1. 对输入的音频进行人声与背景音乐分离