Cross-Teager Energy Cepstral Coefficients For Dysarthric Severity-Level Classification (# Paper ID:10)

Authors: Anand Therattil, Aastha Kachhi, Hemant A. Patil

Speech for Social Good (S4SG) Workshop of INTERSPEECH 2022

Dhirubhai Ambani Speech Research Lab, DA-IICT, Ganghinagar, India.
25-September-2022

Outline

- What is Dysarthria?
- Teager Energy Operator (TEO)
- Cross-Teager Energy Operator (CTEO)
- CTEO Feature Extraction
- Noise Suppression by CTECC$_{min}$
- Experimental Results
- Experimental Analysis
- Summary & Conclusions

What is Dysarthria?

- Dysarthria is a neurological condition that affects human speech.
- It affects the coordination between brain and speech production muscles.
- Few ailments which can induce dysarthria are Cerebral palsy, muscular dystrophy, and stroke.
- The impact and damage of the neurological injury determines the severity level of dysarthria.
- Severity-level of dysarthria are determined by Speech–language pathologist.

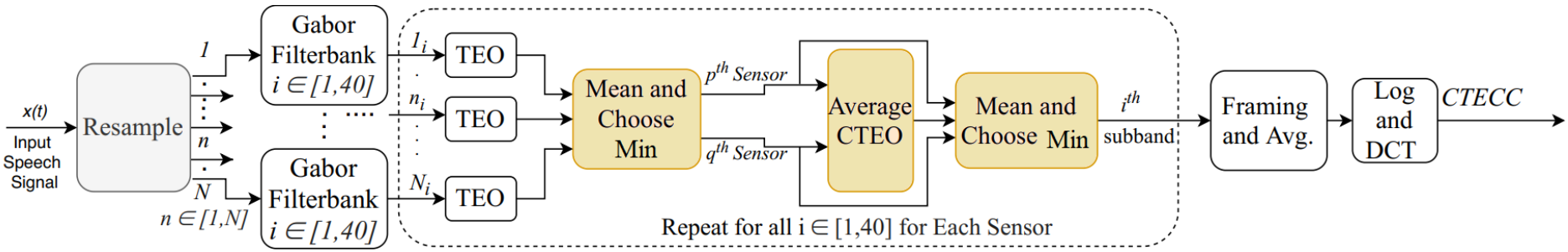
Teager Energy Operator (TEO)

- TEO is known to track the instantaneous energy of the speech signal.
- $\Psi[x(t)] = [\dot{x}(t)]^2 - x(t)\ddot{x}(t) = A^2 sin^2(\omega) \approx A^2\omega^2$
  where $\dot{x}(t)$ is $\frac{dx}{dt}$ and $\ddot{x}(t)$ is $\frac{d^2x}{dt^2}$
- $\Psi\{x(n)\} = x^2(n) - x(n+1)x(n-1)$
- TEO utilizes single channel input (i.e., single microphone as input).

Cross-Teager Energy Operator (CTEO)

- CTEO is an extension of the TEO for multi-channel input.
- CTEO estimates the relative rate of change of energies between signals.
- $\Psi_{cr}[x(t), y(t)] = (\dot{x}(t)\dot{y}(t)) - (x(t)\ddot{y}(t))$
- The Average CTEO ($\Psi_{cr}^{avg}[\cdot]$) is estimated as:
  $\Psi_{cr}^{avg}[x(t), y(t)] = \frac{1}{2}(\Psi_{cr}[x(t), y(t)] + \Psi_{cr}[y(t), x(t)])$.
- The Average CTEO in discrete domain is defined as:
  $\Psi_{cr}^{avg}\{x(n), y(n)\} = x(n)y(n) - 0.5[x(n+1)y(n-1) + x(n-1)y(n+1)]$

CTEO Feature Extraction



- Functional Block Diagram of CTECC$_{min}$ Feature Extraction.

CTEO Feature Extraction Continued

- Input signal is decomposed into 40 subband-filtered signal using Gabor filterband.
  $x_{i_j}(t) = x_i(t) * g_j(t)$
- For j^{th} subband signal Minimum Teager Energy (TE) is estimated.
- CTEO is estimated be selecting 2 channel having Minimum TE represented as: $MES =$
  $min_{(p,q)}(E\{\Psi^{avg}_{cr}[x_{p_j}(t), x_{q_j}(t)]\}, E\{\Psi_{cr}[x_{p_j}(t)]\}, E\{\Psi_{cr}[x_{q_j}(t)]\})$.
- Signal with the minimum energy is selected and converted to Cepstral Domain.

Noise Suppression by CTECC$_{min}$

- Each speech signal $xi(t)$ is represented as:
  $x_i(t) = s_i(t) + n_i(t)$, where i=1,2,...,N, represents the N-channel microphone.
- Cross-Teager Energy (CTE) is expressed as:
  $\Psi_{cr}[x_{p_j}(t), x_{q_j}(t)] = \Psi_{cr}[s_j(t)] + \Psi_{cr}[n_{p_j}(t), n_{q_j}(t)] +$
  $\Psi_{cr}[s_j(t), n_{q_j}(t)] + \Psi_{cr}[n_{p_j}(t), s_j(t)]$
- Average CTE equation represented as:
  $E\{\Psi_{cr}[x_{p_j}(t), x_{q_j}(t)]\} = E\{\Psi_{cr}[s_j(t)]\} + E\{\Psi_{cr}[n_{p_j}(t), n_{q_j}(t)]\}.$
  where $E\{\Psi_{cr}[n_{p_j}(t), n_{q_j}(t)]\} = error \approx 0$

Experimental Setup

Class-wise patient details.

	Female	**Male**	**Number of Samples**
High	F03	M01	751
Medium	F02	M07	930
Low	F04	M05	926
Very Low	F05	M09	930

- For training, we used 90% of data, which comprises of 837, 837, 833, and 676 utterances of very low, low, medium, and high severity-level.
- 10% of the data is utilized, consisting of total 354 utterances.

Experimental Results

% Classification Accuracy for Baseline STFT and CTECC Feature Set.

Feature Set	CNN
Spectrogram	91.72
CTECC_max	91.24
CTECC_min	**95.76**

Performance Evaluation for Various Feature Set

Feature Set	F1-Score	MCC	Jaccard Index	Hamming Loss
STFT	0.87	0.83	0.776	0.124
CTECC_max	0.91	0.88	0.84	0.087
CTECC_min	**0.96**	**0.94**	**0.91**	**0.042**

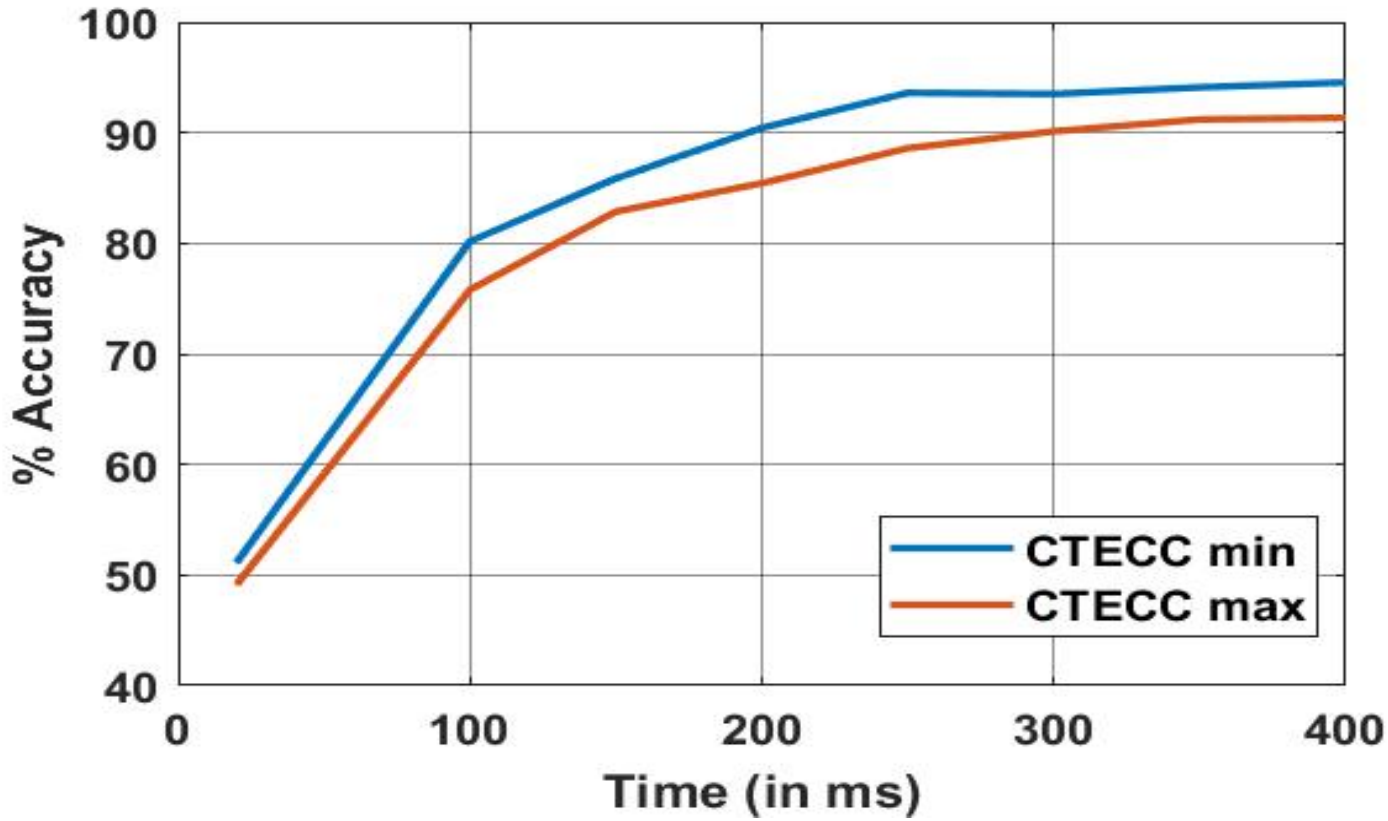
- CTECC$_{min}$ outperforms the baseline STFT and CTECC$_{max}$ feature sets.

Experimental Analysis

Confusion Matrix for STFT, and CTECC Feature Set

Feature Set	Severity	High	Medium	Low	Very Low
STFT	**High**	63	6	3	3
	Medium	10	79	3	1
	Low	3	4	79	7
	Very Low	1	2	1	89
CTECC (Max)	**High**	62	10	2	1
	Medium	4	85	1	1
	Low	1	3	88	1
	Very Low	1	4	2	86
CTECC (Min)	**High**	**70**	3	2	0
	Medium	3	**90**	0	0
	Low	1	3	**87**	2
	Very Low	0	1	0	**92**

Analysis of Latency Period



- Analysis of Latency Period for CTECC min and CTECC max.

Summary and Conclusions

- The discriminative power of CTECC$_{min}$ in dysarthric severity-level classification.
- The extraction of the CTECC$_{min}$ features is computationally expensive.
- CTECC$_{min}$ captures the linguistic information more effectively.
- In future, CTECC$_{min}$ will be futher validated using TORGO and Homeservice corpus.

Acknowledgements

- The organizers of UA Speech Corpus for making UA-Speech corpus publicly available.
- Ministry of Electronics and Information Technology (MeitY), India for sponsoring Speech Technologies in Indian Languages' under 'National Language Translation Mission (NLTM): BHASHINI',
- The consortium leaders Prof. Hema A. Murthy, Prof. S. Umesh (IIT Madraas), and the authorities of DA-IICT Gandhinagar, India.

Selected References

- J.F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Albuquerque, NM, USA, 1990.
- S. Gupta, *et al.*, "Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments," Neural Networks 2021.
- J. F. Kaiser, "Some useful properties of Teager's energy operators," in 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Minneapolis, MN, USA,1993.
- R. Acharya, H. Kotta, A. T. Patil, and H. A. Patil, "Cross-Teager energy cepstral coefficients for replay spoof detection on voice assistants," in ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Ontario, Canada, 2021.
- P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with application to speech analysis," IEEE Transactions on Signal Processing, 1993.

Thank You