

基於深度學習之聽視覺語音增強技術

本院覽號

公告日期

32T-1070207

摘要

語音增強的目的地為消除背景噪音，還原目標語者的語音信號。在本技術中，我們結合聽覺訊號(即目標語者的語音)與視覺訊號(即目標語者的唇型變化)，利用深度學習的技術，開發出一基於深度學習的聽視覺語音增強系統。

技術優勢

許多語音增強技術僅專注於處理語音信號。藉由加入影像資訊，我們提出的基於深度學習的語音增強系統，在主觀與客觀的噪音消除表現上，都明顯地勝過只處理語音的語音增強系統，不論他們是傳統方法或是基於最先進的架構。這樣的差距在背景噪音為非穩定噪音(即說話聲)時尤為明顯，處理非穩定噪音正是傳統語音增強方法所不擅長的。簡言之，我們提出的聽視覺語音增強系統，比起只處理語音的模型，更能有效消除背景雜訊。

應用範圍

有鑑於影像與多媒體資訊越來越容易取得，並且能提升語音品質而加強後端語音辨識系統之辨識能力，此聽視覺語音增強技術在未來有可能被廣泛地應用於目前所有基於語音界面的智慧裝置，例如車內語音助理，智慧家庭語音助理(即Amazon Alexa, Google Home)等等。

創作人

曹昱、王新民、侯人誠、王緒翔



中央研究院
ACADEMIA SINICA