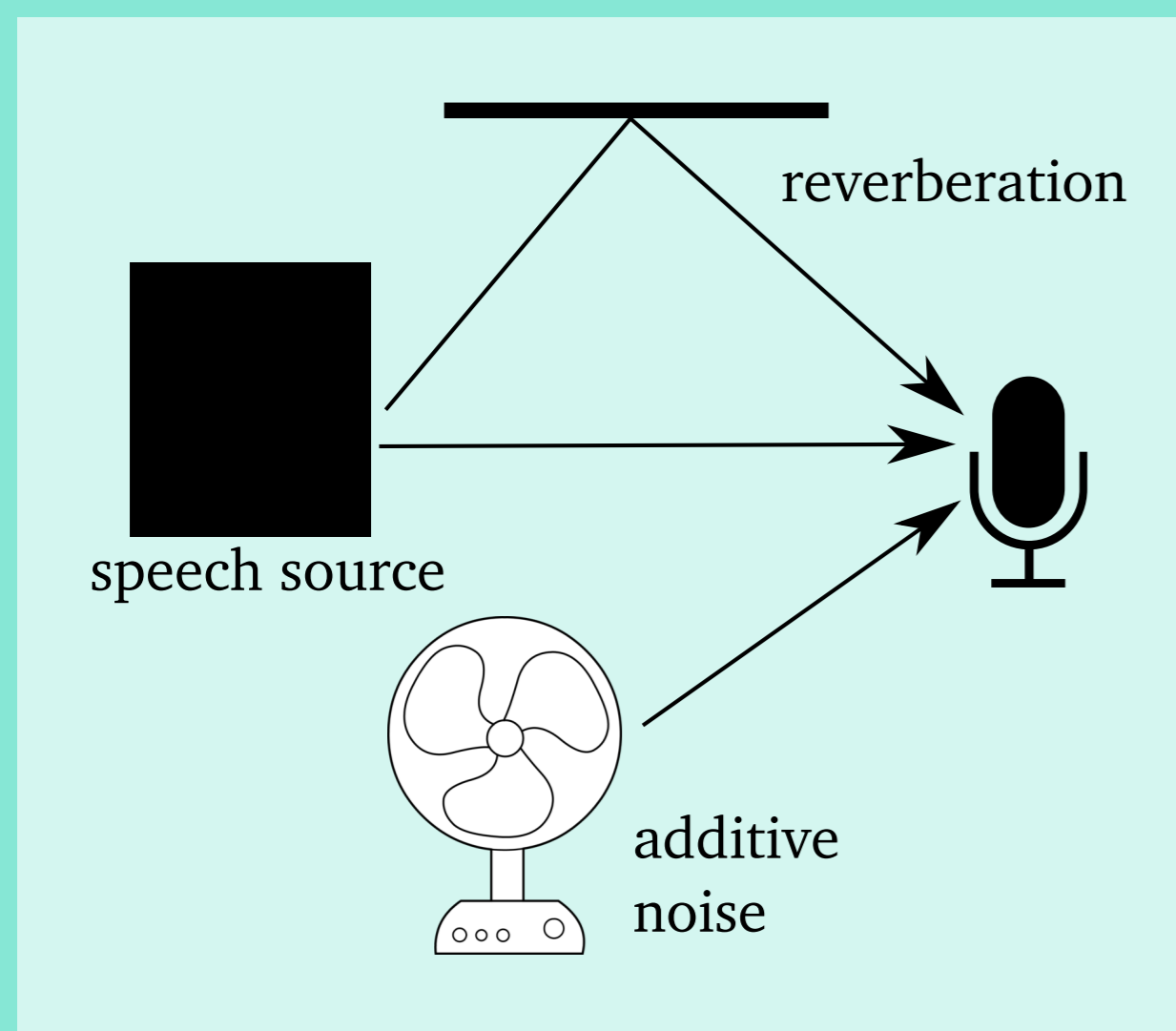


## Problem

Using distant microphones introduces lots of distortions which degrade the accuracy of the speech recognition system. This degradation is caused mainly by **additive noise** and **reverberation**.



## Solution

Common way to reduce the problem is the usage of microphone arrays instead of a single microphone. This enables to spatially select the signal of interest.

To combine the signals from multiple microphones we used **Delay-and-sum** and **Minimum variance distortionless response beamforming**.

To dereverberate the signal we applied **Weighted prediction error** method.

## Results

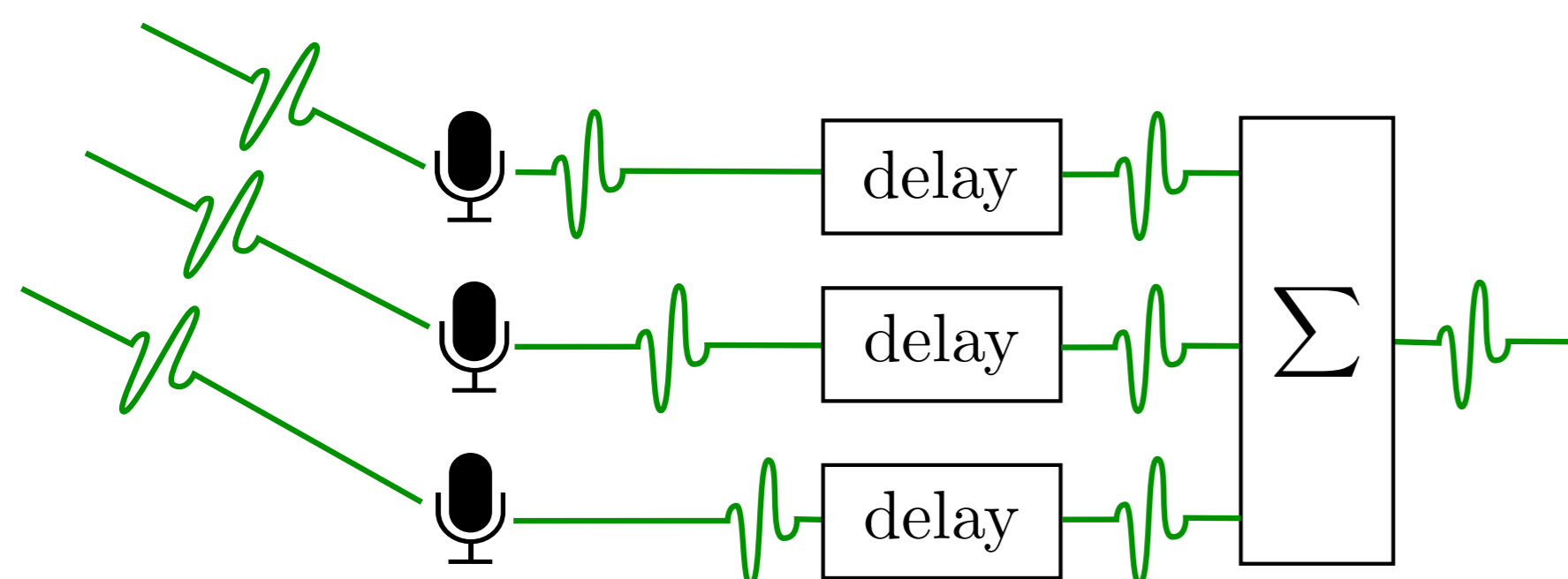
We evaluated the techniques on 3 speech recognition tasks - **CHiME3**, **AMI**, **REVERB**.

	CHiME3	
	simu	real
1-best channel	19.85	25.19
DS	15.99	18.15
MVDR	15.69	
DS + WPE	15.46	17.81
MVDR + WPE	<b>15.37</b>	<b>17.32</b>

	AMI
1-best channel	63.3
DS	58.35
MVDR	57.93
DS + WPE	58.39
MVDR + WPE	<b>57.89</b>

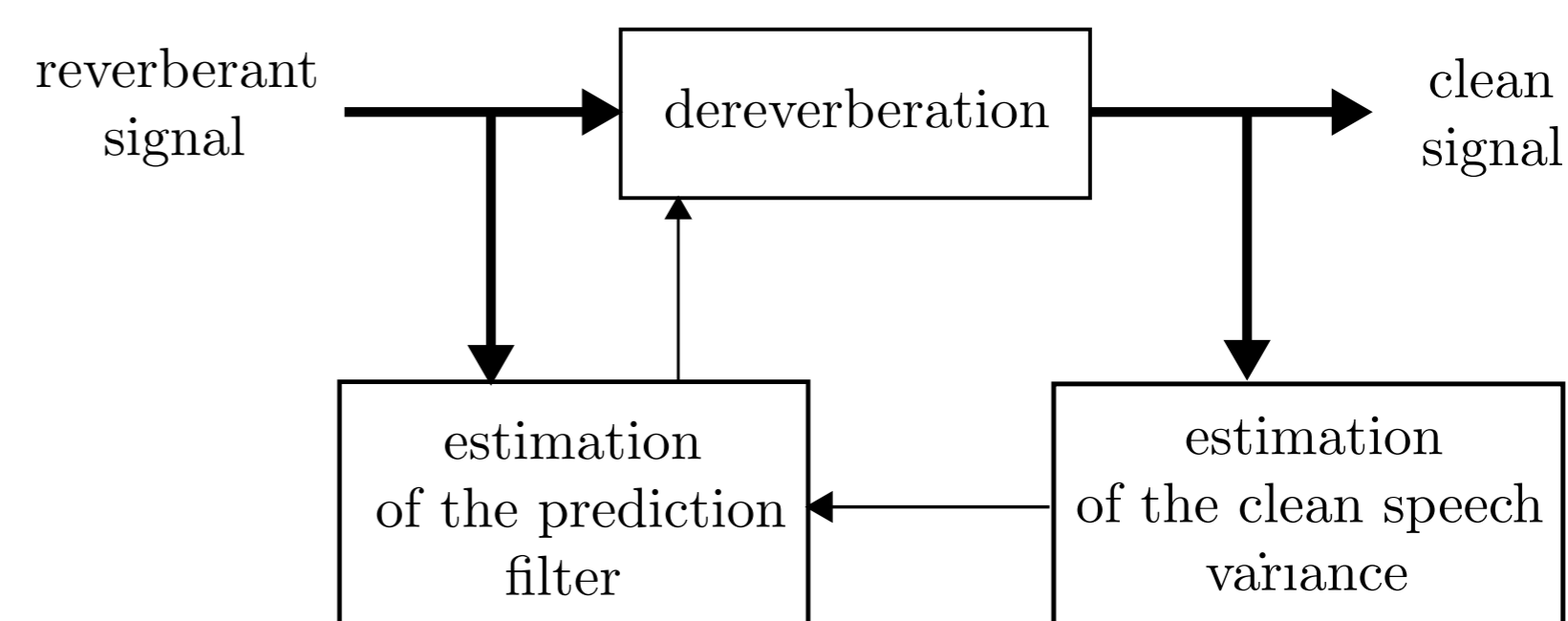
	REVERB	
	simu	real
1-best channel	12.01	30.07
DS	8.06	23.30
DS + WPE	<b>5.65</b>	<b>15.48</b>

## Delay-and-sum



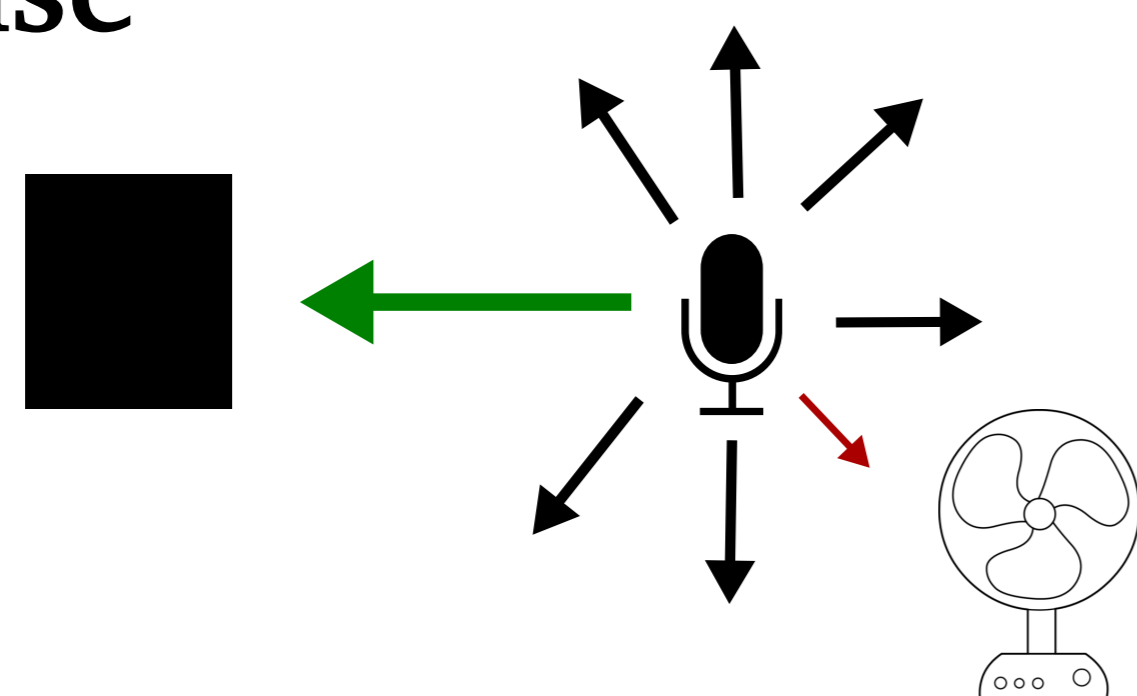
- simple and straight-forward method
- aligns the desired signal in all channels and averages shifted signals
- uses time delay estimation techniques based on cross correlation

## Weighted prediction error



- can effectively make use of multichannel signals
- predicts the reverberant component using the previous observations
- based on linear prediction but assumes a model of the target speech signal with time-varying variance

## Minimum variance distortionless response



- considers not only the position of the desired source but also the positions of the interfering sounds
- explicitly aims to minimize the effect of noise
- uses an estimate of noise covariance matrix

## Beamforming modifications

- the most notable improvements applied on top of the basic algorithms to achieve the aforementioned results

	CHiME3	
	simu	real
no modifications	19.81	23.15
weighting channels	17.94	21.64
skipping unreliable delays	18.63	21.35
fractional delays	18.10	22.50
PHAT weighting	18.13	22.27
all together	<b>15.99</b>	<b>18.15</b>