

Algorithms

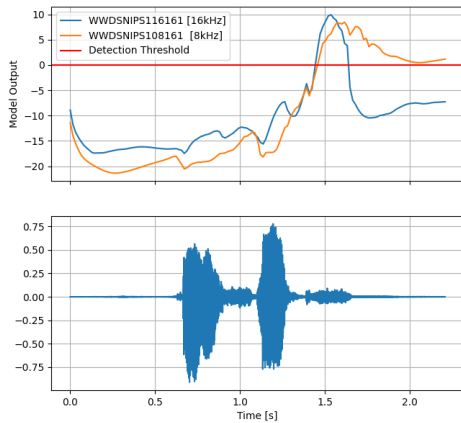
Femtosense includes the following pre-trained algorithms as part of its evaluation kit:

- Single-microphone “Hey Snips” Wake Word Detection
- Single-microphone AI Noise Reduction

Reference	
Model Name	Description
WWDSNIPS108161	Wake Word Detection “Hey Snips” 1 microphone 8 kHz sampling rate 16 ms hop-size Version 1.0
WWDSNIPS116161	Wake Word Detection “Hey Snips” 1 microphone 16 kHz sampling rate 16 ms hop-size Version 1.0
AINRGP11608161	AI Noise Reduction “General Purpose” 1 microphone 16 kHz sampling rate 8 ms hop-size 16 ms algorithmic latency Version 1.0
AINRGP11604081	AI Noise Reduction “General Purpose” 1 microphone 16 kHz sampling rate 4 ms hop-size 8 ms algorithmic latency Version 1.0

“Hey Snips” Wakeword Detection Algorithms

These wakeword detection algorithm is trained to recognize the phrase “Hey Snips.” The model input is a sequence of raw waveform frames, and its output is a sequence of probabilities of the presence of the keyword at each frame. The model has been trained against indoor environmental noise, competing speech, and room reverberance. The model uses an int16 pcm audio format.



Wakeword detection algorithms return scalar values for each input frame of audio (every 16ms). When the output exceeds the detection threshold (default 0), it is interpreted as a positive detection of the keyword.

Variants exist for 16kHz and 8kHz sampling rates.

Model Naming Convention Example

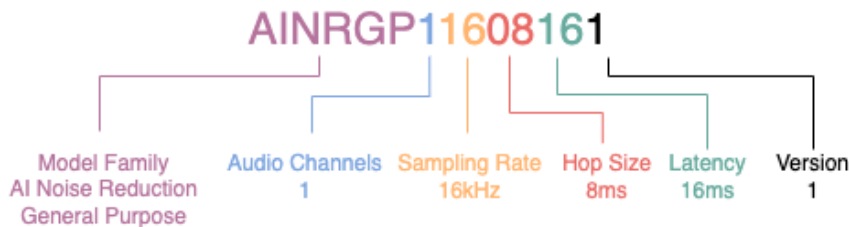


- Audio Channels: 1
- Input audio sampling rate: 8kHz
- Hop size: 16ms
- Algorithm version: 1.0

AI Noise Reduction Algorithms

The general purpose AI Noise Reduction algorithm removes background noise while preserving human speech. Speech can be in any language. The model's input is a sequence of noisy raw waveform frames, and its output is a sequence of enhanced waveform frames. The model uses an int16 pcm input/output audio format.

Model Naming Convention Example



- Audio Channels: 1
- Input/output audio sampling rate: 16kHz
- Hop size: 8ms
- Algorithmic latency: 16ms
- Algorithm version: 1.0

Proposed Evaluation

AI Noise Reduction

Our algorithm removes background noise while preserving the speech. We recommend evaluating the algorithm in the following conditions:

- Noise Environments: car noise, babble noise (restaurant/coffee background), transient sounds
- Signal to Noise Ratios: The algorithm should provide good performance above -3 dB SNR. The algorithm should work without voice distortions above 0 dB SNR.
- Distance: Play source audio at a distance from 0 to 3 meters from the microphone.

Specifications:

- Audio In:
 - 16 kHz Sampling Rate
 - Monaural
 - 16 bits (pcm)
- Audio out:
 - 16 kHz Sampling Rate
 - Monaural
 - 16 bits (pcm)

Our algorithm is not trained for a specific microphone model. For reference, the microphone we used for testing had the following specifications.

- MEMS microphone
- SNR: 67.5 dB
- AOP: 123 dB
- Sensitivity: -38 +- 3 dBFS

Testing should be conducted in a non-reverberant environment, otherwise the effective SNRs levels will be lower.

Model Performance:

Works well down to 0dB SNR without voice distortion. At -3dB, voice is still preserved but with mild distortion.

Table 1: SISDR Improvement in speech babble noise from AINRGP11608161 algorithm. Each SNR is tested with a wide variety of speech babble files from the WHAM dataset.

Input SNR (dB)	-6	-5	-4	-3	-2	-1
SISDR (dB)	10.11	9.54	9.81	10.00	9.74	9.55

Input SNR (dB)	0	1	2	3	4	5
SISDR (dB)	8.74	8.90	8.63	8.24	8.32	7.75

Wake Word Detection

Our algorithm detects `Hey Snips` in a large variety of environments. We recommend evaluating the algorithm in the following conditions:

- Noise Environments: music noise, competing speech
- Signal to Noise Ratios: 3dB SNR or higher
- Distance: Play source audio at a distance from 0 to 2 meters from the microphone.

The model was trained on a small dataset of speakers with American accents. Performance may degrade for speakers with other accents. To aid in the evaluation, we provide a small set of validation audio files for testing purposes. These audio samples were not used during the training of the model.

Specifications:

- Audio In:
 - 8 kHz Sampling Rate
 - Monaural
 - 16 bits (pcm)
- Output: scalar probability of keyword

Testing should be conducted in a non-reverberant environment when mixing with noise, otherwise the effective SNRs levels will be lower.