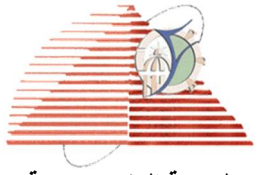

كليـــة الهندســـــة
قسم الاتصالات والكترونيات الهندسية

تحسين كفاءة ودقة أنظمة التعرف علي الأصوات من خلال إستخدام وسائل متعدده مرئيه ومسموعه للإشاره الصوتية

مقدمة من

م/ إسلام عيد علي محمد المغربي

رسالة مقدمة إلي كلية الهندسة - جامعة الفيوم
كجزء من متطلبات الحصول علي درجة الدكتوراه

في
هندسة الإتصالات و الإلكترونيات
قسم الهندسة الكهربية- كلية الهندسة
جامعة الفيوم

كلية الهندسة، جامعة الفيوم
الفيوم- جمهورية مصر العربية
٢٠٢٠

كليـــة الهندســـــة
قسم الاتصالات والكترونيات الهندسية

جـــامعة الفـــيوم
Fayoum University

تحسين كفاءة ودقة أنظمة التعرف علي الأصوات من خلال إستخدام وسائل متعدده مرئيه ومسموعه للإشاره الصوتية

مقدمة من

م/ إسلام عيد علي محمد المغربي

رسالة مقدمة إلي كلية الهندسة - جامعة الفيوم
كجزء من متطلبات الحصول علي درجة الدكتوراه

في
هندسة الإتصالات و الإلكترونيات
قسم الهندسة الكهربية- كلية الهندسة
جامعة الفيوم

يعتمد من لجنة الممتحنين:

أ.د. عمرو محمد رفعت	(المشرف الرئيسي)
أ.د. محمد هشام فاروق	مشرفا
أ.د. محمد فتحى ابو اليزيد	كلية الهندسة – جامعة القاهره
أ.د/ رانيا احمد ابوالسعود	كلية الهندسة – جامعة الفيوم

كلية الهندسة - جامعة الفيوم
الفيوم - جمهورية مصر العربية
٢٠٢٠

كليـــة الهندســـــة
قسم الاتصالات والكترونيات الهندسية

جـــامعة الفـــيوم
Fayoum University

تحسين كفاءة ودقة أنظمة التعرف علي الأصوات من خلال إستخدام وسائل متعدده مرئيه ومسموعه للإشاره الصوتية

مقدمة من

م/ إسلام عيد علي محمد المغربي

رسالة مقدمة إلي كلية الهندسة - جامعة الفيوم
كجزء من متطلبات الحصول علي درجة الدكتوراه

في
هندسة الإتصالات و الإلكترونيات
قسم الهندسة الكهربية- كلية الهندسة
جامعة الفيوم

تحت اشراف

أ.د. عمرو محمد رفعت

أستاذ الاشارات الرقميه في قسم الهندسة الكهربية تخصص هندسه الالكترونيات و الاتصالات الكهربيه
كلية الهندسة
جامعة الفيوم

أ.د. محمد هشام فاروق

أستاذ قسم الرياضيات والفزيقا الهندسية
كلية الهندسة
جامعة القاهره

كلية الهندسة - جامعة الفيوم
الفيوم - جمهورية مصر العربية
٢٠٢٠

ملخص الرسالة

أنظمة التعرف علي الأصوات تعتبر مجال بحثي متميز في مجال تحليل الاشارات يجذب الكثير من الباحثين لهذا المجال. ويستخدم في الكثير من المجالات مثل التفاعل الآدمي مع الحاسبات HCI و الانسان الآلي.

بالإضافة إلي كفاءة تكنولوجيا التعرف علي الأصوات لكنها مازالت تحتاج إلي الكثير من العمل لتحسين النتائج التي تم الوصول إليها. وبالرغم من الجهود المبذولة خلال العقود الماضية للوصول الي اعلي درجات التعرف علي الأصوات فمازالت تلك الانظمة غير دقيقه وغير مناسبة للتطبيقات الحياتيه الحقيقية خاصا تلك التي توجد في أوساط بها الكثير من الضوضاء . أيضاً لان الكلام لم يكن ينتج منفصلاً عن بعض الحركات الايمائية المصاحبة للحديث مثل حركة الشفاه و حركة العين ولذلك إستخدام المعلومات المصاحبة للحديث مثل حركة الشفاه قد تساعد في رفع كفاءة انظمة التعرف علي الأصوات مقارنة بالنتائج التي يتم الحصول عليها من خلال إستخدام الإشارة الصوتية فقط . إضافة الاشارة المرئية للصوت وحركة الشفاه لا يتأثر بالضوضاء المصاحبة للصوت.

في أنظمة التعرف علي الكلام من خلال إستخدام الصوت والصورة للشخص المتحدث AVSR يتم تسجيل كلا من الحركة والصوت للشخص المتكلم ويتم فصل الصورة عن الصوت بإستخدام أنظمة مختلفة. يتم الحصول علي الخصائص التي تمكن النظام من التعرف علي الصوت من خلال الإشارة الصوتية والمرئية ويتم في النهاية دمج المعلومات للوصول إلي اعلي درجة من الكفاءة والدقة للتعرف علي الصوت.

نحاول من خلال هذا البحث أن نجد طرق مختلفة لتحسين عملية التعرف علي الصوت من خلال إضافة المعلومات التي يتم الحصول عليها من المعلومات المرئية للشخص المتحدث ومن خلال إختيار أفضل الطرق للوصول إلي استخراج الخصائص الصوتيه والمرئية للكلام وطرق دمج المعلومات في النهاية.

يقوم هذا البحث بتصميم نظام للتعرف علي الاصوات معتمداً علي الاشارة الصوتية بالاضافة الي الاشارة المرئية المصاحبة للصوت المأخوذه من حركة الشفاه للمتكلم. الدراسات السابقة اثبتت ان إضافة حركة الشفاه إلي الإشاره الصوتيه من الممكن أن يؤدي إلي زيادة دقة التعرف علي الاصوات خاصة في حالة وجود ضوضاء مؤثرة علي الإشارة الصوتية.. من خلال النظام المقترح في هذا البحث يتم إستخراج خصائص الإشارة الصوتيه بإستخدام خاصية MFCC وإستخدام خاصية DCT لإستخراج الخصائص من صورة حركة الشفاه المصاحبة للصوت. . يتم دمج الخصائص المستخرجه من الصوت والصوره المصاحبه له لتدريب نظام التعرف علي الاصوات باستخدام واحدة من اهم انواع Deep Learning ألا وهي BiLSTM التي تتميز بكفاءتها في تصنيف الإشاره الصوتيه لأنها تأخذ في إعتبارها جميع خصائص الصوت . هذا البحث يجري مقارنه بين النتائج التي تم التوصل إليها من خلال إستخدام BiLSTM للتعرف علي الصوت بإستخدام الخصائص التي تم إستخراجها من الصوت والصوره معاً والنتائج التي تم الحصول عليها بإستخدام الطريقة المستخدمة من قبل لتصنيف الاصوات وهيHMM عن طريق استخدام اداة HTK. تم اختبار كفاءة النظام المقترح من خلال تطبيق باستخدام اثنان من اشهر قواعد البيانات للصوت والصورة معا وهي قواعد البيانات AVletter و GRID. من خلال تحليل النتائج للنظام المقترح المعتمد علي الصوت والصوره معا وإستخدام BLSTM في مرحلة التصنيف نجد انه يقدم كفاءة اعلي ومعدل تعرف اكبر بمقدار٩.٢٨% عن استخدام الصوت فقط.