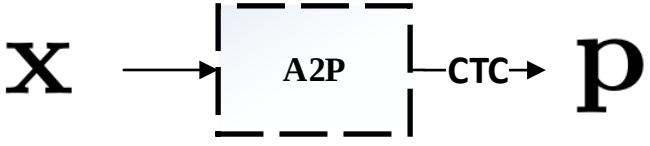
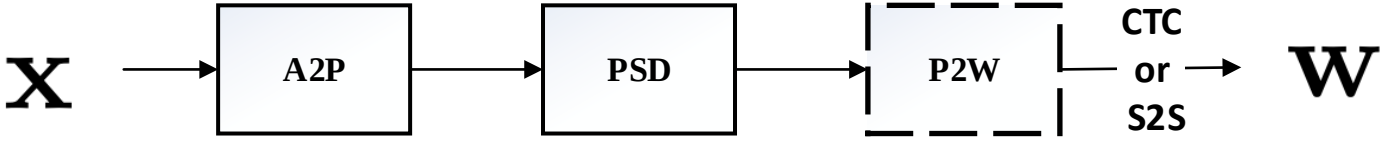


(a) Acoustic-to-phoneme Module

(b) Phoneme-to-word Module



(c) PSD-based Joint Training