Impact of speaking rate on the source filter interaction in speech: A study

Poster ID: 5177

Tilak Purohit, Achuth Rao M V and Prasanta Kumar Ghosh

SPIRE Lab, Electrical Engineering, Indian Institute of Science (IISc), Bangalore, India

Introduction

- **Source filter interaction (SFI)** explains the **drop in pitch** caused due to the constriction in the vocal tract during **voiced consonant production** in a vowel-consonant-vowel (VCV) sequence.
- **Voiced consonants** require the vocal cords to vibrate, hence producing their signature sounds; voiceless consonants do not.
  - /b/ & /g/ are voiced consonants; and vowel sounds like (/a/, /e/, /i/, /o/, /u/) are all voiced.
  - /p/ & /f/ are voiceless consonants.
- **Our hypothesis:**
  - The pitch drop during a voiced consonant in the context of the different vowels could be different given that the articulatory shapes for different vowels are different.
  - The speaking rate affects the movement of the articulators and can have a significant effect on the pitch drop as well.
- **Motivation:**
  - Most work on SFI focus on VCV (or CV) sequence with the vowel being /a/, we extend the study by using VCV sequences with five different vowels.
  - No work present in the literature that investigates SFI in context of speaking rate.

Dataset

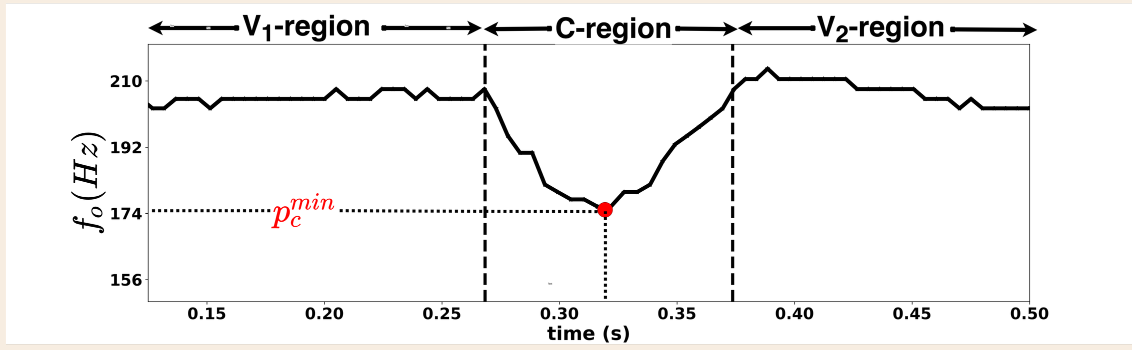
	EGG Corpus	SPIRE VCV Corpus
Modalities	Acoustic + EGG	Acoustics only
Speaking Rate	Normal	Slow, Normal, Fast
Vowels	/a/	/a/, /e/, /i/, /o/, /u/
Consonants	/b/, /g/, /j/, /v/, /z/	/b/, /d/, /g/, /v/, /z/
Audio	16 kHz	16 kHz
Subjects	9 (3F + 6M)	6 (3F + 3M)
Avg. Age	23-25 years	18-22 years
Samples	315	450 × 3 rates
VCV Boundary	manually annotated	manually annotated
Outliers	98 samples	220 samples (13.6 % slow 15.5 % normal 19.8% fast)

- Electroglottography(EGG) signals were recorded at 16kHz.
- **EGG corpus:** For finding out which glottal closure instance (GCI) detection algorithm is the best to estimate the pitch from the acoustic signal in the context of SFI study.
- **SPIRE VCV:** For investigate the pitch drop trend across different vowels and consonant combinations in three different speaking rate.
- **Outliers:** Those samples where the voicing signature were absent in the consonant region were considered as the outliers. These were detected by manually inspecting the audio waveform.

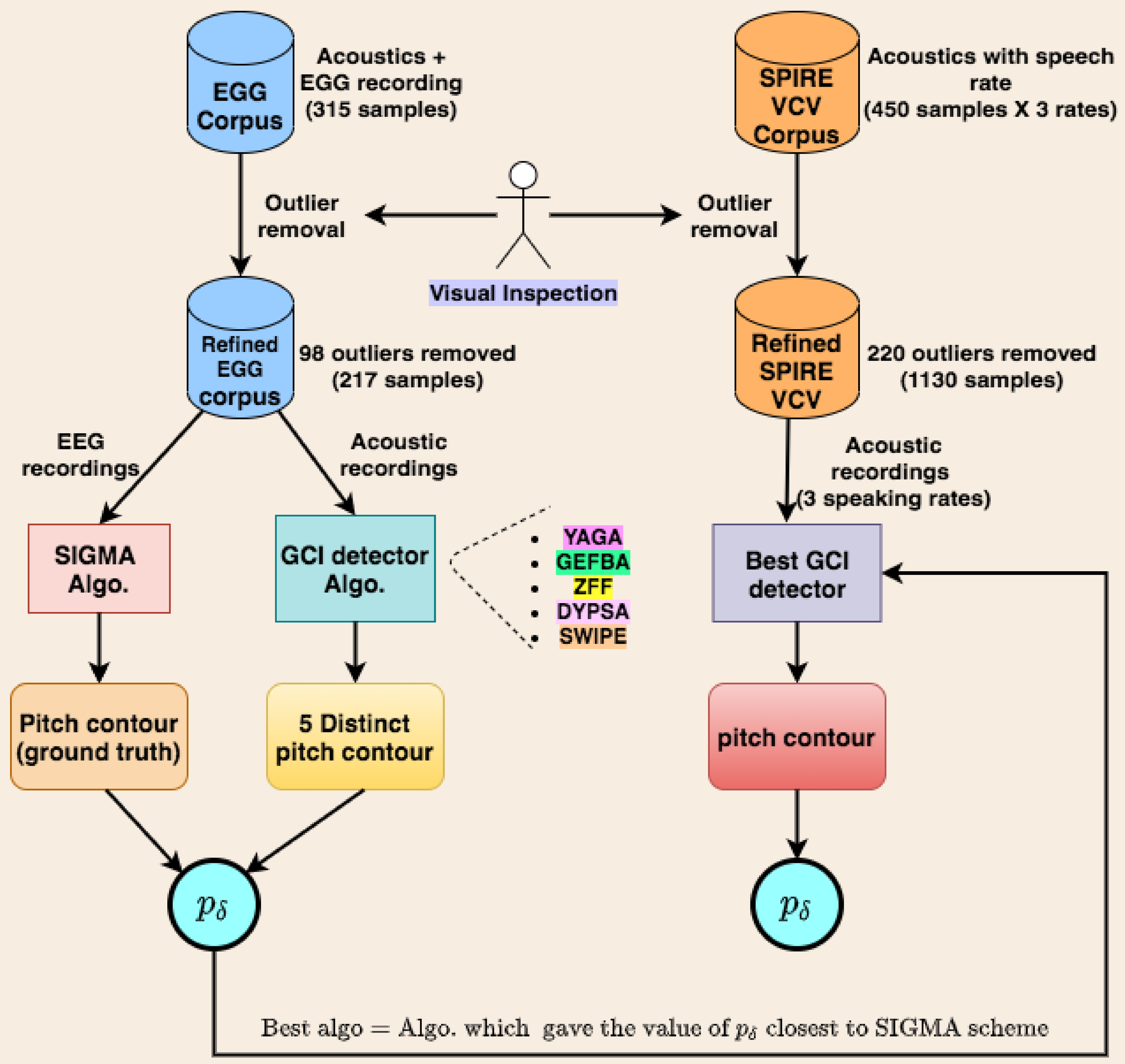
Proposed method

- Compute the pitch contour from acoustic signal using a pitch estimation technique.
- Compute the measure of SFI as a percentage of pitch drop (p_δ) in the C−region compared to the V_1−region:

$$p_\delta = \frac{(p^{med}_{V_1} - p^{min}_{C}) \times 100}{p^{med}_{V_1}}$$



Experimental Setup



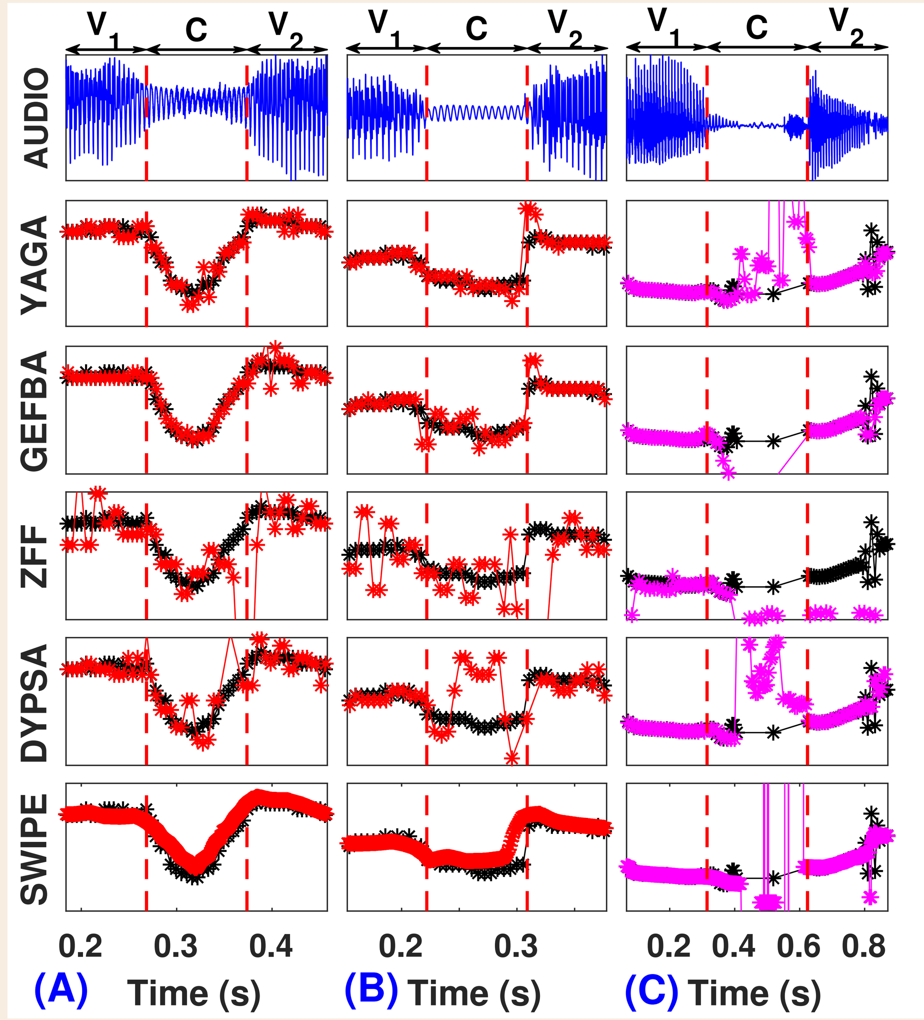
References

- Rao et al: "Effect of source filter interaction on isolated vowel-consonant-vowel perception." JASA 144.2 (2018): EL95-EL99.

Acknowledgement

We thank all the subjects for their participation in the data collection, and the **Department of Science & Technology (DST), Govt. of India** for their support !!

Result 1: The best GCI detector

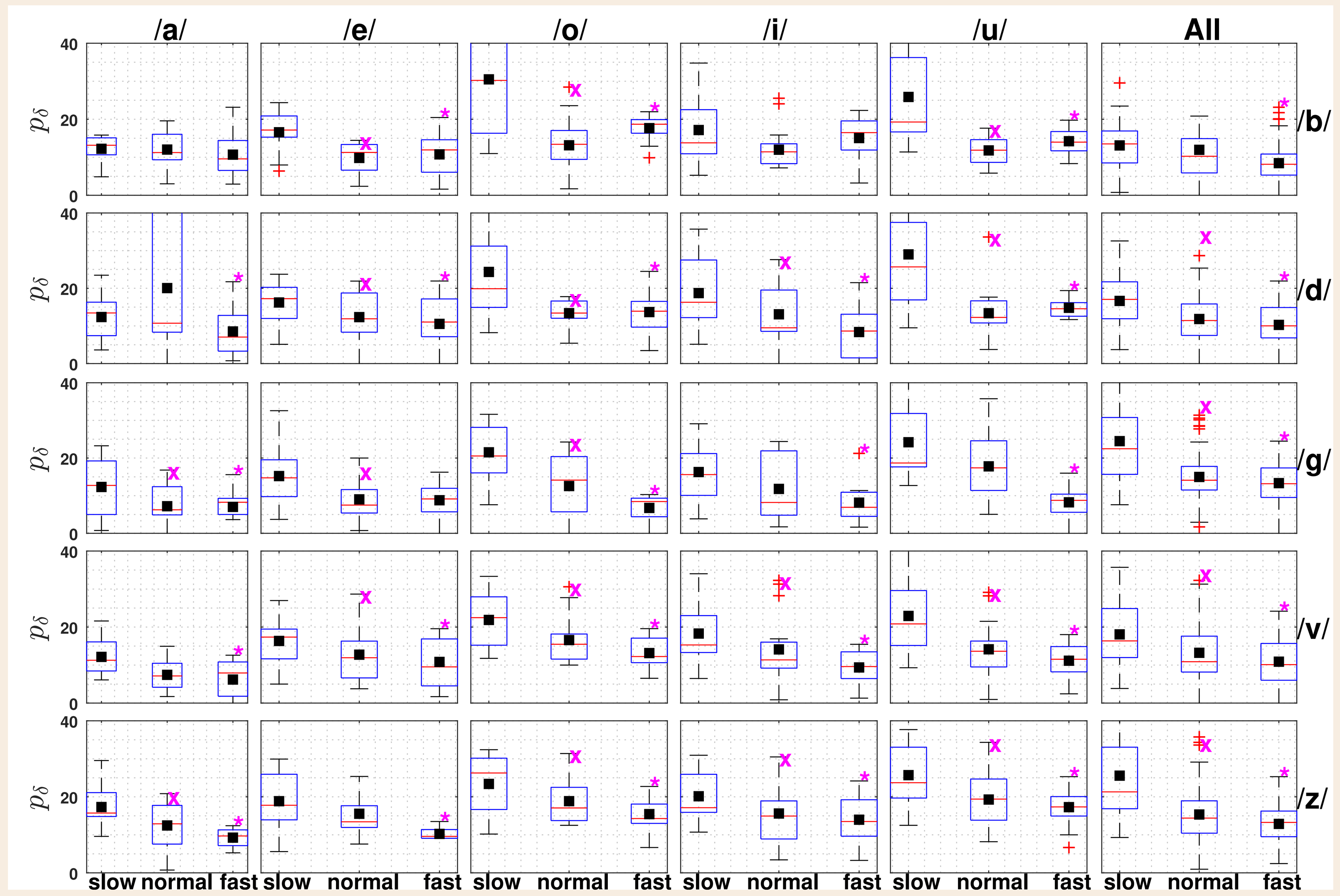


- **YAGA is found to be best GCI detector in the context of SFI study.**

	p_δ (EGG corpus)	p value
SIGMA	13.42 (11.98)	-
YAGA	12.28 (6.79)	0.30
GEFBA	12.06 (7.71)	0.25
ZFF	11.98 (7.92)	0.23
DYPSA	16.20 (9.96)	0.04
SWIPE	9.86 (9.62)	0.004

Table: The average of p_δ across all 217 considered samples in EGG corpus, (·) indicates the standard deviation.

Result 2: Impact of rate on SFI



- It is clear from the figure that the mean and median value of p_δ increase as the speaking rate reduces.

Conclusion

- Using the methodology discussed in this work, **YAGA was determined as the best GCI detector** cum pitch estimation scheme for studying SFI.
- Upon examining the effect of SFI on speaking rate, the study reveals a **significant difference in the pitch drop values between slow and fast rates, with increasing pitch drop as the speaking rate reduces**.

{tilakpurohit,achuhtr,prasantg}@iisc.ac.in