Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science

# Impaired Children Speech Recognition of Isolated Arabic Words Using Deep Learning

**A thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment for the Requirements
of the Degree of Doctor of Philosophy
in
Computer Science**

**By**

**Sura Ramzi Shareef Ahmed**

**Supervised by**

**Asst. Prof. Dr. Yusra Faisal Mohammed**

---

**2023 A.D.** **1444 A.H.**

# ABSTRACT

Many children have speech impairment due to hyperkinetic dysphonia, autism spectrum disorder (ASD), or some other conditions, which results in noisy speech due to incorrect pronunciation and various disorders. That makes it hard for children to benefit from voice-controlled technology, and they struggle to communicate with other people. ASR adopts techniques to recognize spoken words via translating the acoustic wave into textual symbols equivalent to what is said. These techniques provide human-machine communication that helps people in different domains via innovative voice-based applications. However, such techniques face a challenging task in recognizing the speech of impaired children. Where their ability for speech recognition becomes limited due to the disability or mispronunciation of the children.

The dissertation aims to cover this research gap through proposed a framework that can recognize Arabic words from the speech of impaired children and support them to use voice-controlled applications and talk to others easily. The dissertation's methodology investigates two trends. First, the adoption of different ASR approaches. These approaches include a suggested novel encoder structure based on collaborative training for a sequence-predictive model as an acoustic encoder able to encode the speech signal into a sequence of hidden representations to handle the temporal patterns and signal variation within the sequence of pronunciation frames that are disordered at the level of a speaking child and the spoken word itself. In addition to End-to-End, hybrid, and traditional ASR approaches that are adopted by this dissertation in order to enhancement of the recognition models for a speech of impaired children. While the second trend, the dissertation investigated the strength of the popular feature extraction techniques in ASR, specifically MFCC, PNCC, and GFCC, to correctly recognize words within speech impairments.

The dissertation's framework was suggested as the first contribution to recognizing impaired children's speech at the level of Arabic speech. The performance experiments are accomplished using an Arabic speech of words for children with speech impairments, which was named Impaired Children's Speech Files (ICSF). It is a unique dataset constructed from the speech of impaired children "students" from schools with special needs in Iraq. The different approaches adopted by this dissertation showed differentiated efficiency in the recognition rate. In general, the results of the approach based on the proposed framework are promising. They produced results that compete with models known within ASR. The best accurate results were achieved, adopt 96% for HMM, and 92% for Naïve Basina, respectively.

وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

# التعرف على كلام الاطفال ضعيفي النطق للكلمات العربية المنفصلة باستخدام التعلم العميق

اطروحة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة الدكتوراة فلسفة في
علوم الحاسوب

من قبل

سرى رمزي شريف احمد

بإشراف

أ. م. د. يسرى فيصل محمد

1444 هـ

2023 م

# الخلاصة

يعاني العديد من الأطفال من ضعف النطق بسبب خلل النطق الناتج عن خلل النطق الحركي ، أو اضطراب طيفية ، أو بعض الحالات الأخرى ، مما ينتج عنها الكلام غير واضح بسبب النطق غير الصحيح والاضطرابات المختلفة. وهذا يجعل من الصعب على الأطفال الاستفادة من تقنية التحكم الصوتي ، ويعانون من اجل التواصل مع الآخرين. تتبنى ASR تقنيات للتعرف على الكلمات المنطوقة من خلال تحويل الموجة الصوتية إلى رموز نصية مكافئة لها. توفر هذه التقنيات اتصالًا بين الإنسان والآلة يساعد الأشخاص في مجالات مختلفة عبر التطبيقات المبتكرة القائمة على الصوت. ومع ذلك ، فإن مثل هذه التقنيات تواجه مهمة صعبة في التعرف على كلام الأطفال الذين يعانون من صعوبة في النطق. حيث تصبح قدرة هذه التطبيقات على التعرف على الكلام محدودة بسبب الإعاقة أو النطق الخاطئ للأطفال.

تهدف الأطروحة إلى تغطية هذه الفجوة البحثية من خلال اقتراح إطار عمل يمكنه التعرف على الكلمات العربية من خطاب الأطفال المعاقين ودعمهم في استخدام التطبيقات التي يتم التحكم فيها بالصوت والتحدث مع الآخرين بسهولة. تحاول منهجية الأطروحة التحقيق في اتجاهين. الاول ، اعتماد مناهج مختلفة من التعرف على الكلام التلقائي. تتضمن هذه الأساليب بنية مشفر جديدة مقترحة تعتمد على التدريب التعاوني لنموذج تنبؤي بالتسلسل كمشفّر صوتي قادر على تشفير إشارة الكلام في سلسلة من التمثيلات المخفية للتعامل مع الأنماط الزمنية وتغير الإشارة ضمن تسلسل إطارات النطق التي هي مضطربة على مستوى نطق الطفل والكلمة المنطوقة نفسها. بالإضافة إلى مناهج التعرف على الكلام الشاملة والهجينة والتقليدية التي تتبناها هذه الرسالة من أجل تعزيز نماذج التعرف على خطاب الأطفال الذين يعانون من صعوبة في النطق. بينما الاتجاه الثاني ، بحثت الأطروحة في قوة الميزات المستخرجة بواسطة تقنيات استخراج الميزات الشائعة في التعرف على

الكلام ، وتحديداً MFCC ، و PNCC ، و GFCC ، للتعرف بشكل صحيح على الكلمات في حالات ضعف الكلام.

يعد إطار العمل المقترح في هذه الأطروحة حسب علمنا كأول مساهمة في التعرف على كلام الأطفال المعاقين على مستوى الكلام العربي. يتم إجراء تجارب الأداء باستخدام تسجيلات كلام للأطفال يعانون من إعاقات في النطق لمجموهة من الكلمات العربية ، والذي أطلق عليه اسم ملفات نطق الأطفال المعاقين (ICSF). وهي مجموعة بيانات فريدة تم إنشاؤها من حديث الأطفال "الطلاب" من مدارس ذوي الاحتياجات الخاصة في العراق. أظهرت الأساليب المختلفة التي اعتمدتها هذه الرسالة كفاءة متباينة في معدل التعرف. بشكل عام ، نتائج النهج القائم على الإطار المقترح واعدة. حيث حققت نتائج تنافس النماذج المعروفة داخل ASR. تم تحقيق أفضل دقة نتائج 96٪ لـ HMM و 92٪ لـ Naïve Basina على التوالي.