Speech Enhancement

Low resource and real-time

- Task: Build a real-time low-resource SE model (<50M Flops, <2MB) for 44.1kHz speech.
- Main efforts:
 - Data augmentations according to bad cases.
 - High reverberation, low frequency losses, bubble and keyboard noise, etc.
 - Input feature selections
 - MFCC, Energy feature, Delta feature and Mel spec.
 - Final model architectures.
 - Two layers of GRUs and linear layers.

Speech Enhancement

With weakly labelled data

- Previous approaches:
 - Need noisy-clean paired data.
 - $f(s+n) \mapsto s$
- Proposed approaches:
 - Utilize AudioSet data for enhancement.

•
$$f(s_1 + s_2 \mid c) \mapsto s_1$$

- Audioset:
 - Weakly labeled with 527 sound classes.
 - · Clip level labeled. No onset-offset information.
 - Each clip may contains multiple labels.

