



$M$		Melody Feature	$M_q$		Quantized Melody	$U_q$		Quantized Unified
$A$		Acoustic Feature	$A_q$		Quantized Acoustic		STAGE 1: Melody Codec Training	
$M_p$		Projected Feature	$\widehat{M}_q$		Quantized Predicted		STAGE 2: Joint Codec Training	