Perceptually Enhanced Single Frequency Filtering for Dysarthric Speech Detection and Intelligibility Assessment

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Krishna Gurugubelli, and Anil Kumar Vuppala
2019 IEEE International Conference on Acoustics, Speech and Signal Processing, Brighton, UK

ICASSP 2019

Objectives

1. Instantaneous spectral representation using single frequency filtering.
2. Dysarthria detection and dysarthric speech intelligibility assessment.

Motivation

Dysarthria as an **abnormalities in instantaneous variations** in speech.
Averaged spectral representation over 20-30 ms.
Growing interest in **AM-FM based modulation feature** representation of speech.

Single Frequency Filter-bank

The transfer function of single frequency filter (SFF) [1] is given by,

$$H(z) = \frac{1}{1 - az^{-1}}. \quad (1)$$

The transfer function of frequency modulated SFF is given by,

$$H_k(z) = \frac{1}{1 - a_k z^{-1}}. \quad (2)$$

Here $a_k = ae^{-jw_k}$, $w_k = \frac{\tilde{w}_k * 2 * \pi}{f_s}$, and $\tilde{w}_k$ represents the k^{th} frequency component. The single frequency filter-bank can be realized as,

$$SF_{filterbank} = \left[\; H_1(w) \; H_2(w) \ldots H_k(w) \; \right]^T \quad (3)$$

Here, k=1,2,3,...M, and the k^{th} filter response and corresponding temporal envelope are given by

$$y_k[n] = \sum_{i=1}^{N} h_k[i] \; x[n-i] \quad (4)$$

$$m_k[n] = \sqrt{y_{kR}^2[n] + y_{kI}^2[n]} \quad (5)$$

Temporal envelopes are together forms the instantaneous frequency distribution, and is further processed by perceptual operations.

PE-SFCC feature extraction

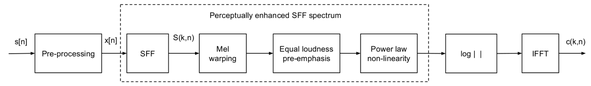

Fig. 1. PE-SFCC feature extraction framework.

SFF Spectral Sepresentation

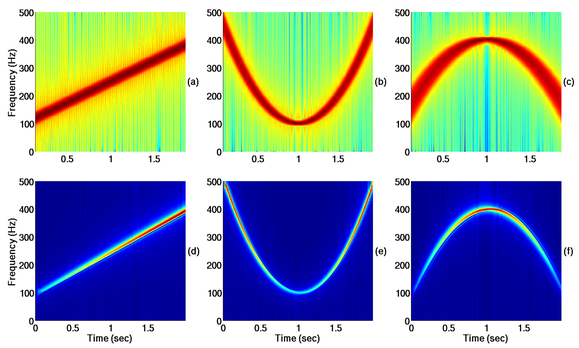

Fig. 2. Time-frequency representation of linear, quadratic, and convex chirp signals. Short time Fourier transform (Top row: (a)-(c)). Single frequency filtering (Bottom row: (d)-(f))

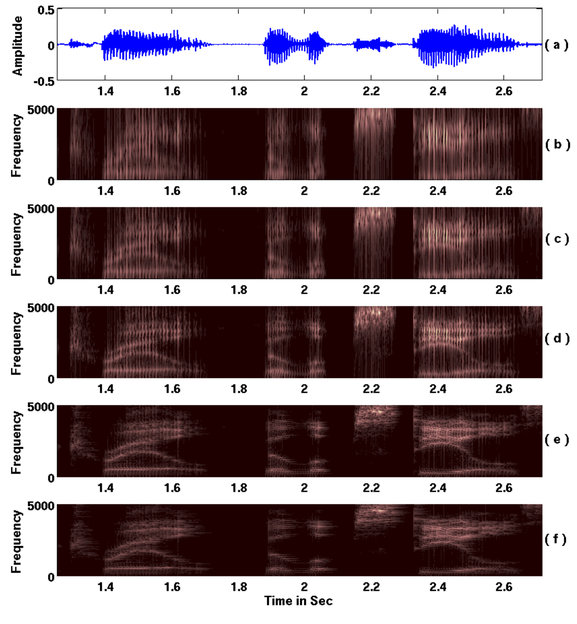

Fig. 3. Time-frequency representation of speech signal with different resolutions. (a) Speech signal. (b)-(f) SFF-TF representations of speech for a = 0.85, 0.90, 0.95, 0.985, and 0.995.

UA-Speech Database

UA-Speech [2] is a publicly available dysarthric speech database.

- 16 dysarthric and 13 healthy speakers.
- 765 isolated words (uncommon words, common words, computer commands, digits, and radio alphabets).
- 4 groups dysarthric speakers based intelligibility (very low: 0-25%, low: 25-50%, medium: 50-75% and high: 75-100%)

Experimental setup

- **List of experiments:**
  1. Dysarthric speech detection (DSD)
  2. Dysarthric speech intelligibility assessment (DSIA)
- **Base line features:** Mel frequency cepstral coefficients (**MFCC**), Perceptual linear prediction (**PLP**) cepstral features, Multi-taper MFCC, Constant-Q cepstral coefficients (**CQCC**)
- **Classifier:** i-vector with probabilistic linear discriminant analysis.
- **Validation:** Leave one speaker out validation.
- Speaker independent dysarthria assessment.
- **Pre-processing:** long silence regions are trimmed to 50ms by using SOX toolkit.

Results

Table 1. Comparison between of PE-SFCC and other state-of-the-art features in DSIA and DSD on UASPEECH database

	DSIA system Accuracy in %	DSD system Accuracy in %
PLP features	45.55	80.01
MFCC	42.07	78.70
Multi-taper MFCC	50.43	88.79
CQCC	49.14	91.38
PE-SFCC	**60.78**	**93.64**

Results contd...

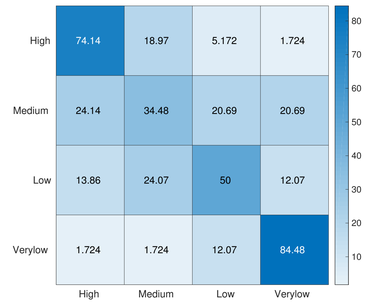

Fig. 4. Confusion matrix of DSIA system with PE-SFCC features.

- High confusion between neighboring classes.
- Poor detection accuracy of the class Medium.

Conclusions

Instantaneous spectral features works better than the conventional frame level features. Proposed PE-SFCC outperformed the state-of-the-art features.

Future Research

- Fusion of Excitation source information for dysarthria assessment.
- Detection of dysarthria type.

References

[1] G Aneeja and B Yegnanarayana. Single frequency filtering approach for discriminating speech and non-speech. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(4):705–717, 2015.

[2] Heejin Kim and et. al. Dysarthric speech database for universal access research. In *Proc. Interspeech*, 2008.

Contact Information

krishna.gugrugubelli@research.iiit.ac.in
anil.vuppala@iiit.ac.in
Speech Processing Laboratory, KCIS,LTRC
IIIT, Hyderabad, India.