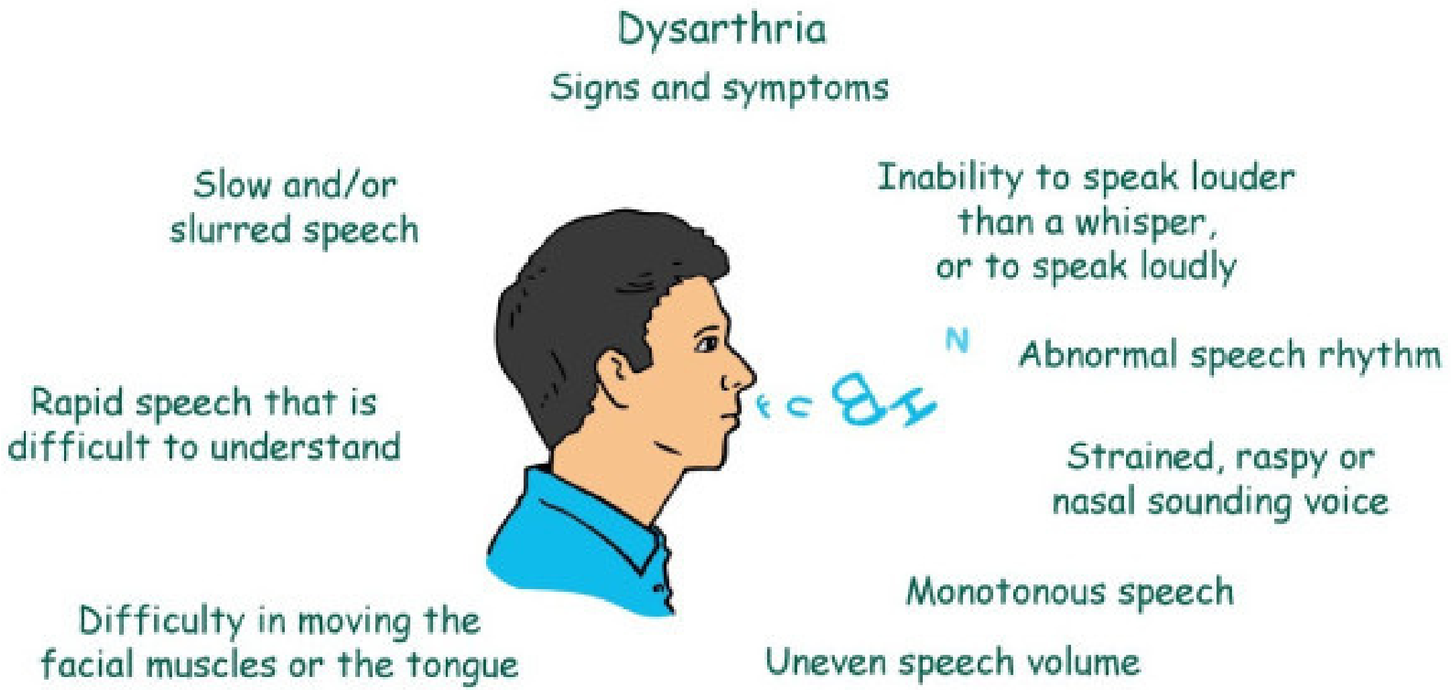
Feifei Xiong · Jon Barker · Heidi Christensen

IEEE ICASSP-2019

Mid-Evaluation Presentation

Saketh Reddy Vemula - 2022114014
Viswanath Vuppala - 2022101084

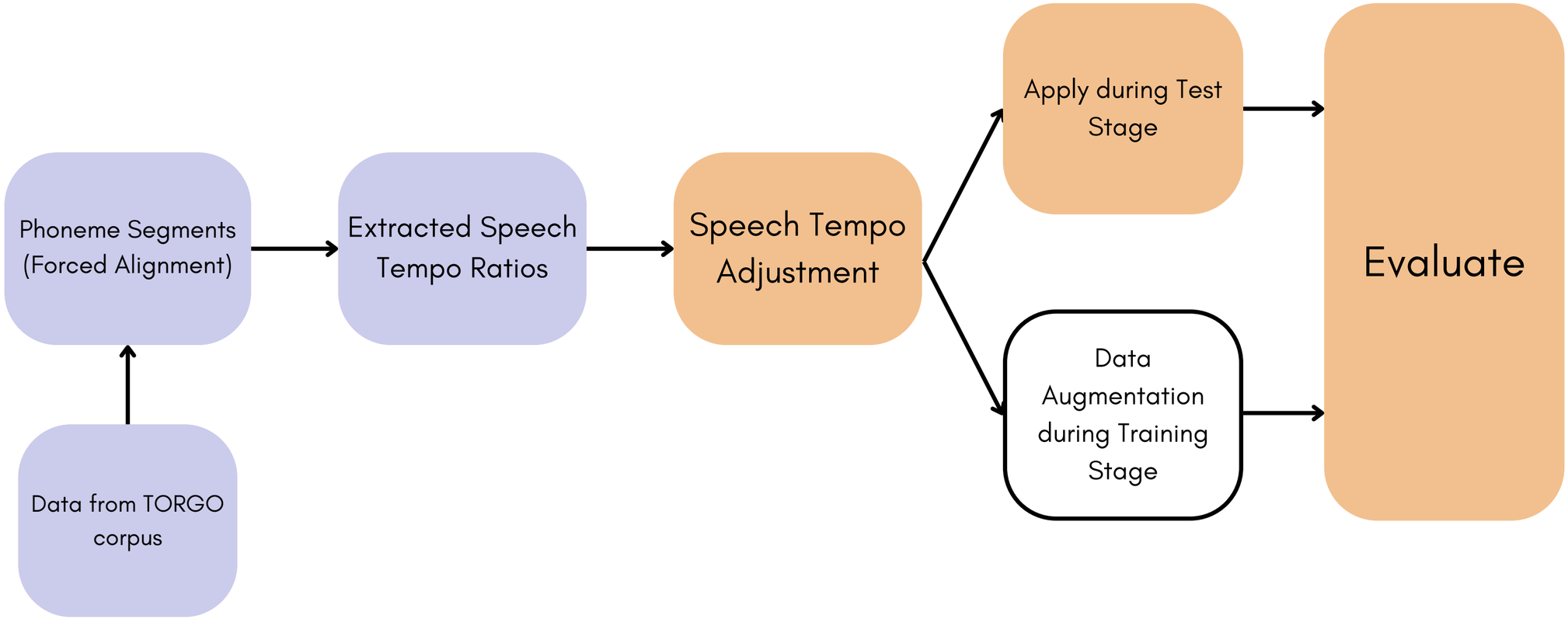
Introduction



- Increased respiration frequency
- Inadequate Pauses
- Breathy or hoarse voice
- Reduced speech
- Deviations in pitch and volume
- Mis-articulated Sounds

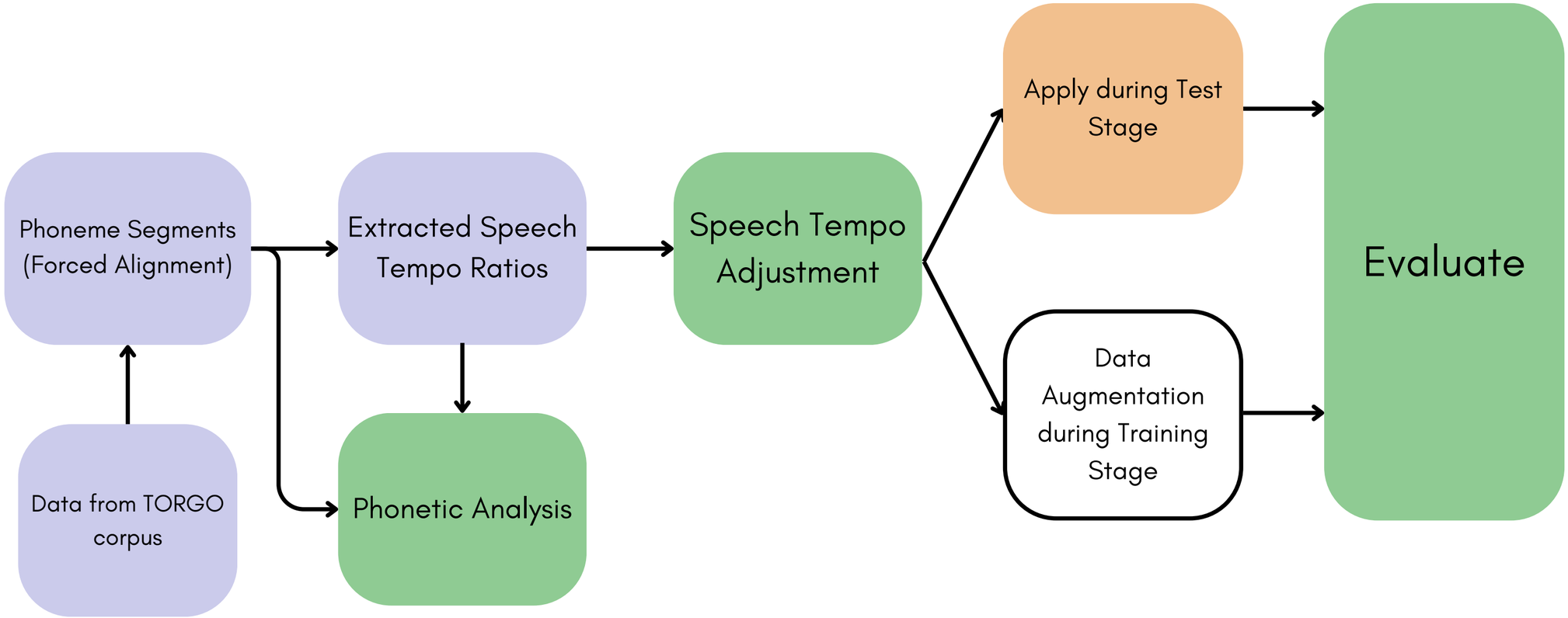
Earlier Conclusions

Presented Speech tempo adjustment for improved performance of ASR trained on typical speech

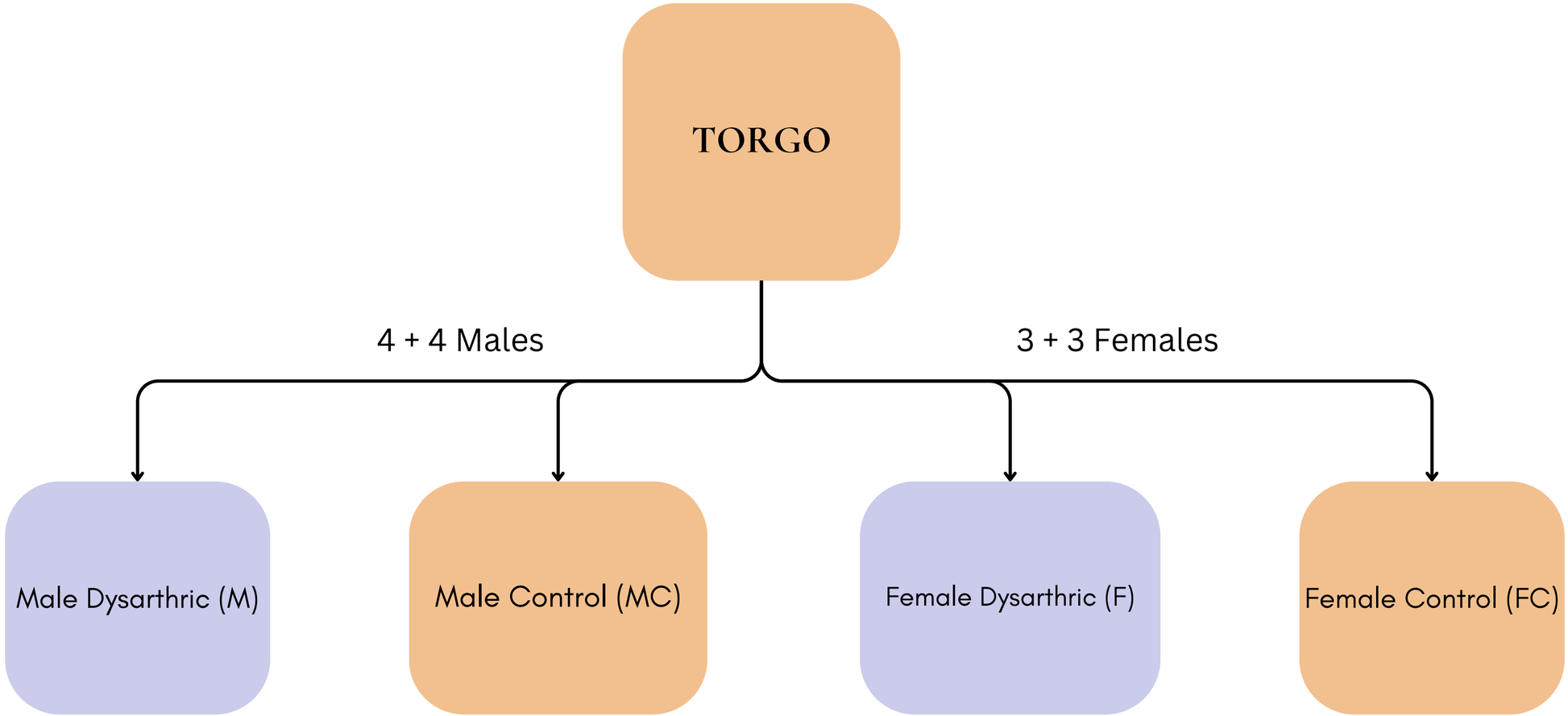


Objectives

- Analyze dysarthric speech at the phonetic level, specifically tempo.
- Implement phoneme-based tempo adjustments to assess their impact on automatic speech recognition (ASR) performance.
- Evaluate the effectiveness of phoneme-based tempo modifications in improving ASR accuracy for dysarthric speakers.



Dataset Overview



TORGO Database: Acoustic and articulatory speech from speakers with dysarthria

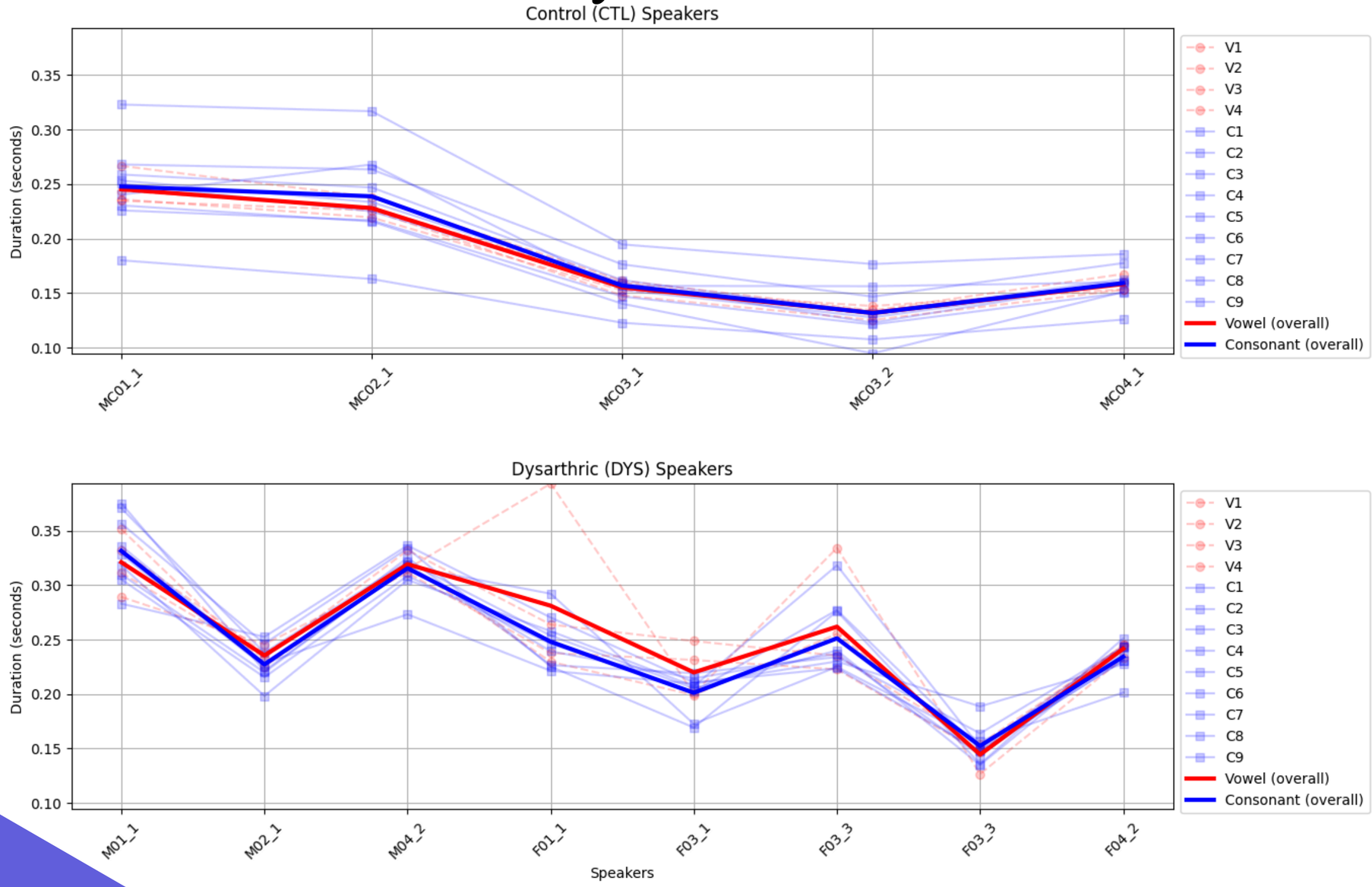
Phoneme Grouping

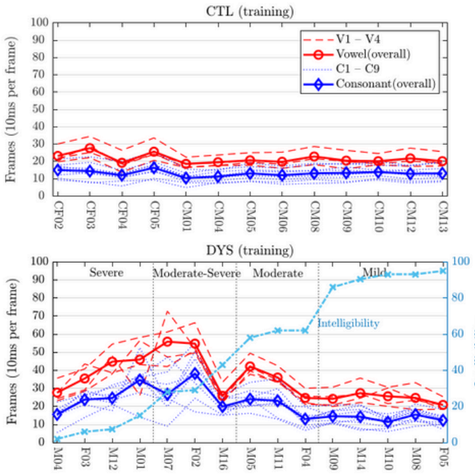
Vowels	V1: Short Vowels	AH AO AX EH IH UH
	V2: Medium Vowels	AE
	V3: Long Vowels	AA ER IY UW
	V4: Diphtongs	AW AY EY OW OY

Consonants	C1: Glides	L R W Y
	C2: Unvoiced Stops	K P T
	C3: Voiced Stops	B D G
	C4: Nasals	M N NG
	C5: Unvoiced Fricatives	F S SH TH
	C6: Voiced Fricatives	DH V Z ZH
	C7: Unvoiced Affricatives	CH
	C8: Voiced Affricatives	JH
	C9: Aspirates	HH
	Silence	bcl cl dcl gcl kcl nol nolo pcl tl
	Inconsistencies	sill lab2 a ahl ahn aor ee et ff h nh o rly

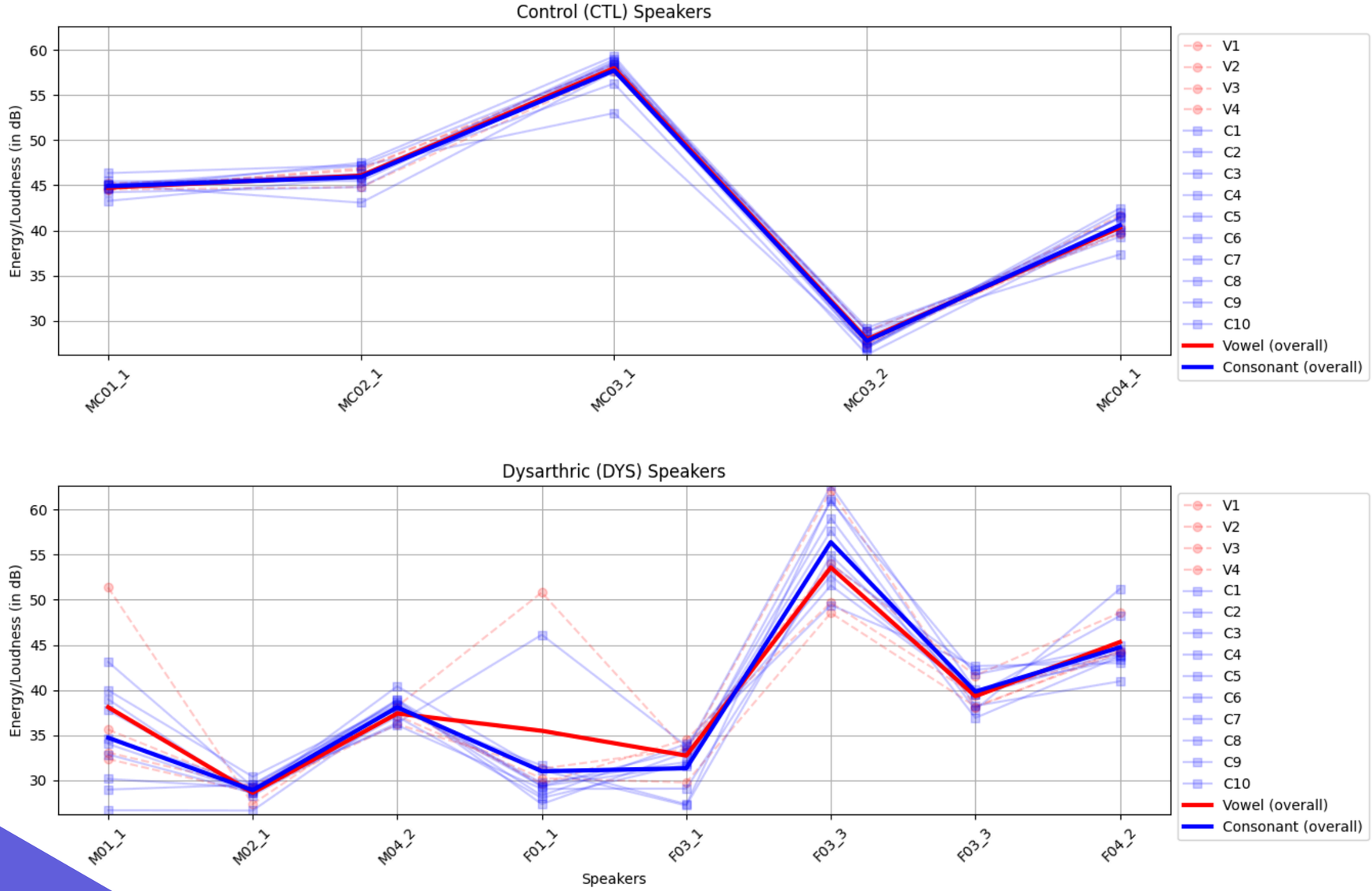
Google Sheet

Phoneme duration Analysis

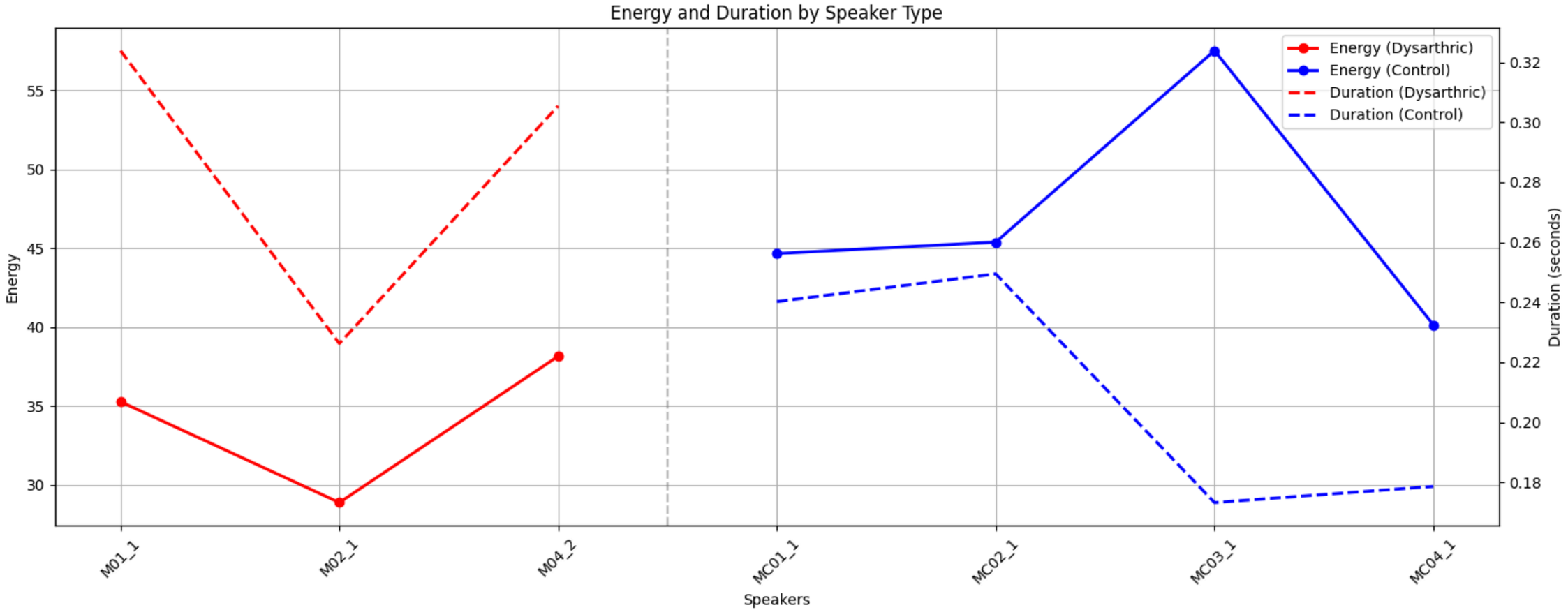




Phoneme Energy (Loudness) Analysis

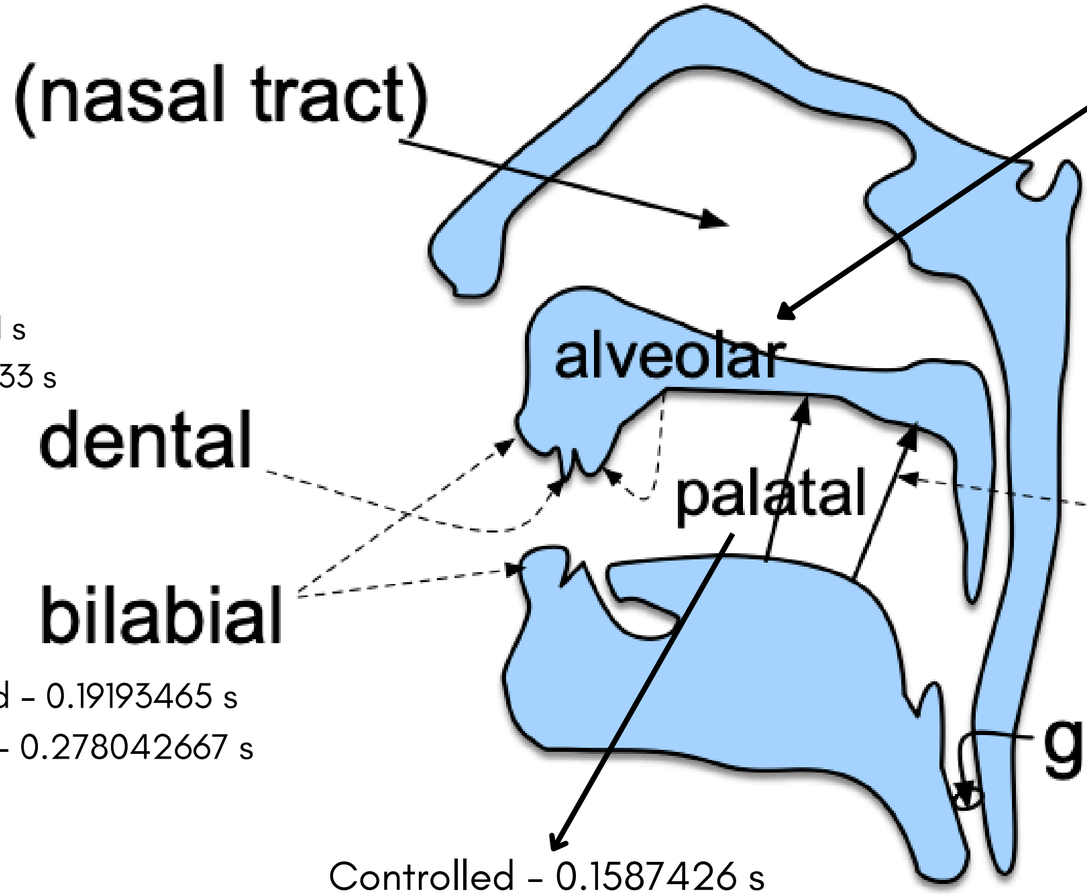


Overall Average Energy and duration



Post Mid Presentation

Place of Articulation Comparision



Controlled - 0.186117171 s
Dysarthric - 0.295834714 s

Controlled - 0.1752311 s
Dysarthric - 0.284252833 s

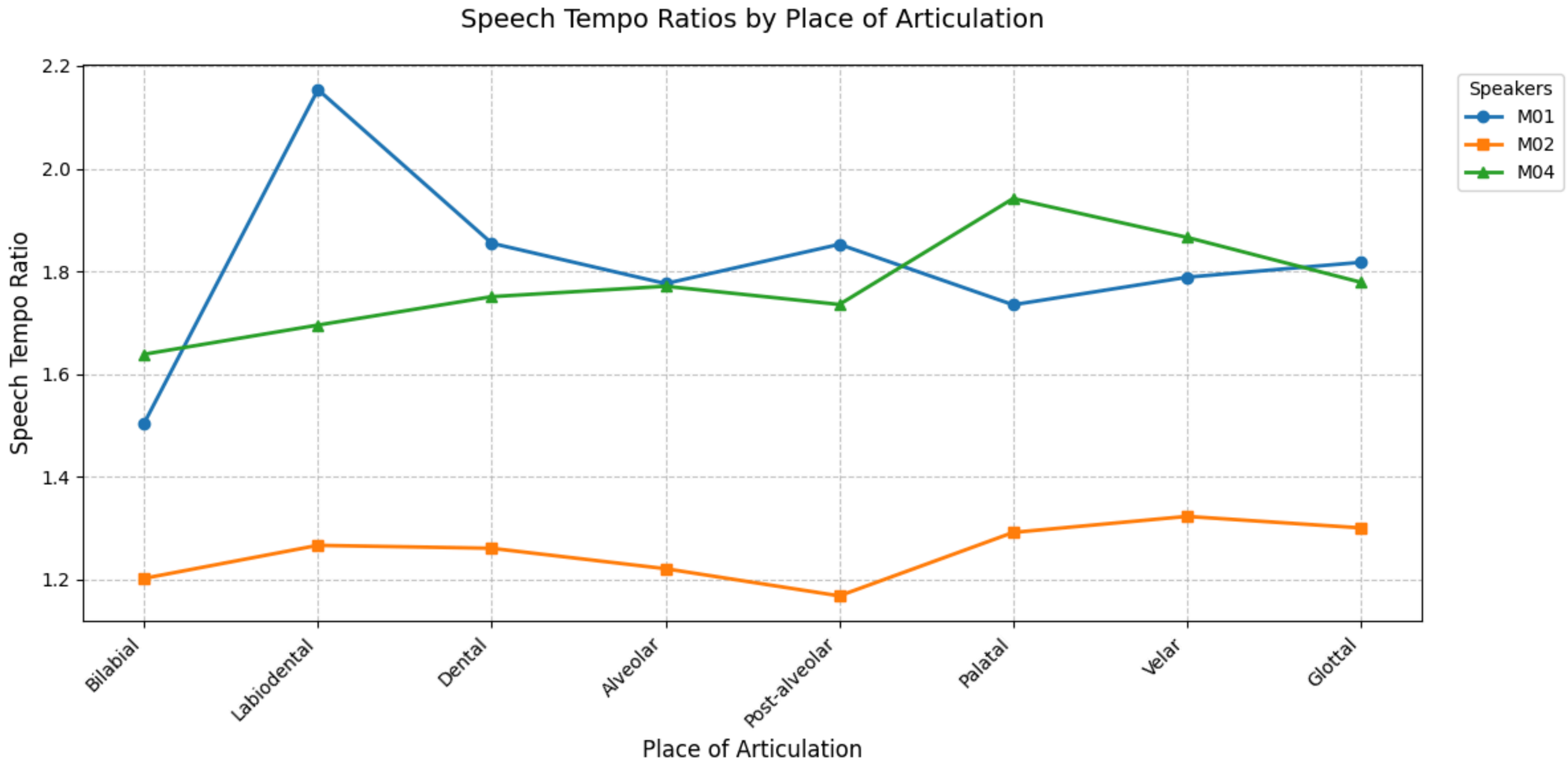
Controlled - 0.173535267 s
Dysarthric - 0.287948333 s

Controlled - 0.19193465 s
Dysarthric - 0.278042667 s

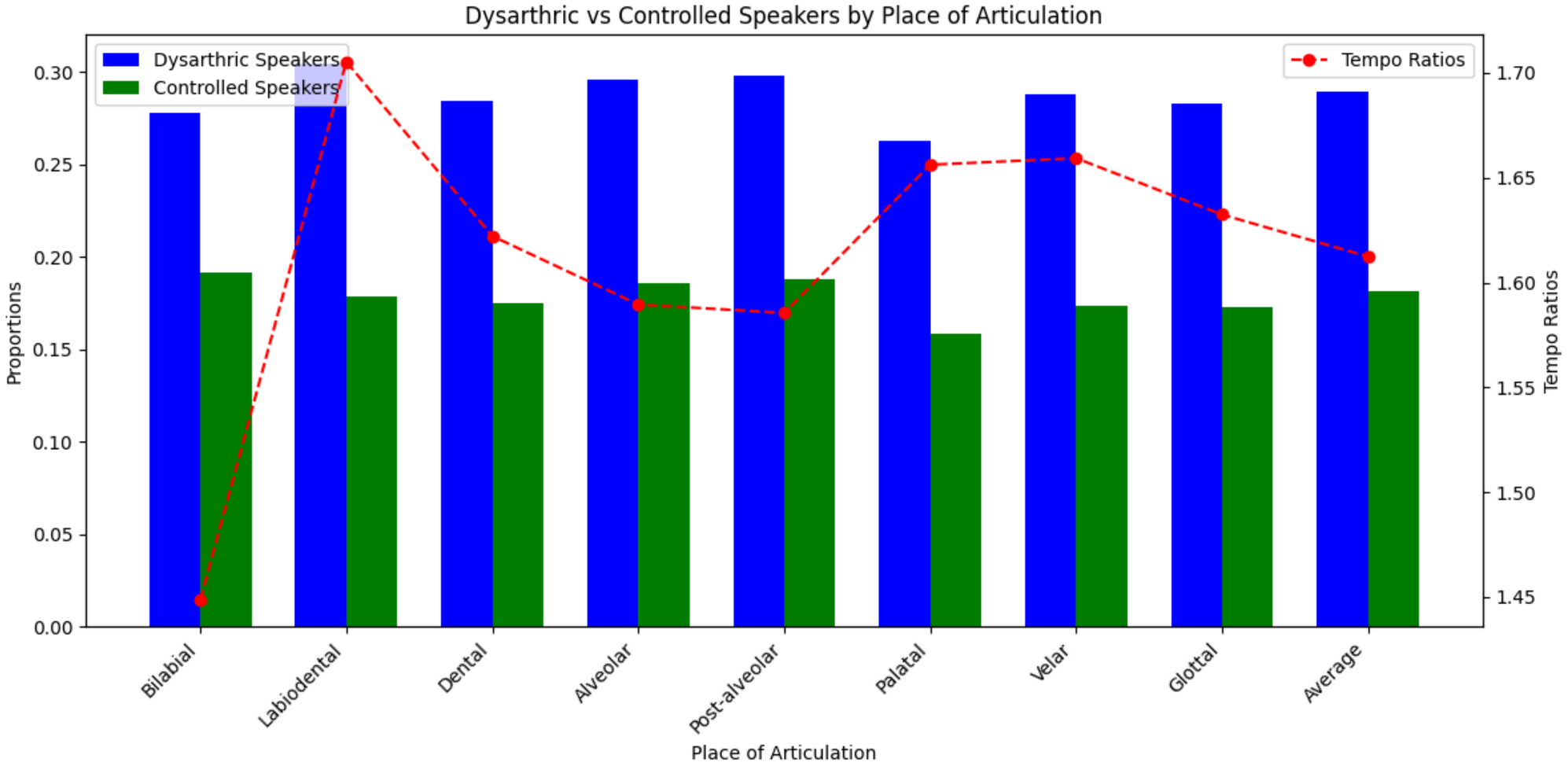
Controlled - 0.172931556 s
Dysarthric - 0.282291778 s

Controlled - 0.1587426 s
Dysarthric - 0.262909 s

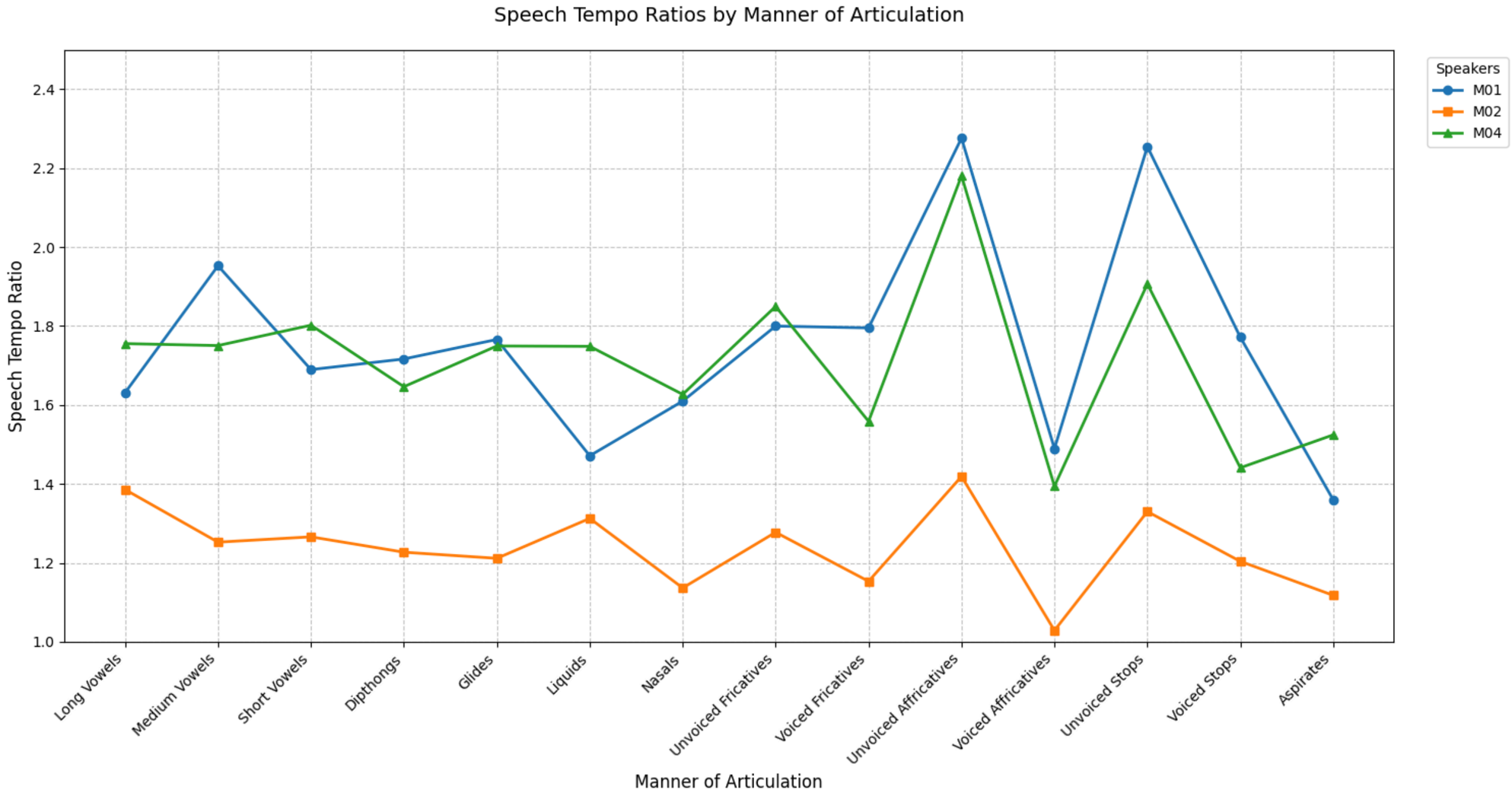
Speaker-wise Place of Articulation Analysis



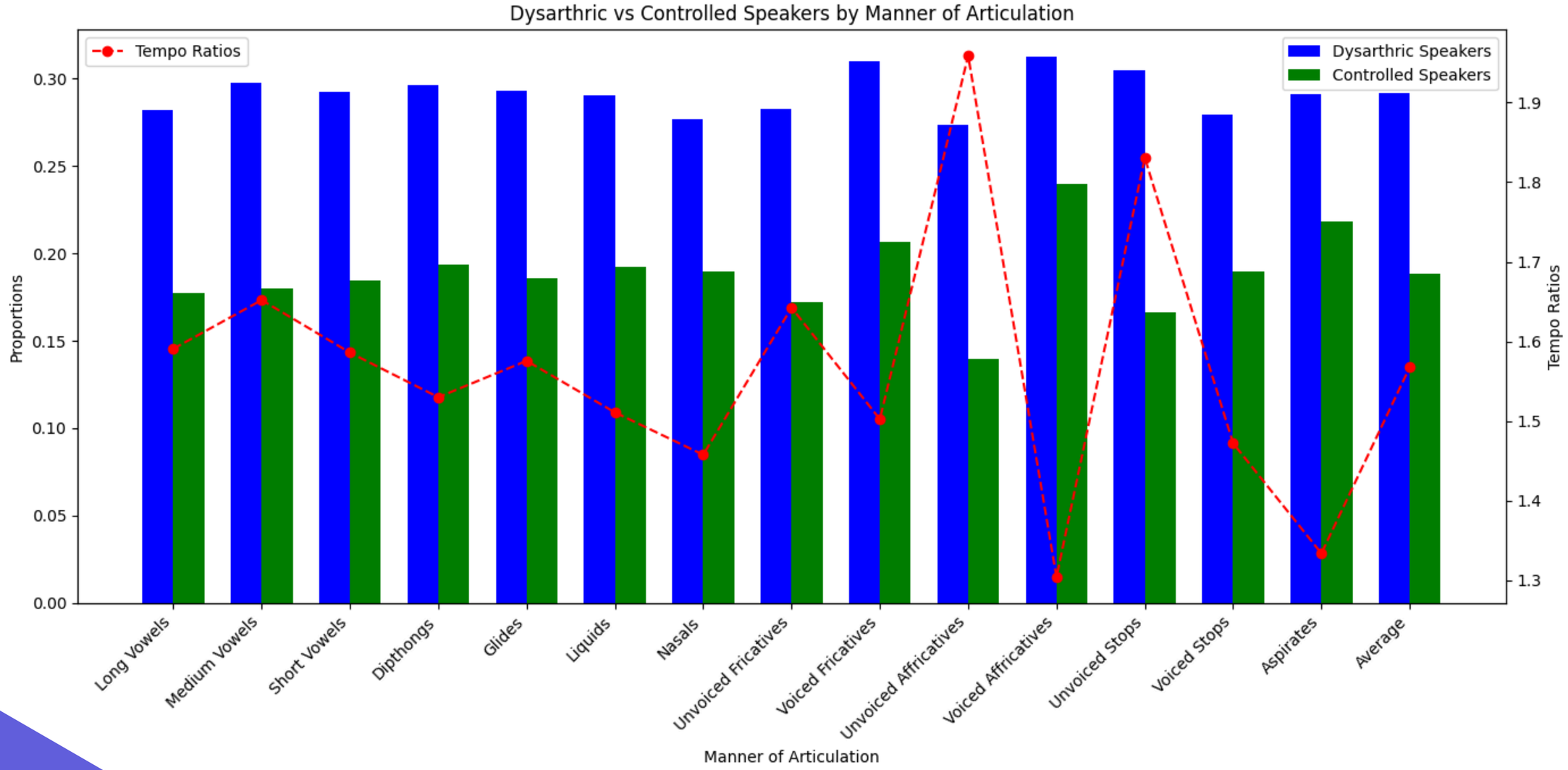
Overall Place of Articulation Comparision



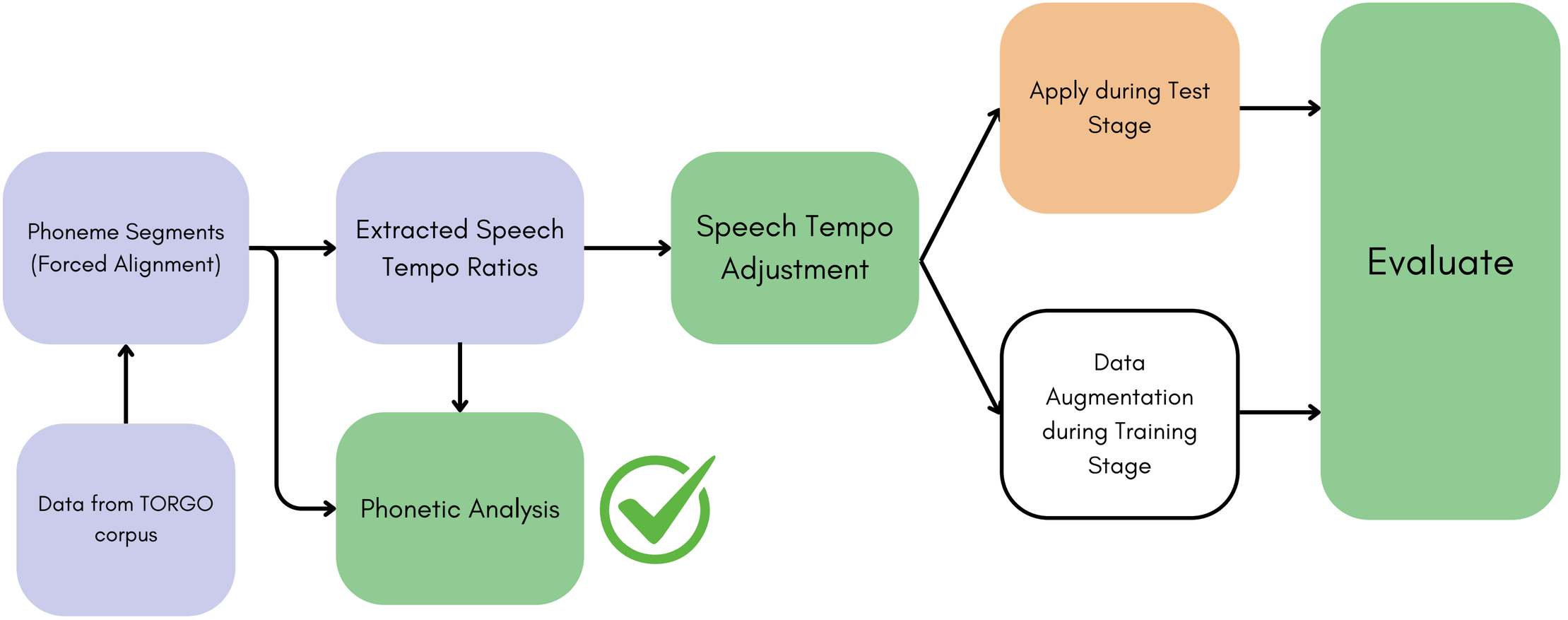
Speaker-wise Manner of Articulation Analysis



Overall Manner of Articulation Comparision

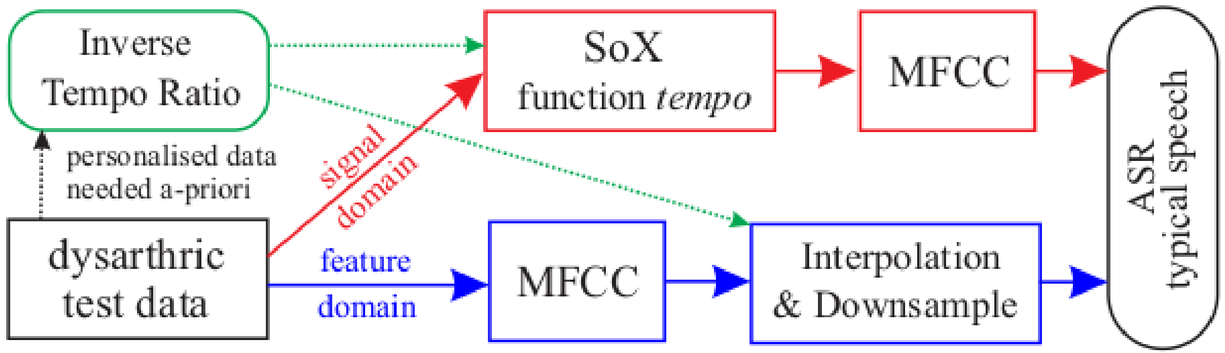


Objectives


Phoneme Segments (Forced Alignment)
Extracted Speech Tempo Ratios
Speech Tempo Adjustment
Apply during Test Stage
Evaluate
Data from TORGO corpus
Phonetic Analysis
Data Augmentation during Training Stage

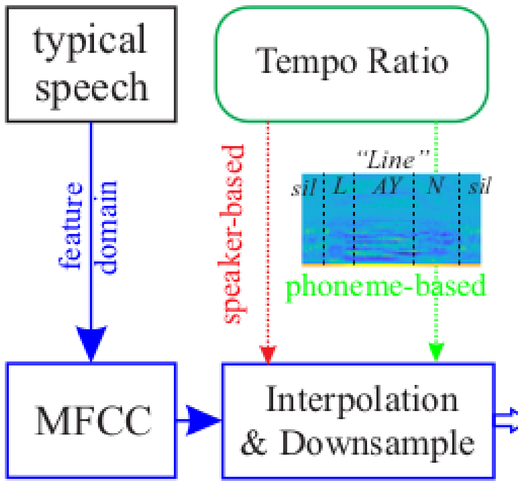
Speech Tempo Adjustment for ASR

Test Stage $\overline{\mathcal{R}_{d \leftarrow c}}$



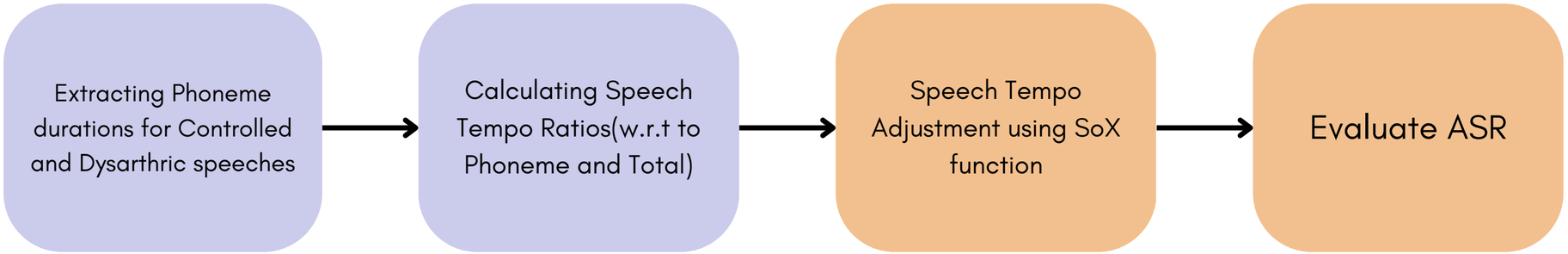
Training Stage

$$\mathcal{R}_{d \leftarrow c}(p) = \frac{T_d(p)}{T_c(p)}$$

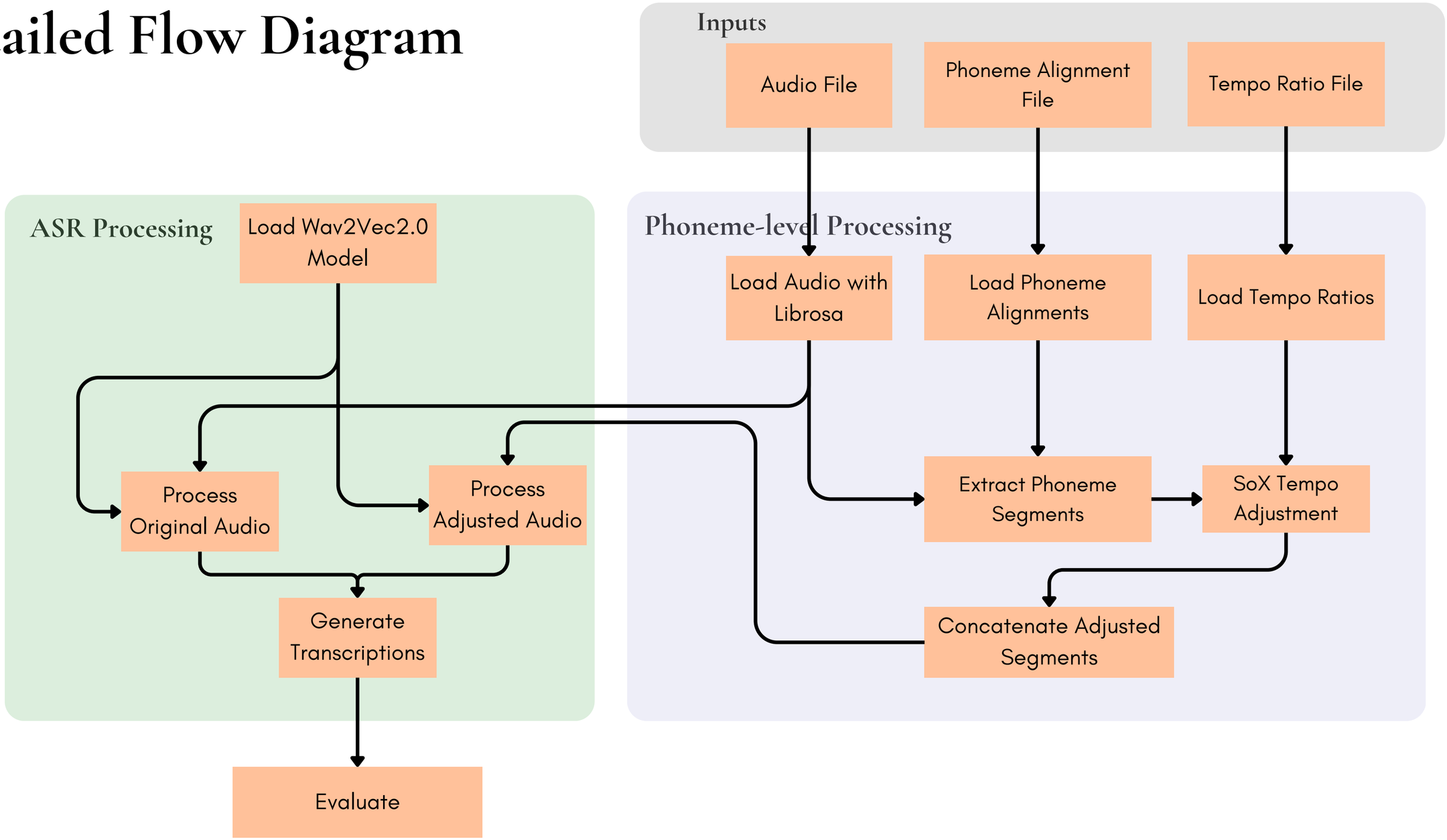


Speech Tempo Adjustment for ASR - Signal Domain

Test Stage $\overline{\mathcal{R}_{d \leftarrow c}}$:



Detailed Flow Diagram



Results

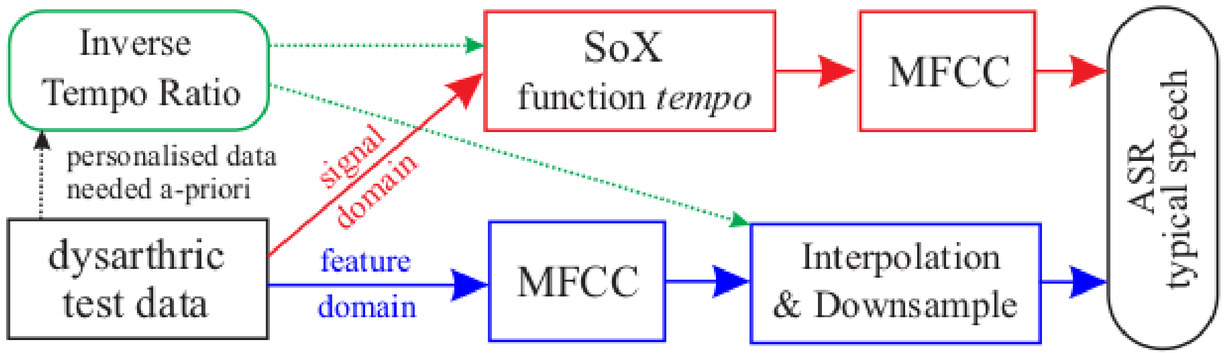
Session	Model	WER before Tempo-Adjustment	WER after Tempo-Adjustment	Average Original Duration (s)	Average Adjusted Duration (s)
M01_Session_1_headMic	facebook/wav2vec2-base-960h	1.0332	1.0165	4.7500	3.9738
M02_Session1_arrayMic	facebook/wav2vec2-base-960h	1.0083	0.9809	2.6993	2.5054
M04_Session2_headMic	facebook/wav2vec2-base-960h	1.0237	1.0327	4.5883	3.6116
M01_Session1_headMic	jonatasgrosman/wav2vec2-large-xlsr-53-english	0.9829	0.9096	4.7500	3.9738
M02_Session1_arrayMic	jonatasgrosman/wav2vec2-large-xlsr-53-english	1.0934	1.1123	2.6993	2.5054
M04_Session2_headMic	jonatasgrosman/wav2vec2-large-xlsr-53-english	1.0302	1.0738	4.6133	3.6295

Results

Future Scope and Limitations

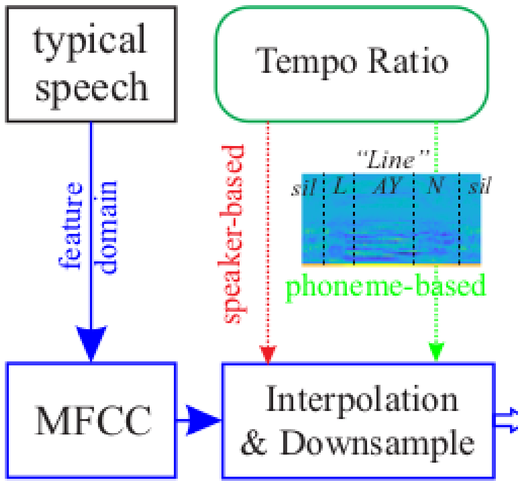
Test Stage

$$\overline{\mathcal{R}_{d \leftarrow c}}$$



Training Stage

$$\mathcal{R}_{d \leftarrow c}(p) = \frac{T_d(p)}{T_c(p)}$$



References

[1] H. Kim, M. H. Johnson, J. Gunderson, A. Perlman, T. Huang, K. Watkin, S. Frame, H. V. Sharma, and X. Zhou. Uaspeech, 2023.

[2] F. Rudzicz, A. K. Namasivayam, and T. Wolff. The torgo database of acoustic and articulatory speech from speakers with dysarthria. Lang. Resour. Eval.,
46(4):523–541, Dec. 2012.

[3] G. Schu, P. Janbakhshi, and I. Kodrasi. On using the ua-speech and torgo databases to validate automatic dysarthric speech classification approaches, 2022.

[4] F. Xiong, J. Barker, and H. Christensen. Phonetic analysis of dysarthric speech tempo and applications to robust personalised dysarthric speech recognition. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5836–5840, 2019.

Codes link

Thank You