Robust Wind Noise Detection and Suppression for Wearable Glass with Multi-Microphone Array

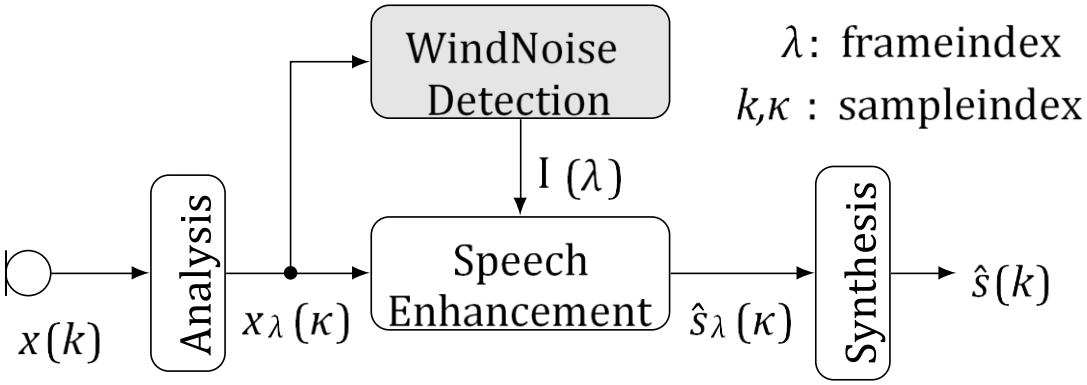
Anthony D. Rhodes
Intel Corporation

- Goals
  - Develop and explore robust, generalizable algorithms capable of real-time deployment for the following tasks:

(I) Accurate wind noise detection

(II) Wind noise suppression for ASR

Contents



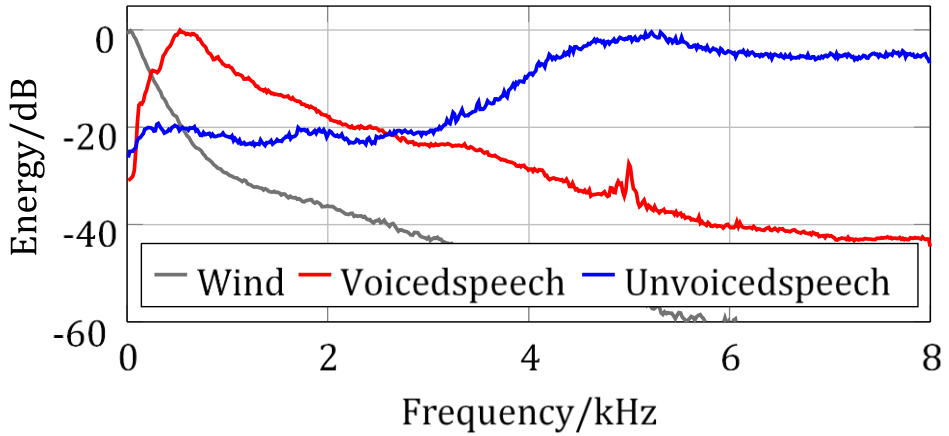
- Noisy signal x(k) for time frame λ is rendered as (DFT) $X_{\lambda}(k)$
- For task (I) we develop wind *noise detection* classifier: I(λ) for **(5) distinct classification regimes**.
- Task (II) we develop several wind *noise suppression systems* **(4) total**.

Wind Noise Detection and Reduction System

I. Wind Detection

- A sufficiently precise detection of wind noise is the first step towards suppression of noise in captured signals [1].
- We seek discriminative (preferably low-dimensional) features that can be used to accurately determine the presence of wind while the WVAD indicates that the wearer is currently speaking.
- Features for wind detection commonly rely on short-term statistics.

- In particular, the spectral energy distribution for very low frequencies (< 10 Hz) for wind is discernable from that of speech.



I. Wind Detection: (WD1) STM Method

- Following [2], we define a wind detection indicator: $I_{STM}(\lambda)$ using *short-term mean* features for time frame λ.

- Motivation: The digital representation of an acoustic signal can be assumed to be obey a *zero-mean* condition (under long-term trajectory conditions).
- The mean value of short segments can be used to detect low frequency parts in a signal.

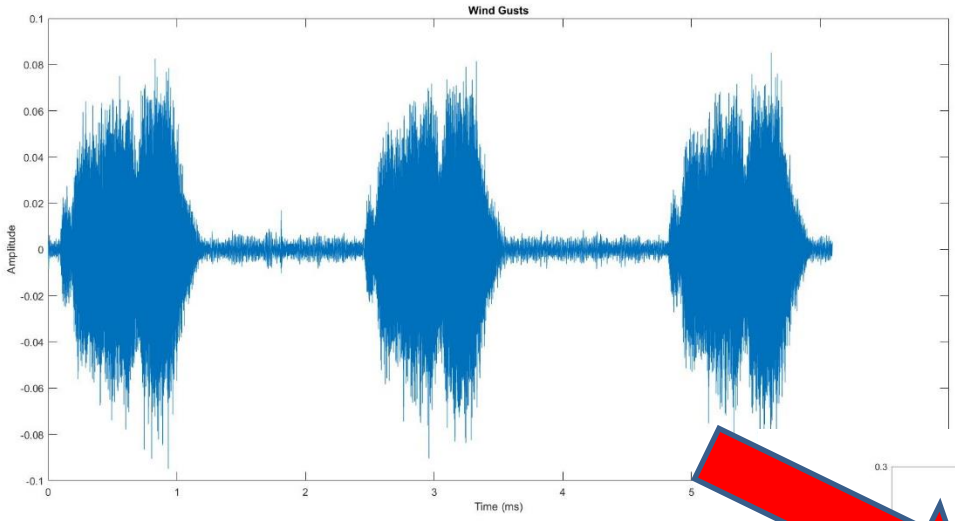
$$I_{STM}(\lambda)=\left|\frac{\sum_{\kappa=0}^{L-1} x_{\lambda}(\kappa)}{\sum_{\kappa=0}^{L-1}|x_{\lambda}(\kappa)|}\right|$$

- The normalization above generates values close to 0 for high frequency components, and close to 1 for low frequency components.

  Note that the short-term duration parameter 'L' needs to be tuned.

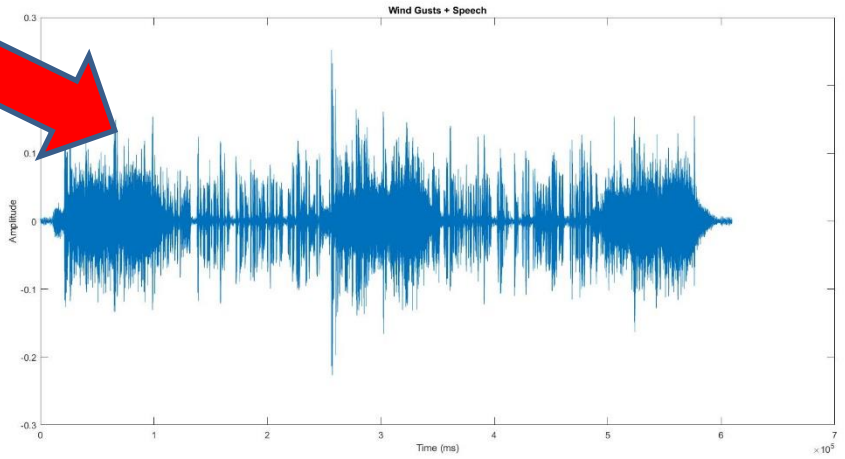
I. Wind Detection: STM Method

- *Test data*: 1 min of continuous speech (4 different speakers) + wind gusts
- *Heuristics*: L=500; threshold I_{STM} @ 0.5.



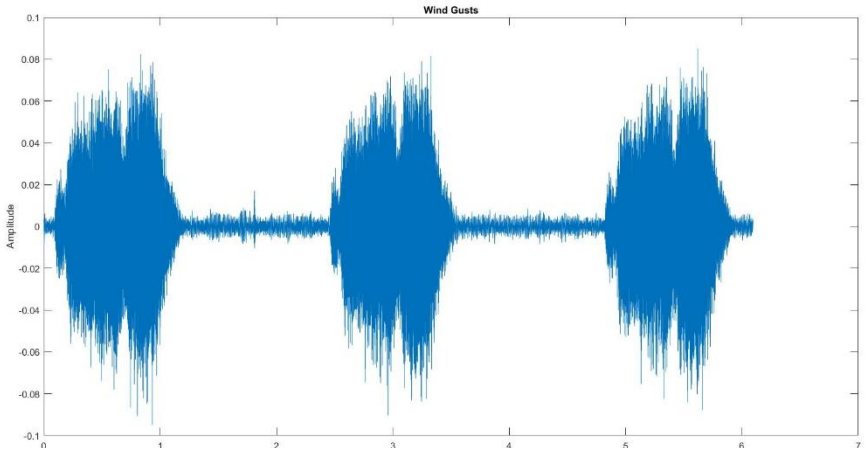
Wind Gusts

Continuous Speech + Wind Gusts

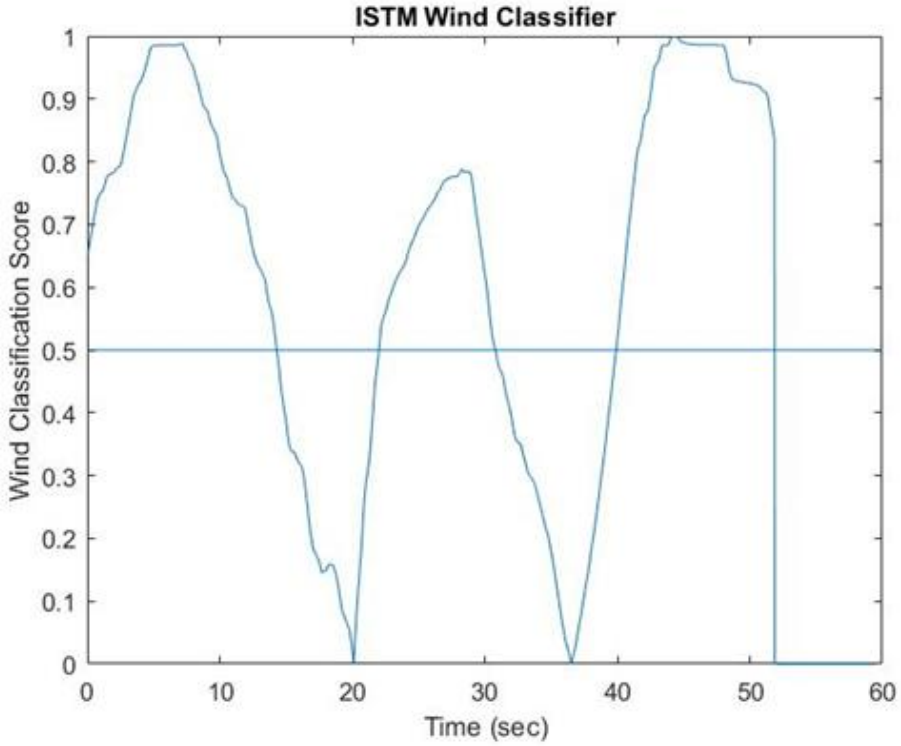


I. Wind Detection: STM Method

Wind (ground-truth)



Continuous Speech + Wind Gusts



I. Wind Detection: (WD2) SSC Method

- [3] Suggests utilizing energy distribution to build wind classifier; in particular, we consider the signal sub-band centroids (SSC) as a discriminative feature for classification.

- Define the spectral centroid for time frame λ for the bin range [$\mu 1$,$\mu 2$]:

$$\Xi_{\mu 1,\, \mu 2}(\lambda)=\frac{\sum_{\mu=\mu 1}^{\mu 2}|X(\lambda,\mu)|^2\cdot\mu}{\sum_{\mu=\mu 1}^{\mu 2}|X(\lambda,\mu)|^2}$$

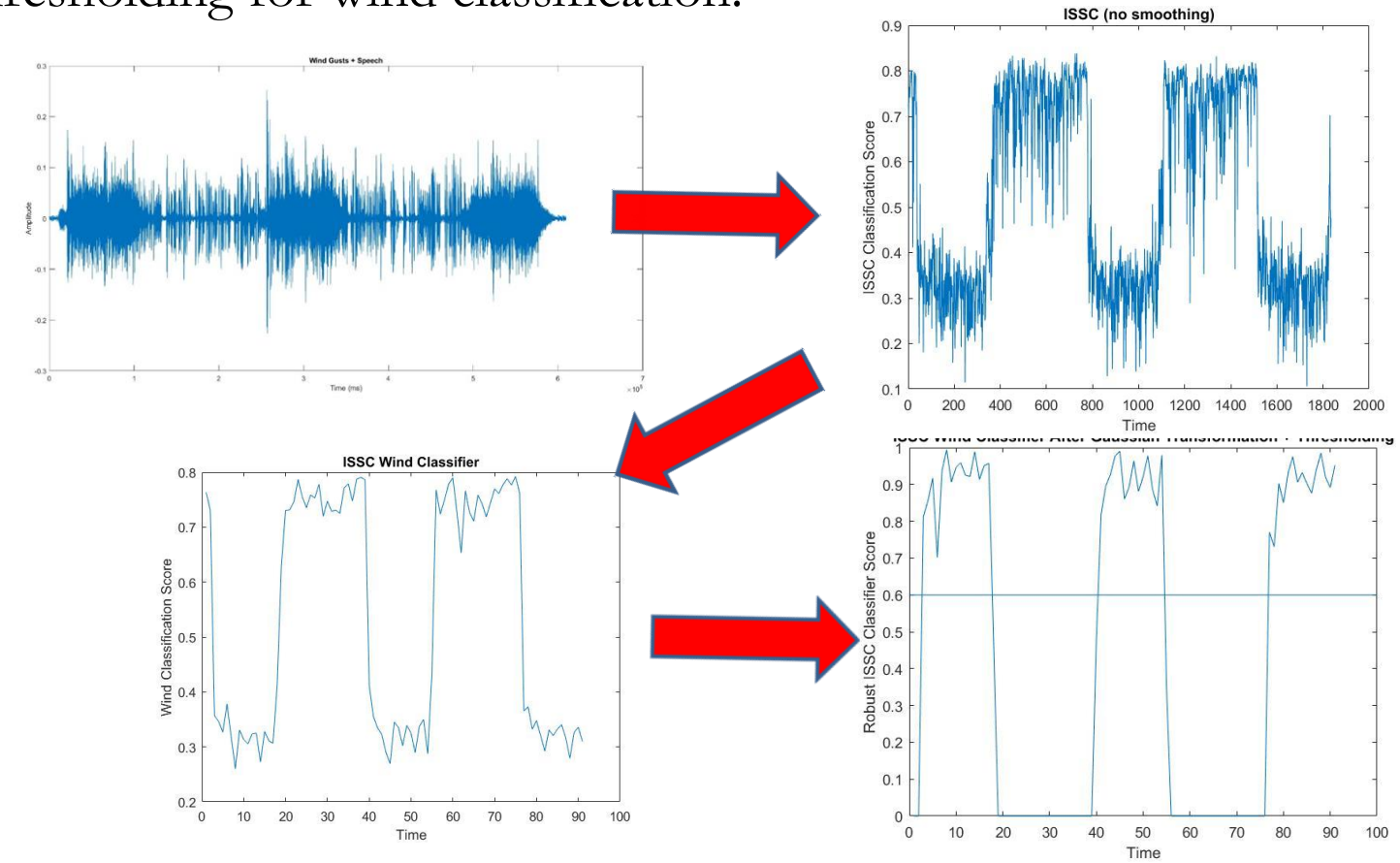
- In order to detect wind, we consider the sub-band range: [0, 10], as in [3].

- Define the SSC-based wind indicator:

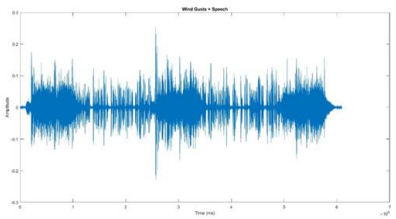
$$I_{SSC}(\lambda)=\frac{\mu 2-\Xi_{\mu 1,\, \mu 2}(\lambda)}{\mu 2}\in[0,1]$$

I. Wind Detection: (WD2) SSC Method

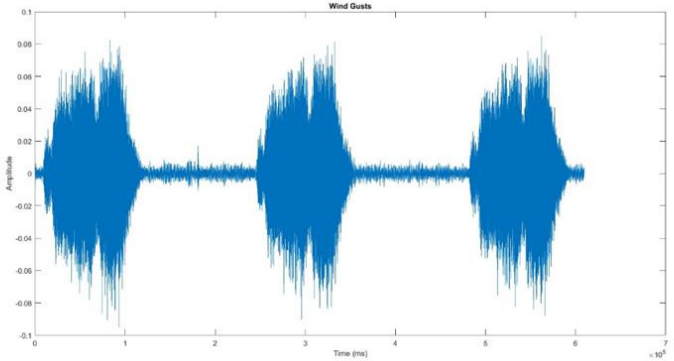
- Because of the low-dimensional spectral representation used for the SSC method, the wind indicator function tends to be very noisy.
- To generate a more robust model, we apply a smoothing procedure (500ms windows), following by a Gaussian fit to the I_{SSC} function + thresholding for wind classification.



I. Wind Detection: (WD2) SSC Method Test Results

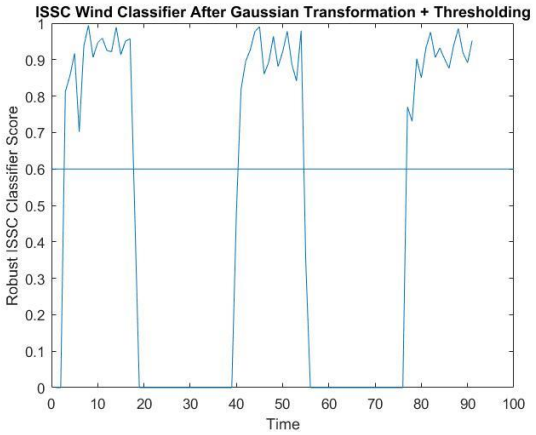


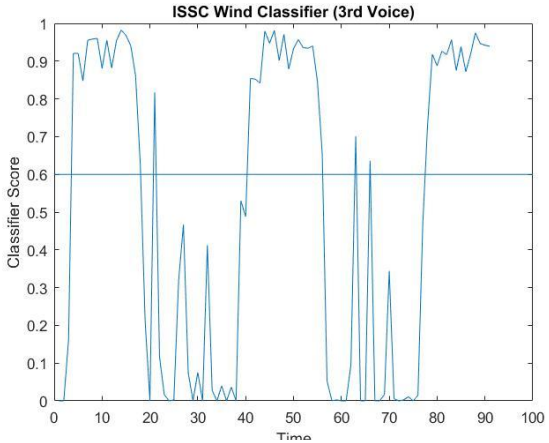
Wind + Speech



Wind (ground-truth)

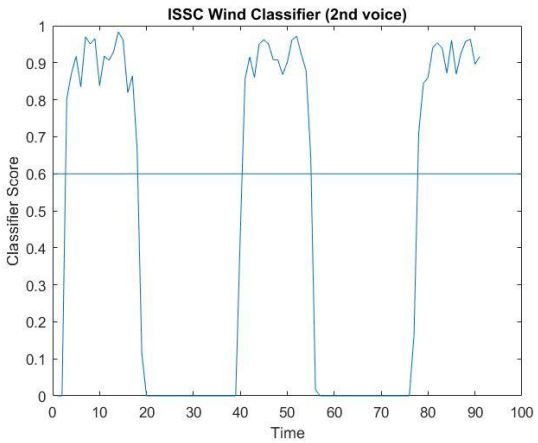
Voice 1

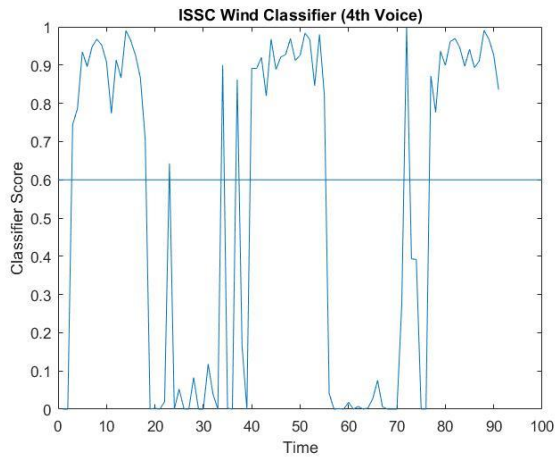




Voice 3

Voice 2





Voice 4

I. Wind Detection: (WD3) NSF

- [4] Proposes a NSF (negative slope fit) approach to wind noise classification.
- NSF relies on the assumption that the magnitude of the spectrum of wind noise can be roughly approximated by a linear decay over the frequency:

$$\dot{X}(\lambda) = a_1 \cdot \boldsymbol{\mu} + a_0 \quad \text{with frequency vector } \boldsymbol{\mu} = [0,1,\dots,N/2]^T$$

One can thus express this approximation as:

$$\dot{X}(\lambda) = M \cdot \mathbf{a} = [\mathbf{1}, \boldsymbol{\mu}] \cdot [a_0, a_1]$$

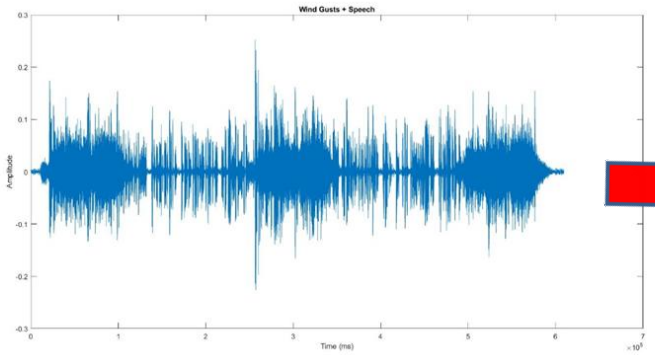
• We can compute **a** with a least-squares approach, where we minimize the error: $e(\lambda) = \|X(\lambda) - \dot{X}(\lambda)\|^2$.

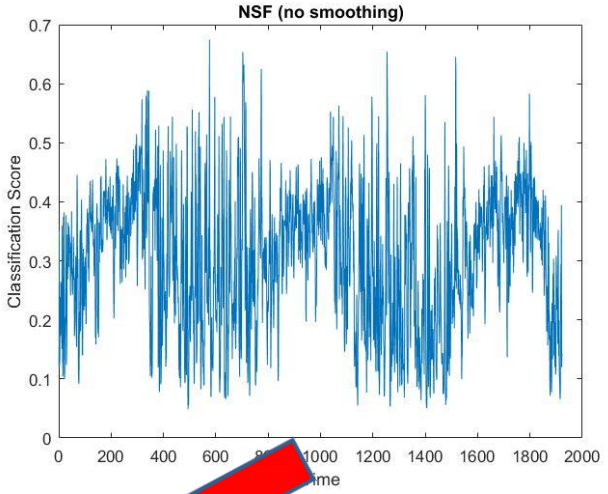
• The optimal solution is given by: $\mathbf{a}_{\text{opt}} = (M^T M)^{-1} \cdot M^T \cdot X(\lambda)$

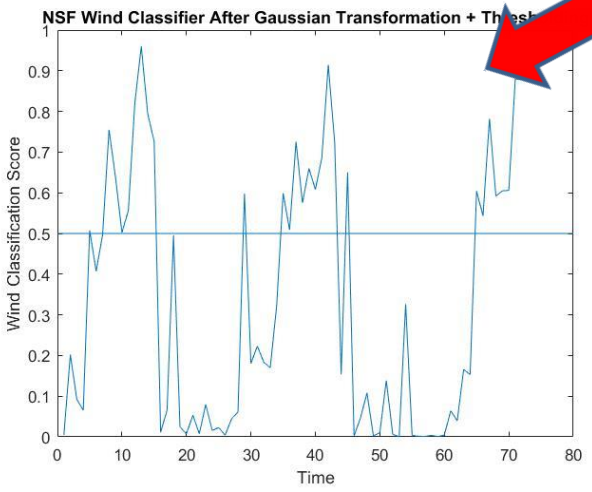
• The NSF-based indicator is defined: $I_{\text{NSF}}(\lambda) = 1 - \frac{e(\lambda)}{\|X(\lambda)\|^2}$

I. Wind Detection: (WD3) NSF Method

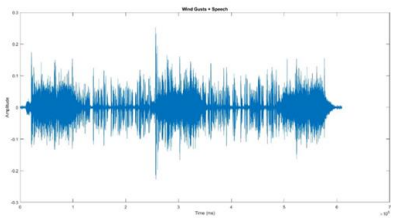
- To generate a more robust model, we apply a smoothing procedure (500ms windows), following by a Gaussian fit to the I_{NSF} function + thresholding for wind classification.



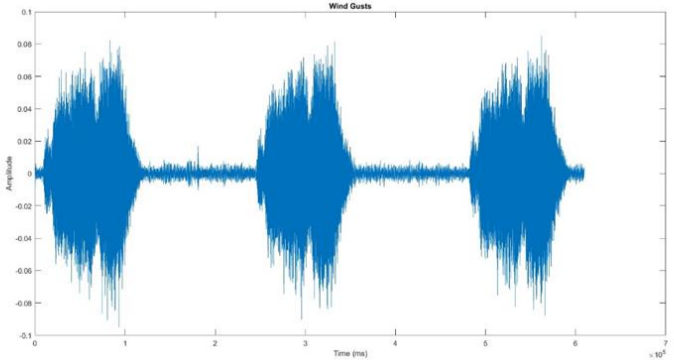




I. Wind Detection: (WD3) NSF Method Test Results

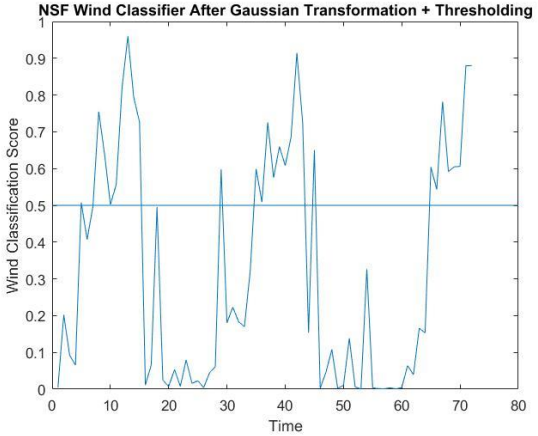


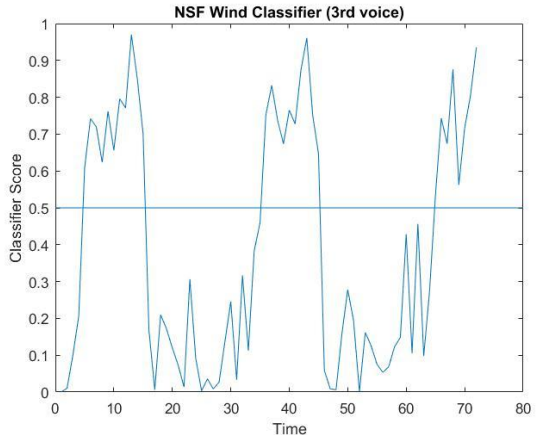
Wind + Speech



Wind (ground-truth)

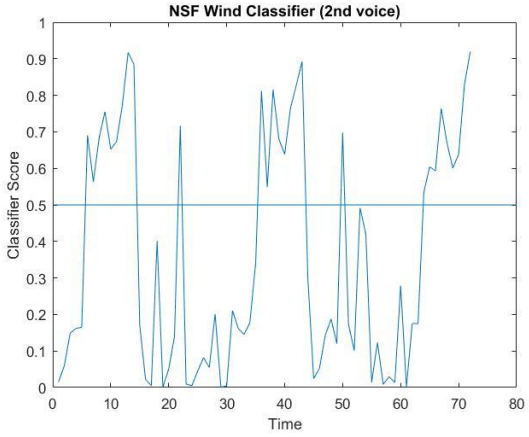
Voice 1

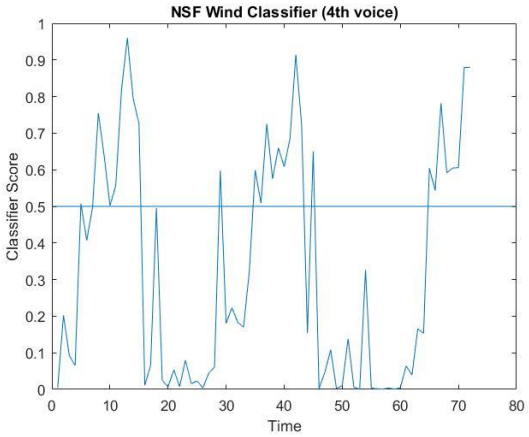




Voice 3

Voice 2





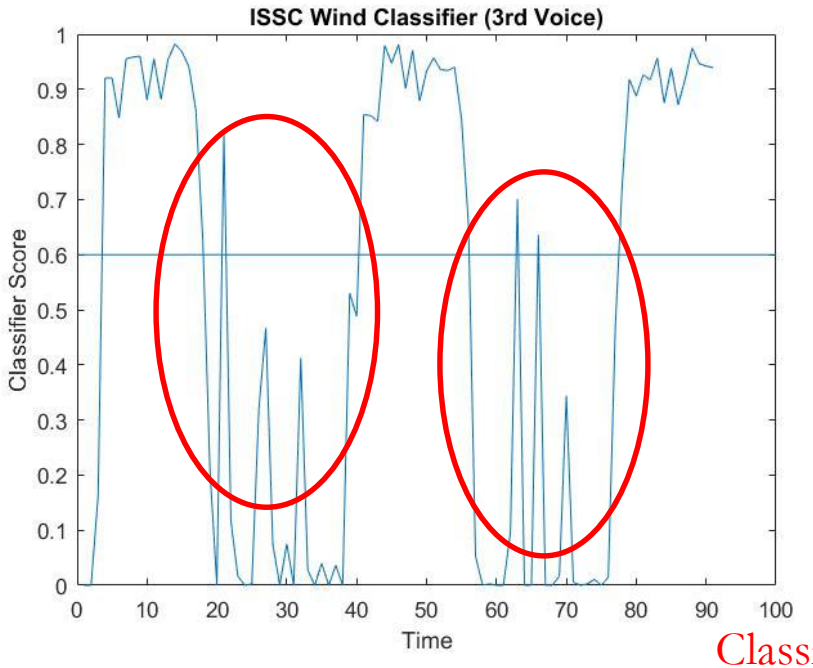
Voice 4

I. Multi-Microphone Arrays for Wind Detection

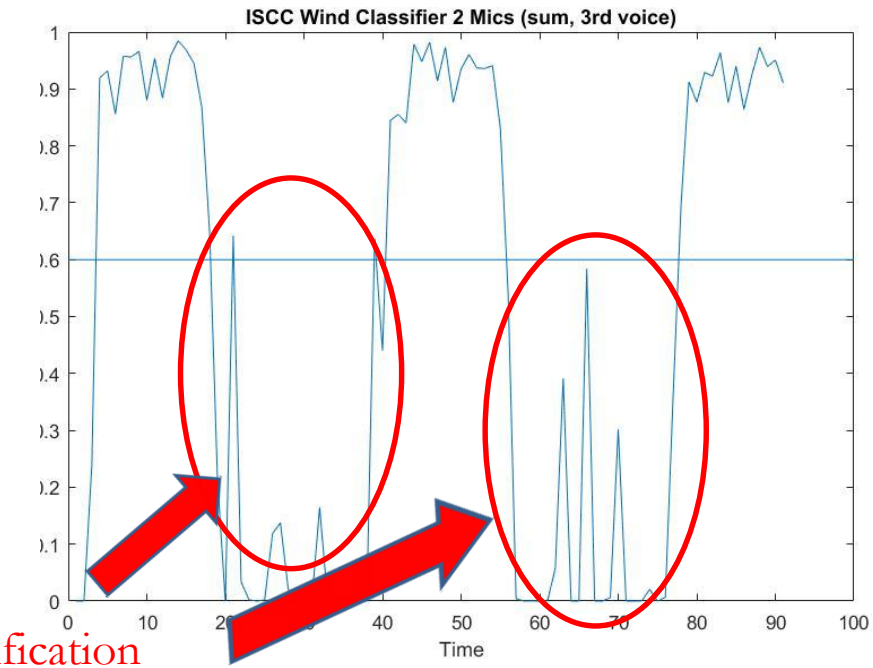
- Q: Does using multiple microphones help for wind detection?

- A: Yes! (using multiple mics does tend to make each classifier more robust).
- Results are shown for naïve 2-channel sum wind classification (more sophisticated techniques such as *delay-sum* can be used).

Single Mic



Two Mics



Classification error reduction

I. Wind Detection: (WD4) Coherence

- Multi-channel coherence features can be used to differentiate between target signal and undesirable noise; see [*] [**].

- Define the *2-channel coherence* as the ratio of the cross power spectral density (CPSD) and auto power spectral densities (APSDs)

$$\Gamma(\lambda, \mu) = \frac{\phi_{x_1 x_2}(\lambda, \mu)}{\sqrt{\phi_{x_1 x_1}(\lambda, \mu)\phi_{x_2 x_2}(\lambda, \mu)}}$$

- Where the power spectral densities (PSDs) are estimated by the recursive smoothed periodgram [*]:

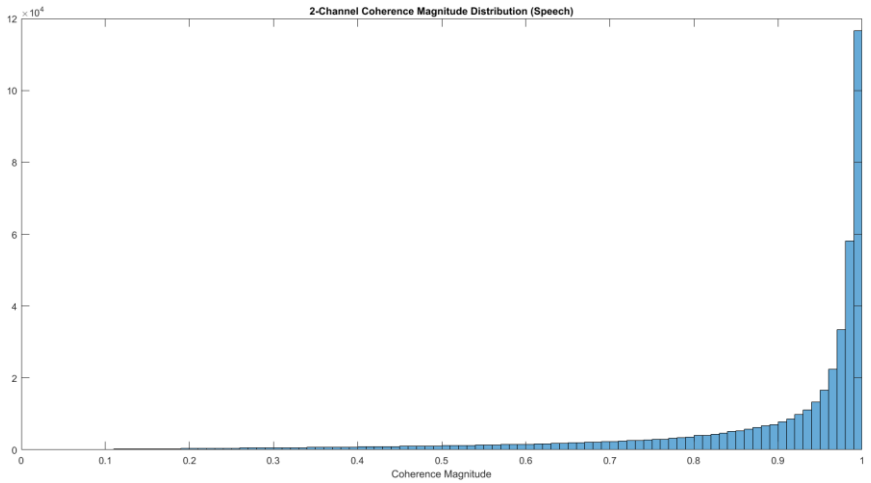
$$\phi_{x_i x_j}(\lambda, \mu) = \alpha_s \phi_{x_i x_j}(\lambda - 1, \mu) + (1 - \alpha_s) X_i(\lambda, \mu)\, X_j{}^H(\lambda, \mu)$$

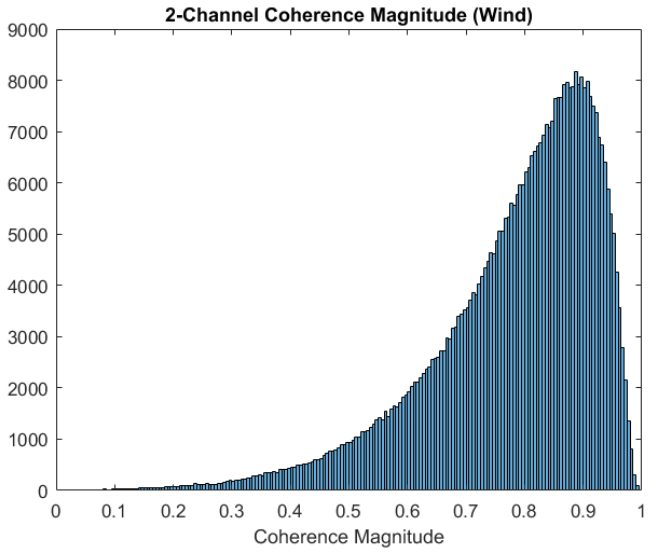
Here α is a smoothing constant set heuristically (α=0.8) and X represents the short time spectrum of the signal.

I. Wind Detection: (WD4) Coherence

- From the 2-channel coherence, we determine the magnitude (MC) and phase as features to separate speech and noise:

$$MC(\lambda, \mu) = |\Gamma(\lambda, \mu)|, \quad \varphi(\lambda, \mu) = \angle\{\Gamma(\lambda, \mu)\}$$



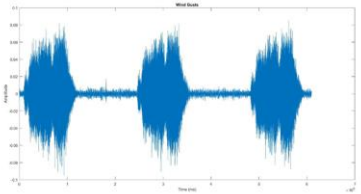


2-Channel Coherence Magnitude Histogram (Speech)

2-Channel Coherence Magnitude Histogram (Wind)

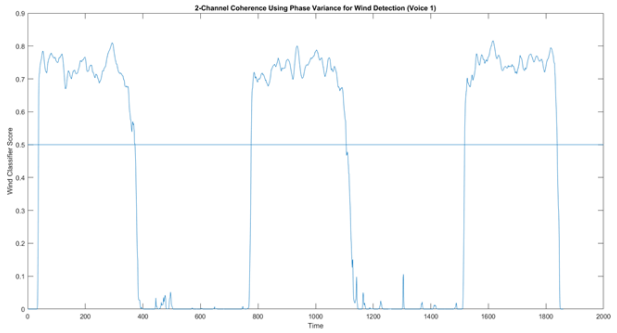
I. Wind Detection: (WD4) Coherence Test Results

- We test 2-channel coherence features (i.e. MC, phase) for wind detection using 4 distinct voices; experiments yielded (near) perfect classification.

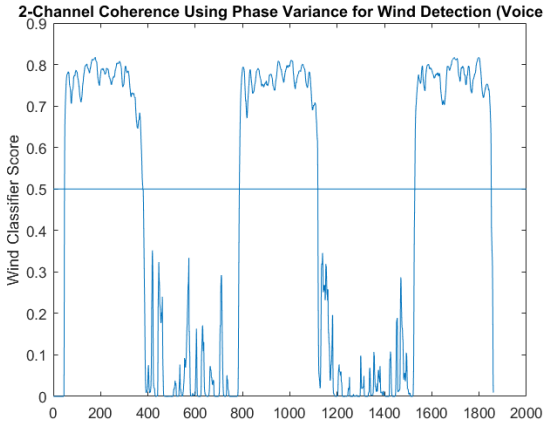


Wind (ground-truth)

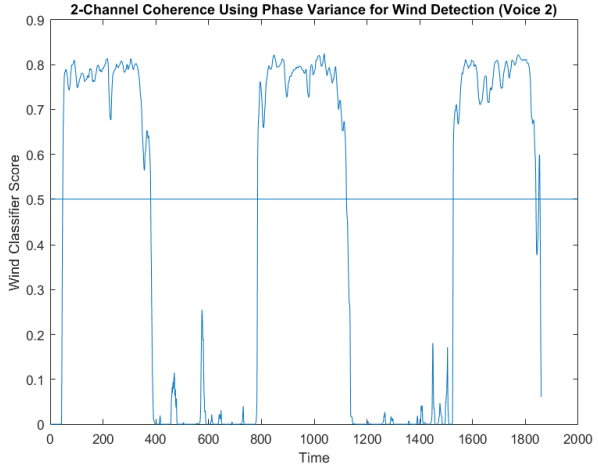
Voice 1



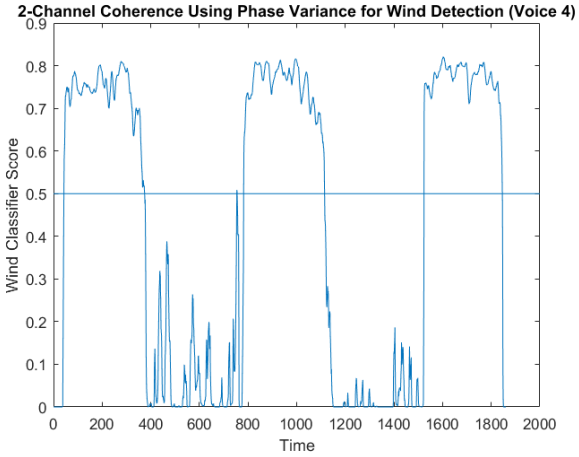
Voice 3



Voice 2



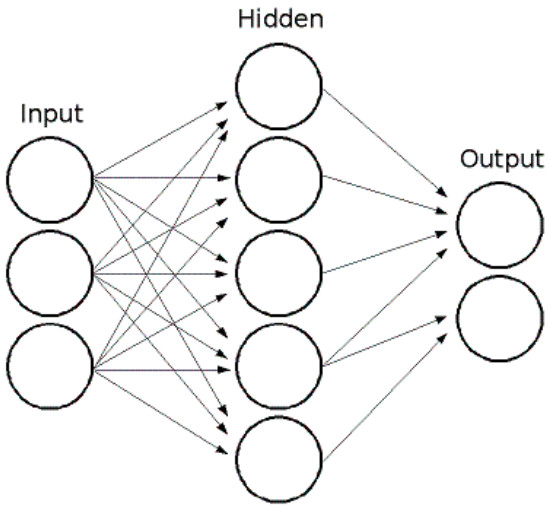
Voice 4



I. Wind Detection: (WD5) Neural Wind Noise Detection with Context-Framing

- Many contemporary, state-of-the-art results in signal processing utilize neural (and related ML) approaches. **Pros**: Capacity for strong pattern recognition and classification results in high dimensions; **Cons**: Strongly data-driven, usually requiring large (labeled) datasets.

- Despite these drawbacks, can NNs be leveraged for wind noise (in addition to other types of noise/interference) detection tasks in low-energy and "small data" experimental regimes? Yes!



I. Wind Detection: (WD5) Neural Wind Noise Detection with Context-Framing

- We develop a feed-forward NN for wind noise detection. The input to the network consists of context-expanded frames of the noisy signal appended with a noise estimate frame (see below) [*], [**], [***].

- These results, including training regime, are general enough to be applied to virtually any noise modality.

- As in [*], [**], we use the *log-power spectra* features of a noisy utterance n^u for the short-time Fourier transform. Let $N(t, f) = log|STFT(n^u)^2|$.

- Let n_t be the t^{th} frame of $N(t, f)$ and the context-expanded frame at t be represented by: $y_t = [n_{t-r}, \dots, n_t, \dots, n_{t+r}]$.

I. Wind Detection: (WD5) Neural Wind Noise Detection with Context-Framing

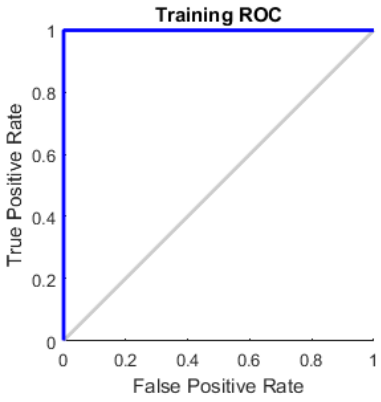
- We implement *noise-aware training* [*] by augmenting the context-expanded input to the NN with an estimate of the noise ($\hat{e}_t$) in frame n_t: where $\hat{e}_t = \frac{1}{F}\sum_{t=1}^{F} n_t$; note that $\hat{e}_t$ is estimated when the WVAD module indicates that the wearer is not speaking.

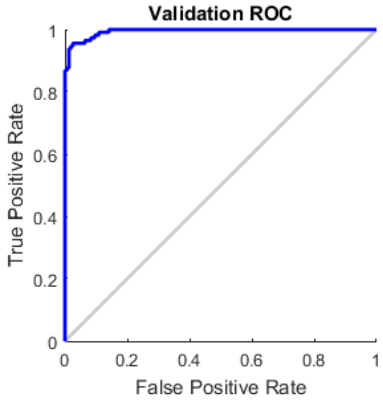
- Finally, the input to the NN is defined: $y'_t = [n_{t-r}, \ldots, n_t, \ldots, n_{t+r}, \hat{e}_t]$.

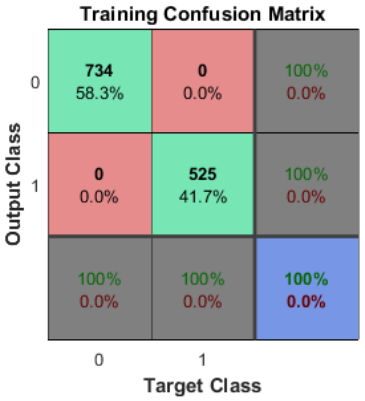
- We train with "context-horizon" parameter r=2, a "shallow" NN (for low-energy applications) with 50 hidden nodes, using conjugate gradient backpropagation; note that only 1 min of noisy speech audio was used for training.

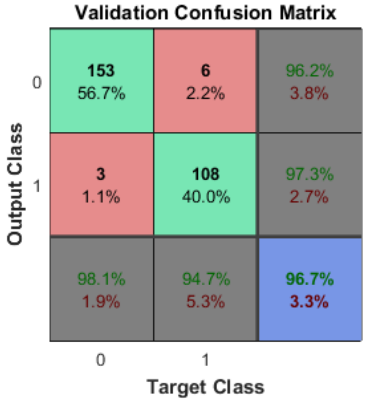
I. Wind Detection: (WD5) Neural Wind Noise Detection with Context-Framing

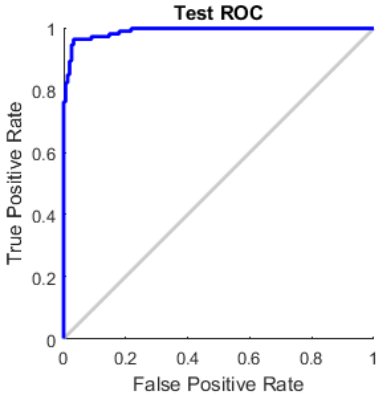
- ROC and confusion matrices for NN trained with context-framed and (approximate) noise-aware data for wind noise detection.

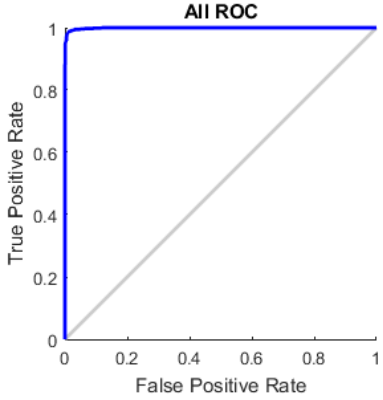


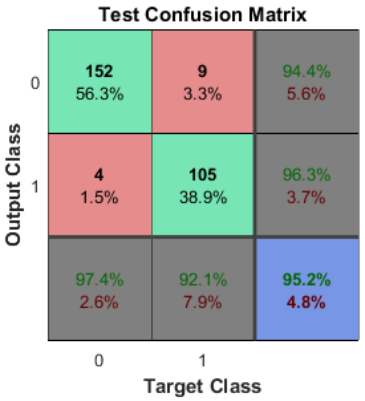


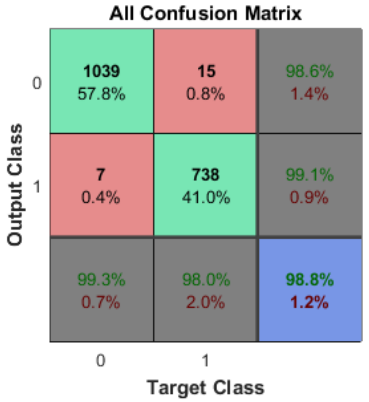








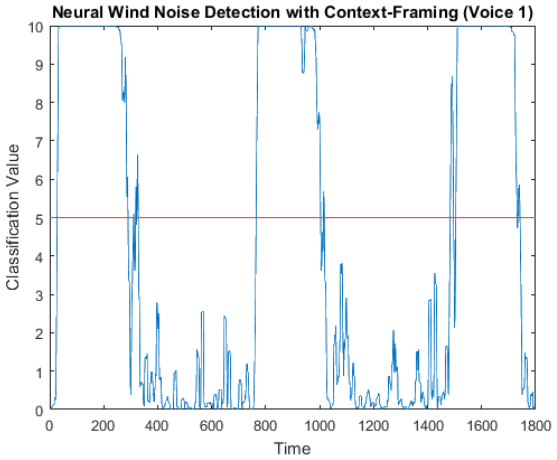


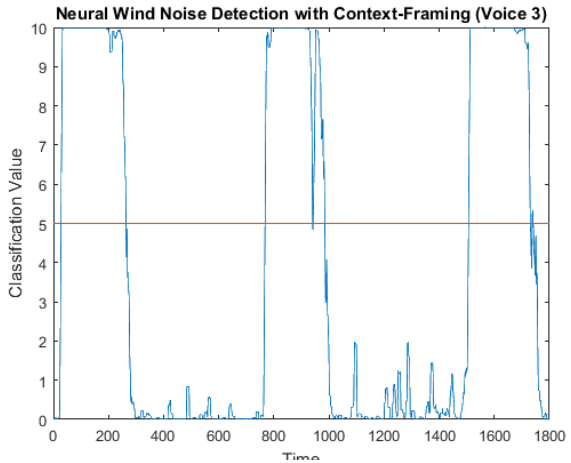


I. Wind Detection: (WD5) Neural Wind Noise Detection with Context-Framing

We test our neural wind noise detection with context-framing procedure using 4 distinct voices. Experiments yielded essentially perfect classification – even for low-energy amenable configurations; 200ms frame-smoothing is applied for optimal classification results.

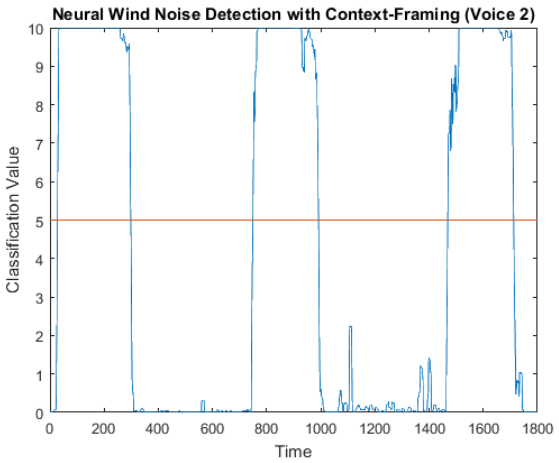
Voice 1

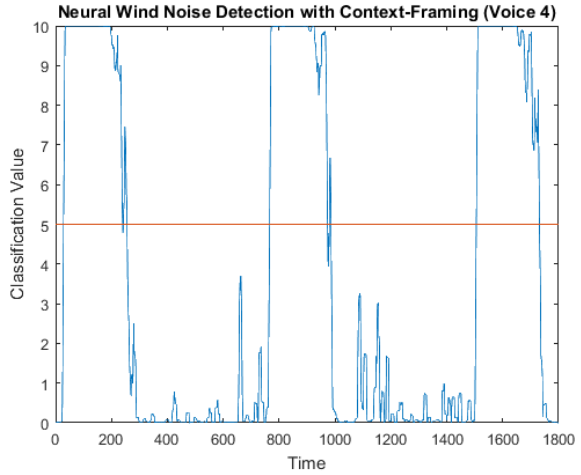




Voice 3

Voice 2





Voice 4

I. Summary of Wind Noise Detection

- *STM* method uses normalized, short-term sample trajectories; used to detect low frequencies.
- **Pros**: Renders smooth, low variance indicator function; **Cons**: tuning of sample trajectory length 'L' possibly required for each source; medium computational overhead.

- *SSC* method computes low-frequency centroids.
- **Pros**: Generates low dimensional, discriminative features; consistent results across multiple speakers; **Cons**: due to low dimensionality, indicator function can be very noisy; stronger results requiring a smoothing procedure, introducing a time-delay in classification.

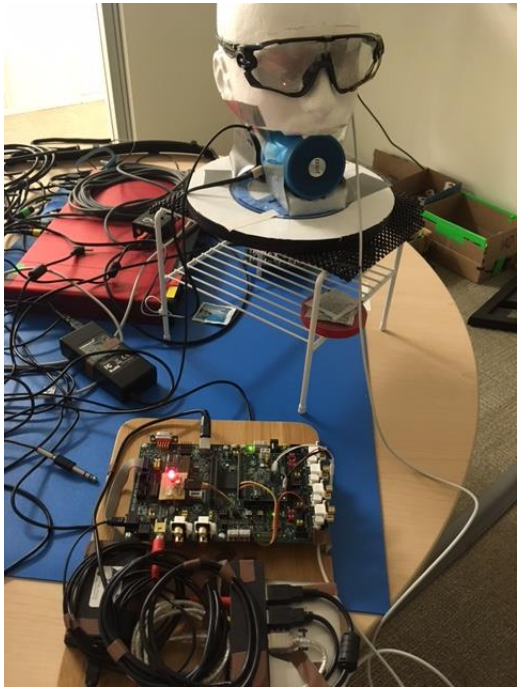
- *NSF* method depends on assumption that wind noise can be approximated as linear decay wrt frequency.
- **Pros**: Consistent results across multiple speakers; **Cons**: indicator function can be noisy, requiring smoothing; medium computational overhead.

I. Summary of Wind Noise Detection (cont'd)

- *Coherence* method uses coherence-based features in conjunction with a recursive, smoothed periodogram estimate of the power spectral densities.
- **Pros**: Takes advantage of multi-channel data; coherence-based features yield (at least two) useful mechanisms to differentiate wind/speech, including magnitude of coherence and phase coherence; requires little to no ostensible training. **Cons**: Requires recursive periodogram estimate (or similar estimate) of spectral densities for reliable results; coherence based features require heuristic – often sensitive – parameter tuning.
- *Neural Wind Noise Detection with Context-Framing* uses a NN with "context" plus noise-aware data frames.
- **Pros**: Model generally appears to require small to moderate amounts of data; test classification performance was strong (95%). **Cons**: Method requires training – for full generalizability, possibly with various "types" of wind/speech. Noise-aware training requires a consistent estimate of noise, which is often difficult with non-stationary noise (e.g. wind).

Toward Robust, Real-Time Wind Detection in Low Power, Multi-Microphone Array Environments

- We develop and implement a novel *real-time wind detection algorithm* (RTWD) suitable for general low power, multi-microphone devices that requires *no training*.
- We test this algorithm using a wind simulator under challenging detection criteria (>= 6mph wind), including a low tolerance for false-positive readings.



RTWD Algorithm

- Together, we combine the use of low-dimensional SSC features with coherence-based features in a multi-channel environment for accurate, real-time wind detection.
- SSC features are desirable for their ability to differentiate wind from voice in a low-computation regime. However, the low-dimensionality comes at a cost, as it can lead to a noisy classification – particularly in the case of mild gusts of wind (e.g. ~ 6mph and less).
- We attempt to improve the robustness of SSC-base classification in general by imposing a two-step smoothing/data transformation.
- In addition, to assist with wind classification in low intensity regimes, we incorporate coherence-based features that effectively reduce the instance oversensitivity to false-positive readings.

RTWD Algorithm (cont'd)

$$\Xi_{\mu 1,\ \mu 2}(\lambda) = \frac{\sum_{\mu=\mu 1}^{\mu 2} |X(\lambda,\mu)|^2 \cdot \mu}{\sum_{\mu=\mu 1}^{\mu 2} |X(\lambda,\mu)|^2}$$

$$I_{SSC}(\lambda) = \frac{\mu 2 - \Xi_{\mu 1,\ \mu 2}(\lambda)}{\mu 2} \in [0,1]$$

C1

C2

2-Channel Audio Input $X(\lambda, \mu)$

FFT

SSC Features

$I_{SSC(C1)}$

$I_{SSC(C2)}$

Smoothing/Data Transformation

Wind Classification

$$\Gamma(\lambda,\mu) = \frac{\phi_{x_1x_2}(\lambda,\mu)}{\sqrt{\phi_{x_1x_1}(\lambda,\mu)\phi_{x_2x_2}(\lambda,\mu)}}$$

$$\phi_{x_ix_j}(\lambda,\mu) = \alpha_s \phi_{x_ix_j}(\lambda-1,\mu) + (1-\alpha_s) X_i(\lambda,\mu)\ X_j{}^H(\lambda,\mu)$$

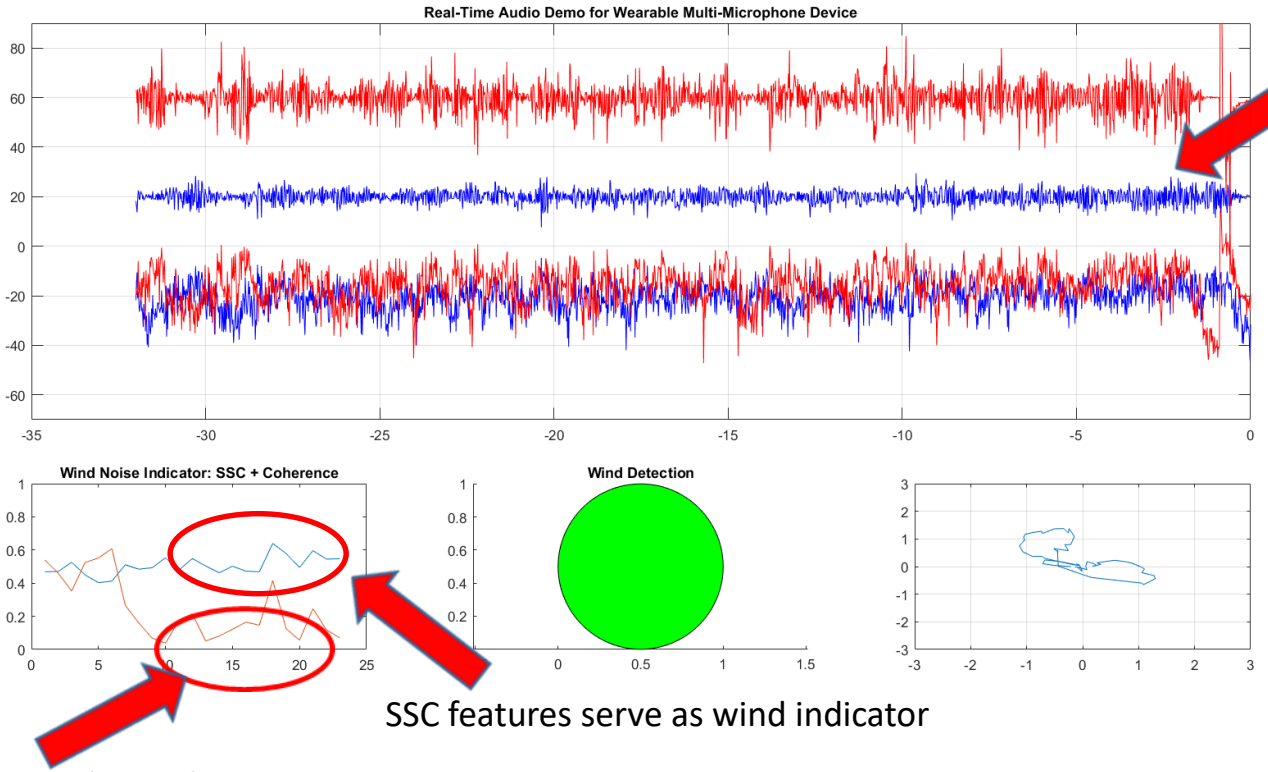
$$MC(\lambda,\mu) = |\Gamma(\lambda,\mu)|$$

Coherence Features

$MC_{(C1,C2)}$

Smoothing

RTWD (cont'd)



Real-Time 2-Channel Audio Stream

SSC features serve as wind indicator

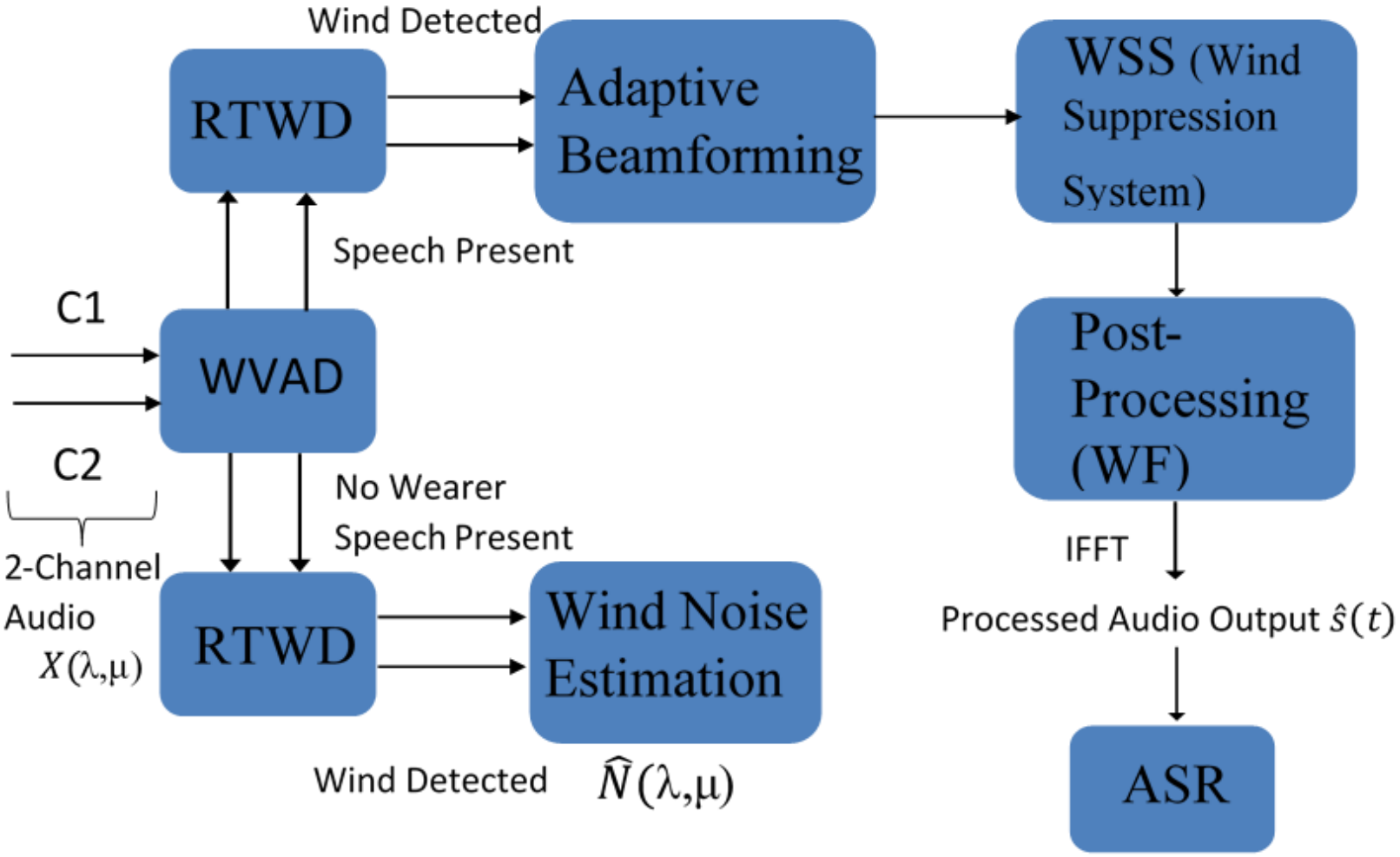
Coherence features indicate degree to which audio stream is "Speech-like"

- RTWD algorithm runs in real-time on low power, multi-channel wearable device.

- Classification performance is strong in presence of low wind intensity (<= 6 mph), roughly 90% correct classification (including false-positives).

- In mild to strong wind intensity regimes (9 mph+) classification was nearly perfect.

Proposed Wind Noise Detection and Suppression System



II. Wind Noise Suppression: Regularized Low-Rank Adaptive Beamforming

- We apply regularized, low-rank adaptive *beamforming* to enhance the noisy multi-channel signal prior to implementation of the WSS.

- The beamformer output is a complex weight vector **w**; we thus compute the enhanced signal estimate as $y(k) = w^H x(k)$ (where H denotes the Hermitian transform); see [*], [**], [*text1], [*text2].

- When the steering vector **a**(θ) is known exactly, which we assume throughout, the optimal weight vector **w** can be obtained by maximizing the beamformer output SNR given as:

- SNR = $\frac{E\left[\left|w^H s\right|^2\right]}{E\left[\left|w^H n\right|^2\right]} = \frac{\sigma_s^2 |w^H \mathbf{a}(\theta)|^2}{w^H R_n w}$, with R_n the (estimated) noise covariance matrix and σ_s^2 the desired signal power.

II. Wind Noise Suppression: Regularized Low-Rank Adaptive Beamforming

- The MVDR beamformer is obtained by minimizing the denominator of the SNR expression, leading to the corresponding optimization problem:

$$min_w \; w^H R_n w \;\; \text{s.t. } w^H \mathbf{a}(\theta)=1$$

- For robust performance, we apply regularization to the objective function above. This yields the following solution with penalty weight γ:

$$w = \left(\hat{R} + \gamma I\right)^{-1} \mathbf{a}(\theta)$$

- We use a (SVD-based) rank-1 estimate of $\hat{R}$ with channel coherence features for improved performance and stability.

- The algorithm is "adaptive" insofar as $\hat{R}$ is computed over non-speaking audio frames based on the WVAD response.

II. (WSS1): Spectral Subtraction

- *Spectral subtraction* is based on the notion that enhanced speech can be generated by subtracting elements of the (estimated) noise spectrum from that of the input noisy signal [*], [**], [***].

- Naively, an approximation of the clean speech in the frequency-domain can be modeled as: $\hat{X}_k(t,f) = Y_k(t,f) - \hat{N}_k(t,f)$.

- A common problem encountered with spectral subtraction for speech enhancement is the presence of broadband noise ("musical noise") in the processed audio.

- To reduce unwanted noise, [Boll] introduced a modification to spectral subtraction with scaling factor (α) and spectral floor (β) parameters.

II. (WSS1): Spectral Subtraction

- Define the estimate of the original speech signal:

$$\left|\hat{X}(t,f)\right|^2 = \begin{cases} \left|Y(t,f)\right|^2 - \alpha\left|\hat{N}(t,f)\right|^2, \ if \left|Y(t,f)\right|^2 - \alpha\left|\hat{N}(t,f)\right|^2 > \beta\left|\hat{N}(t,f)\right|^2 \\ \beta\left|\hat{N}(t,f)\right|^2, otherwise \end{cases}$$

- Above, the harshness of the subtraction is varied according to the parameter α, while the introduction of the spectral floor prohibits the subtraction operation from producing a signal below a reasonable threshold (and thereby diminishes the presence of musical noise).

- Spectral subtraction can also be framed as a filtration procedure: $\left|\hat{X}_k(t,f)\right|^2 = H(t,f)|Y_k(t,f)|$.

- We experiment with an adaptive Weiner filter, where: $H_{WF} = \dfrac{E\left[\left|Y(t,f)\right|^2\right] - E\left[\left|\hat{N}(t,f)\right|^2\right]}{E\left[\left|Y(t,f)\right|^2\right]}$

II. (WSS2): Non-Negative Sparse Coding

- Non-negative sparse coding (NNSC) can be used to build a *speaker independent* noise suppression system; see: [*], [**].

- The key assumption with NNSC is that the signal and noise are **additive** in the time domain (i.e. x(t)=s(t)+n(t)). As with previous methods, we represent the signal in the as an element-wise exponentiated short time Fourier transform: X= $|STFT(\mathrm{x(t)})|^{\gamma}$.

- We aim to factorize the signal matrix as X≈DH, for non-negative matrices *D* and *H* – referred to as the *dictionary* and *code* matrices, respectively.

II. (WSS2): Non-Negative Sparse Coding

- The columns of the dictionary matrix comprise a source-specific basis and the sparse code matrix consists of corresponding weights.

- It has been shown [*] that imposing non-negative constraints leads to a parts-based representation, because only additive combinations are allowed. Sparsity, while generally supported by biological and low-energy systems [*], also tends to encourage model robustness by encouraging specificity in the dictionary.

- Under the aforementioned additive assumption, we have:

$$X = Xs + Xn \approx [D_s\ Dn] \begin{bmatrix} H_s \\ H_n \end{bmatrix} = DH$$

II. (WSS2): Non-Negative Sparse Coding

- [*] Propose precomputing the source dictionary for wind noise, and to learn the dictionary of the speech directly from the noisy data. This results in a method that is speaker independent.
- The iterative update rules for modified NNSC method with precomputed noise dictionary (D_n) are as follows:

$$H_s \leftarrow H_s \cdot \frac{\bar{D}_s^T X}{\bar{D}_s^T \bar{D} H + l_s}, H_n \leftarrow H_n \cdot \frac{\bar{D}_n^T X}{\bar{D}_n^T \bar{D} H + l_n}$$

$$D_s \leftarrow \bar{D}_s \cdot \frac{X H_s^T + \bar{D}_s \cdot \left(1\left(\bar{D} H H_s^T \cdot \bar{D}_s\right)\right)}{\bar{D} H H_s^T \bar{D}_s \cdot \left(1\left(X H_s^T \cdot \bar{D}_s\right)\right)}$$

- Where above, the bar operation represents a column-wise normalized matrix; 1 is a square matrix of all ones; dot and division operations are performed pointwise; l_s and l_n are signal and noise sparsity parameters, respectively, which we tune heuristically.
- Finally, the estimate for the clean speech signal is: $\hat{X}_s = \bar{D}_s H_s$.

II. (WSS3): Subspace Methods: SVD/GSVD

Noise Reduction for ASR: ANWSS Algorithm

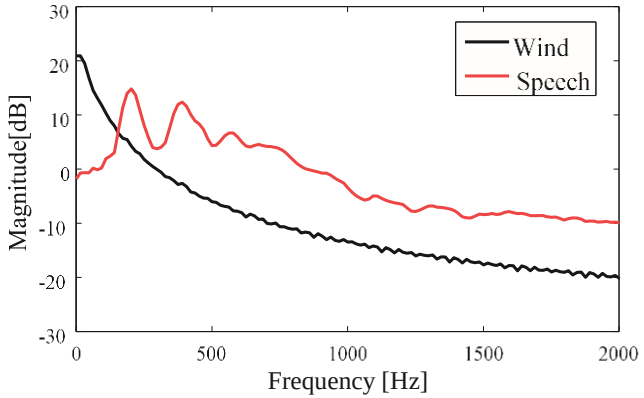
- We devise a novel wind suppression algorithm: *ANWSS* (attentive neural wind suppression system) for use with low-computation, multiple-microphone devices.

- To begin, our algorithm **"attends"** to the frequency range(s) of a noisy speech signal that are most adversely affected by wind noise.

- This spectral identification can either be done in: (1) an *a priori* fashion (i.e. no learning from data is involved) or (2) *a posteriori*, where a noise approximation is first made (using the classification/detection algorithm described previously), and then relevant frequency bins are identified according to a separate feature/spectral analysis.

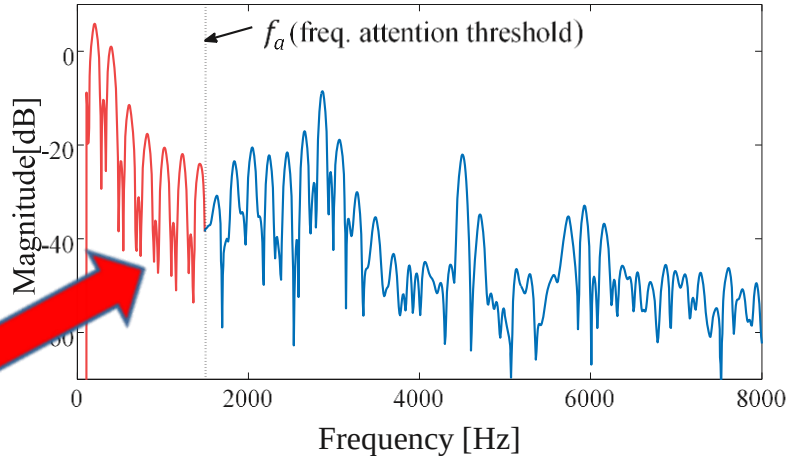
ANWSS Algorithm (cont'd)

- In the current algorithm iteration, we rely on the prior knowledge that wind commonly "overwhelms" speech [see Nelke *et al.*] in the extreme lower frequencies.



- We accordingly direct the ANWSS algorithm toward learning a neural model that reconstructs corrupted speech in the *attentive spectral region*.

Reconstructed Signal



ANWSS Algorithm (cont'd)

- This approach bears several general advantages for the noise reduction task: (1) the model can be learned with a relatively small amount of data; (2) the data representation is low-dimensional; (3) generally, the speech signal remains uncorrupted.

- We develop a shallow, low-dimensional, feed-forward NN for wind noise suppression. The input to the network consists of context-expanded frames of the noisy signal appended with a noise estimate frame (see below) [*], [**], [***].

- As in [*], [**], we use the *log-power spectra* features of a noisy utterance n^u for the short-time Fourier transform. Let $N(t, f) = log|STFT(n^u)^2|$.

- Let n_t be the t^{th} frame of $N(t, f)$ and the context-expanded frame at t be represented by: $y_t = [n_{t-r}, \dots, n_t, \dots, n_{t+r}]$.

ANWSS Algorithm (cont'd)

- Let n_t be the t^{th} frame of $\mathrm{N}(t,f)$ and the context-expanded frame at t be represented by. The input to the NN is the vector: $y_t = \left[n_{t-r}^{(1)}, \ldots, n_t^{(1)}, \ldots, n_{t+r}^{(1)}, n_{t-r}^{(2)}, \ldots, n_{t+r}^{(2)}\right]$; where the parameter r represents the "context-horizon", and the superscripts here indicate the channel identification.

- We train with r=3, a "shallow" NN (for low-energy applications) with 150 hidden nodes, using conjugate gradient backpropagation; note that only 1 min of noisy speech audio was used for training.

- The reconstructed signal is obtained by applying the following "inverse" operation sequence to the output of the NN: $\hat{s} = exp(Y(t,f)/2) \cdot exp\big(i\angle N(t,f)\big)$, where Y(t,f) is the NN output.

ANWSS Algorithm (cont'd)

Noise Corrupted Audio

C1

C2

2-Channel Audio Input $X(\lambda, \mu)$

FFT

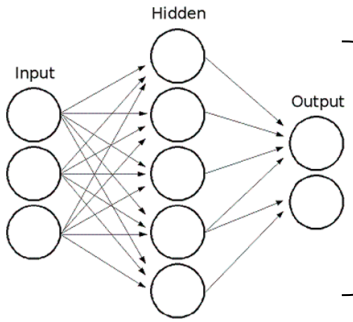
Log-Power Spectra Features

$N(t,f)_1$

$N(t,f)_2$

Context-Aware Data Processing

$y_t = \left[n_{t-r}^{(1)}, \ldots, n_t^{(1)}, \ldots, n_{t+r}^{(1)}, n_{t-r}^{(2)}, \ldots, n_{t+r}^{(2)}\right]$



Shallow feed-forward NN

$Y(t,f)$

Inverse Log-Power Spectra Transformation

$\hat{s} = exp(Y(t,f)/2) \cdot exp\left(i\angle N(t,f)\right)$

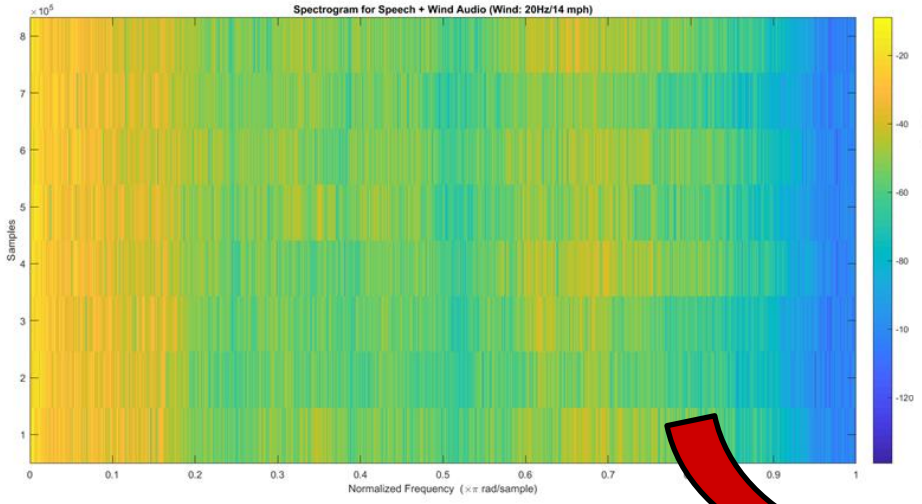
IFFT

$N(t,f) = log|STFT(n^u)^2|$

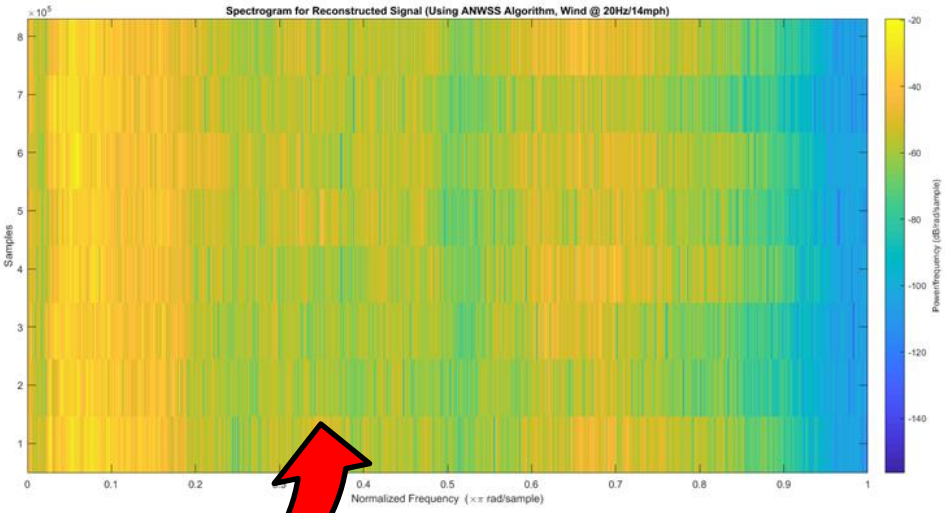
ANWSS Algorithm (cont'd)

- The spectrogram analysis (wind @ 20Hz) for noisy and subsequently reconstructed signals is shown.

- The reconstruction clearly mitigates the low frequency dominance (i.e. the wind noise) present in the original corrupted signal.

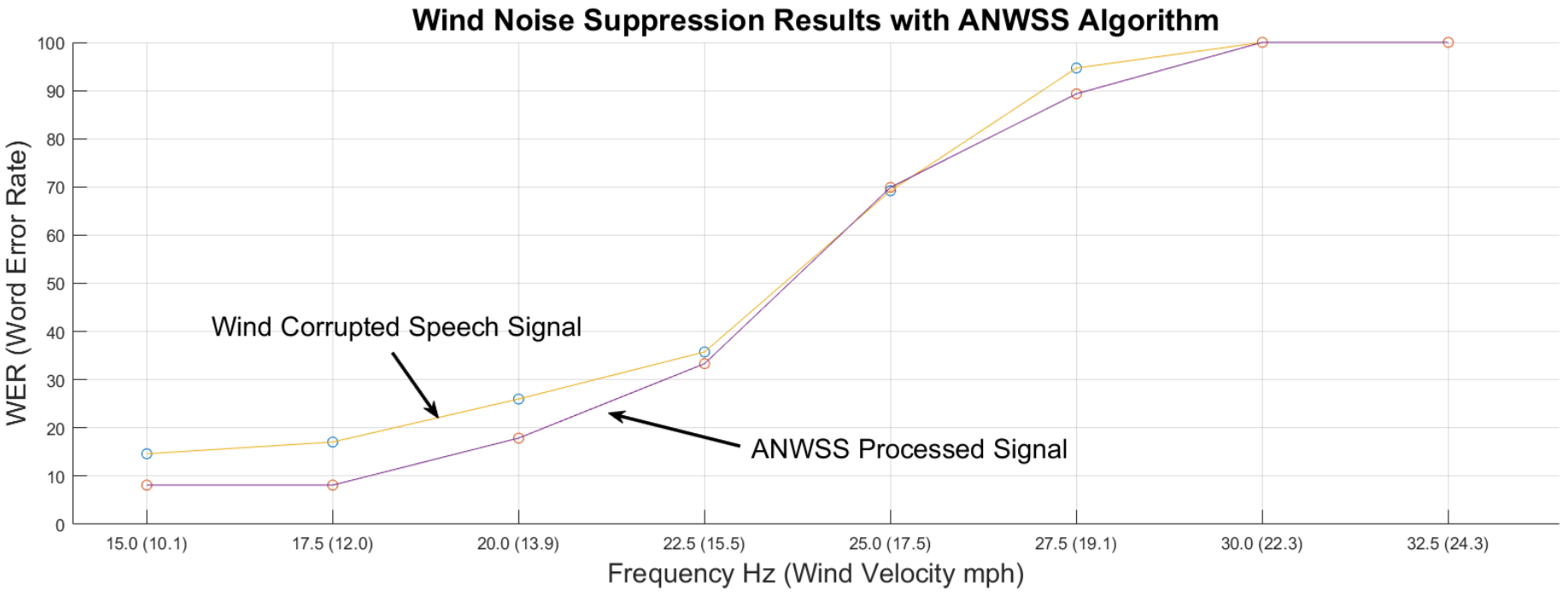


Spectrogram of Noisy Signal

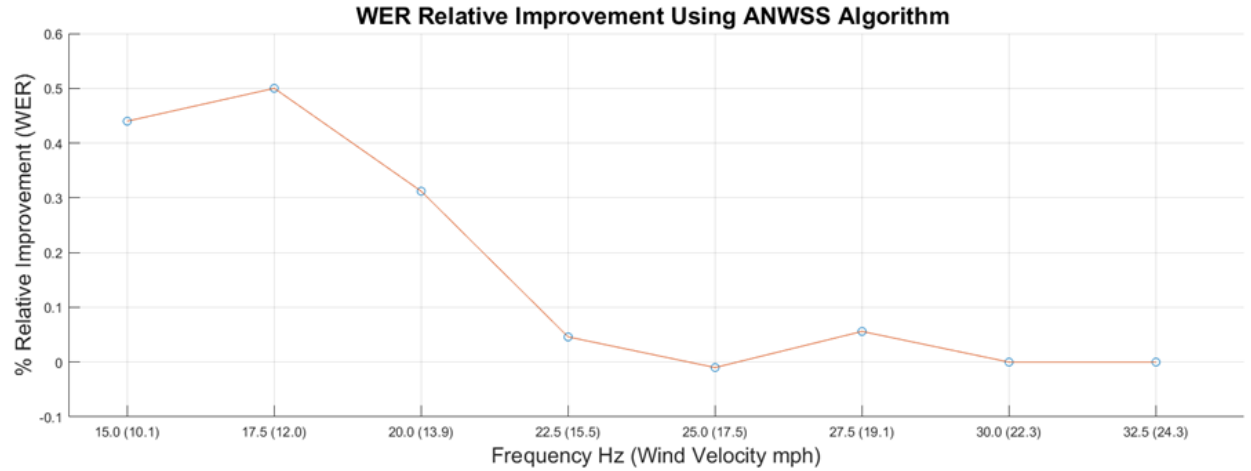


Spectrogram of Reconstructed Signal Using ANWSS Algorithm

ANWSS Algorithm (cont'd)



ANWSS Algorithm (cont'd)



- WER for ASR were significantly reduced using the ANWSS algorithm, showing the considerable potential of this method.
- In particular, the algorithm performs very well in moderate to strong wind regimes for which ASR degradation is most precipitous.
- Although accurate ASR in severe wind conditions (25 mph+) may be generally unfeasible, the ANWSS-based reconstructed audio under these extreme conditions is nonetheless still commonly comprehensible to a human listener, indicating the potential further utility of ANWSS as a noise suppression method for human-to-human audio communications.

Conclusions/Future Work Directions (cont'd)

- We successfully developed a novel, robust and competitive system for wind noise detection and suppression for a multi-microphone, wearable device under challenging, low-power constraints.

- Our detection system (RTWD) uses two sets of low-dimensional features (SSC and coherence) for robust wind detection requiring no training.

- For wind noise suppression, we advance an innovative procedure (ANWSS) that "attends" to frequency regions in which wind interference is dominant, and reconstructs the corrupted signal in these bins using a context-aware, shallow NN.

Conclusions/Future Work Directions

Potential Future Directions:

- Patent classification/suppression work; submit paper to signal processing conference.
- Further refine use of coherence features (e.g. for 3 or more microphones) for RTWD.
- Apply dynamic "attention" modeling using real-time noise source approximation for ANWSS.
- Apply detection/suppression algorithms in tandem with adaptive beamforming, filtering procedures – among other low-computation techniques.
- Apply methods developed here toward a robust noise classification system with the potential for learning audio scene-context, object localization, etc., applicable to a future, fully-realized "intelligent" audio system.

Conclusions/Future Work Directions (cont'd)

- We successfully developed a novel, robust and competitive system for wind noise detection and suppression for a multi-microphone, wearable device under challenging, low-power constraints.
- Our detection system (RTWD) uses two sets of low-dimensional features (SSC and coherence) for robust wind detection requiring no training.
- For wind noise suppression, we advance an innovative procedure (ANWSS) that "attends" to frequency regions in which wind interference is dominant, and reconstructs the corrupted signal in these bins using a context-aware, shallow NN.
- calibration

- Cross validation
- In tandem with adaptive beamforming, filtering among others

References

[1] Nelke Christoph, et al. Wind Noise Detection: Signal Processing Concepts for Speech Communication. DAGA 2016

[2] C. Nelke. "Wind noise short term power spectrum estimation using pitch adaptive inverse binary masks", in Proc. of IEEE Intern. Conference on Acoustics, Speech and Signal Processing, 2015.

[3] P. Vary and R. Martin, Digital Speech Transmission. Enchancement, Coding and Error Concealment, Wiley Verlag, 2006.

[4] E. Nemer, W. LeBlanc, et al. "Single-microphone wind noise suppression," Patent 2010/00209, 2010.

[5] Weile, Julie and Martin Andersen. "Wind Noise Management", 2016.

[6] Maj, Jean-Baptitse et al. SVD-Based Optimal Filtering Techniques for Noise Reduction in Hearing Aids Using Two Microhpones. EURASIP Journal on Applied Signal Processing 2002:4, 434-445.

[7] Docla, Smin and Marc Moonen. Robustness of SVD-Based Optimal Filtering For Noise Reduction in Multi-Microphone Speech Signals.

[8] Lu, Xugang et al. Speech Enhancement Based on Deep Denoising Autoencoder. 2013

References

[9] Nelke Christoph, et al. Dual Microphone Wind Noise Reduction by Exploiting the Complex Coherence. Speech Communication, 2014.

[10] Schmidt, Mikkel and Jan Larsen. Reduction of Non-Stationary Noise Using A Non-Negative Latent Variable Decomposition. Machine Learning for Signal Processing, 2008.

[11] Bagchi, Deblin et al. Combining Spectral Feature Mapping and Multi-Channel Model-Based Source Separation for Noise-Robust Automatic Speech Recognition. Speech and Computer: 18th annual conference, 2016.

[12] Chen, Wenjun et al. SVD-Based Technique for Interference Cancellation and Noise Reduction in NMR Measurement of Time-Dependent Magnetic Fields MDPI, 2016.

[13] Fischer, Dörte and Timo Gerkmann. Singl-Microphone Speech Enhancement Using MVDR Filtering and Wiener Post-Filtering. ICASSP, 2016.

[14] Doclo, Simon. Robustness of SVD-Based Optimal Filtering For Noise Reduction in Multi-Microphone Speech Signals. IWAENC, 1999.

[15] Sun, Meng et al. Speech Enhancement Under Low SNR Conditions via Noise Estimation Using Sparse and Low-Rank NMF with Kullback-Leibler Divergence. IEEE Transactions of Audio, Speech and Language Processing, Vol. 23, 2015.

References

[16] Gerkmann, Timo et al. Phase Processing for Single-Channel Speech Enhancement. IEEE Signal Processing Magazine, March 2015.

[17] Alexandre, Enrique. Exploring the Feasibility of a Two-Layer NN-Based Sound Classifier for Hearing Aids. EUSIPCO 2007.

[18] Ochiai, Tsubasa. Multichannel End-toend Speech Recognition. arXiv, March 2017.

[19] Heymann, Jahn et al. Neural Network Based Spectral Mask Estimation for Acoustic Beamforming. ICASSP 2016.

[20] Shao, Wei and Wei-cheng Wang. A New GSC based MVDR Beamformer with CS-LMS Algorithm for Adaptive Weights Optimization. 2011 4th International Congress on Image and Signal Processing.

[21] Xie, Junyuan et al. Image Denoising and Inpainting with Deep Neural Networks. NIPS 2012.

[22] Per, Hansen and SØren Jensen. Subspace-Based Noise Reduction for Speech Signals via Diagonal Triangular Matrix Decompositions. EURASIP 2007.

[23] Thomas, Mark R. P., Jens Ahrens and Ivan Tashev. Optimal 3D Beamforming Using Measured Microphone Directivity Patterns. IWAENC 2012.

References

[24] Schmidt, Mikkel, Jan Larsen and Fu-Tien Hsiao. Wind Noise Reduction Using Non-Negative Sparse Coding. Machine Learning for Signal Processing, 2007.

[25] Vaseghi, Saeed. Advanced Signal Processing and Noise Reduction, "Spectral Subtraction." Wily & Sons, 2000.

[26] Verteletskaya, Ekaterina and Boris Simak. Noise Reduction Based on Modified Spectral Subtraction Method. IAENG, 2011.

[27] Kumar, Anurag and Dinei Florencio. Speech Enhancement in Multiple-Noise Conditions Using Deep Neural Networks. INTERSPEECH 2016.

[28] Park, Jinuk et al. Coherence-Based Dual Microphone Wind Noise Reduction by Wiener Filtering. ICSPS 2016.

[29] Ghasemi, Jamal and M.R. Karami Mollaei. A New Approach Based on SVD for Speech Enhancement. 2011 IEEE 7th Internation Colloquium on Signal Processing and Its Applications.

[30] Vorobyov, Sergiy. Principles of Minimum Variance Robust Adaptive Beamforming Design. Signal Processing 93, 2013.

References

[31] Leng, S. and W. Ser. Adaptive Null Steering Beamformer Implementation for Flexible Broad Null Control. Signal Processing 91, 2011.

[32] Tim Roughgarden & Gregory Valiant: "The Modern Algorithmic Toolbox, Lecture #9: The Singular Value Decomposition and Low-Rank Matrix Approximations." 2015

[33] Doclo Simon et al. A Novel Iterative Signal Enhancement Algorithm for Noise Reduction in Speech. ICSLP 1998.

[34] Lilly, B.T. and K.K. Paliwal. Robust Speech Recognition Using Singular Value Decomposition Based Speech Enhancement. TENCON 1997.

[35] Xu, Yong et al. A Regression Approach to Speech Enhancement Based on Deep Neural Networks. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 23, No.1, 2015.

[36] Loizou, Philipos. Speech Enhancement: Theory and Practice. CRC Press, 2013.