جامعة الموصل

كلية علوم الحاسوب والرياضيات

فصل مصادر الكلام باستخدام الشبكة العصبية التكرارية العميقة المتكيفة

رهام جاسم عيسى حمدان

رسالة ماجستير

علوم الحاسوب

بإشراف

د. يسرى فيصل الأرحيم

أستاذ مساعد

1442 هـ

2020 م

الملخـص

تعد مشكلة تداخل الاشارات الكلامية مع بعضها خلال عملية التسجيل باستخدام لاقط صوتي واحد والتي تسمى بمشكلة حفل الكوكتيل (Cocktail party)، من اكثر المسائل البحثية صعوبة في مجال معالجة الاشارات الصوتية، حيث تنتج هذه المشكلة خليطا من الاصوات المختلفة هذه الاشارة تسمى بالإشارة المدمجة. يتطلب حل هذه المشكلة استرجاع الاشارات الصوتية المكونة للإشارة المدمجة الى مصادرها الاصلية حيث يطلق على هذه المهمة بعملية فصل المصادر الصوتية، حيث اقتصرت افضل النتائج التي وصل اليها الباحثون على فصل اشارتين كلاميتين بنتائج مقبولة، ولكن بتعقيد حسابي كبير، كما تتطلب ان تكون الاشارة مدمجة لحظياً. تم في هذا البحث، استخدام الذكاء الاصطناعي لبناء نموذج لفصل الاشارات الصوتية المدمجة احادية القناة مكونة من اشارتين صوتيتين مختلفتين (ذكر– انثى) بواسطة الشبكات العصبية العميقة الخاضعة للإشراف -Supervised Deep Recurrent Neural Network) Bidirectional Long Short Term Memory (DRNN-BLSTM)) تعمل على تقدير نوعين من اقنعة التردد الزمني وهي ((Ideal Ratio Mask (IRM) و Optimal ratio) Mask (ORM)) والتي تستخدم لفصل المصادر من الاشارات المدمجة لحظياً.تم تدريب الشبكة باستخدام (500) اشارة مدمجة لحظيا. كما تم تقييم ومقارنة كفاءة نظام الفصل من خلال تطبيق ثلاث انواع من المقاييس وهي نسبة الاشارة الى التشوه Signal to Distortion) Ratio (SDR)) ونسبة الاشارة الى التداخل ((Signal to Interference Ratio(SIR) و نسبة الاشارة الى الزيف ((Signal to Artifact Ratio (SAR) للإشارات الناتجة من عملية الفصل لكل مرشح بشكل منفصل. وقد اظهرت النتائج تحسنا في كفاءة اداء النموذج في عملية الفصل باستخدام قناع (ORM) مقارنة عن استخدام المرشح (IRM) . حيث حققت عملية

القصل باستخدام قناع (IRM) نسبة (SDR=6.16 db) و (SIR=7.189db) و (SAR=14.020db) ، بينما حقق القناع (ORM) نسبة (SDR=6.347db) و (SIR=7.387db) و (SAR=14.150db) تم برمجة النموذج لغة (بايثون Python-3.8.3-amd64).

University of Mosul
College of Computer Sciences
and Mathematics

Separating Speech Sources Using Adapted Deep Recurrent Neural Network

Riham Jassim Issa Hamdan

M.Sc./Thesis

Computer Sciences

Supervised by

Dr.Yusra Faisal Al-Irhayim

Assistant Professor

2020 A.D **1442 A.H**

Abstract

The matter of overlapping set of speech signals recorded from a single microphone, which is called (cocktail party problem) , is one of the most challenge in the field of audio signal processing .To solved this problem , we should restore the source signals that make up the Mixture signal, this task is called audio source separation process. and the results are still limited to separate two instantaneously mixed speakers with acceptable results and high computational complexity. In this research, and deep learning model was designed to separate two instantaneously mixed speakers.(male-female) Supervised Deep Recurrent Neural Network-Bidirectional Long Short Term Memory (DRNN-BLSTM) that evaluates two time frequency masks: Ideal Ratio Mask (IRM) and Optimal Ratio Mask (ORM) , which was used to separate the required audio sources from the instantaneous signal. The efficiency of the separation system was also evaluated and compared through the application of three types of measures: Signal to Distortion Ratio (SDR), Signal to Interference Ratio (SIR) and Signal to Artifact Ratio (SAR) for signals. Resulting from the separation process for each Mask separately. The results showed The results show that the (optimal ratio mask) outperforms other training target (Ideal Ratio Mask). The Model with IRM training Target achieve SDR = 6.16dB, SIR = 7.189db and SAR = 14.020db ,while ORM achieved SDR = 6.347db and SIR = 7.387db.) And (SAR = 14.150 dB). The model is programmed By (Python-3.8.3-amd64).