King Saud University
College of Computer and Information Sciences
Department of Computer Science

Deep Learning-based Pronunciation Error Detection for Non-native Learners of the Arabic Language

By
Mohammed Mahdi Ahmed Algabri

Supervisor
Prof. Hassan I. Mathkour

Co-Supervisor
Prof. Mansour M. Alsulaiman

Submitted in Partial Fulfilment of the Requirements
for the Degree of Doctor of Philosophy in the Department of Computer Science at the College of Computer and Information Sciences.

Riyadh, Kingdom of Saudi Arabia

January, 2022

Thesis Approval

Student name: Mohammed Mahdi Ahmed Algabri.

Thesis title: Deep Learning-based Pronunciation Error Detection for Non-native Learners of the Arabic Language.

Defence and approval date: Wednesday, 23/6/1443 H, corresponding to 26/1/2022 G.

Dissertation Committee Members:

Supervisor: Prof. Hassan Mathkour.

Signature: mathkour

Co-Supervisor: Prof. Mansour Alsulaiman.

Signature:

Examiner: Prof. Hatim A. AboAlsamh.

Signature:

Examiner: Prof. Abdulmalik S. AlSalman.

Signature:

External Examiner: Prof. Mohamed Osman Hegazi.

Signature: Hegazi

King Saud University

2022

ملخص الرسالة

حصلت أنظمة التدريب على النطق بمساعدة الحاسوب (CAPT) على إهتمام كبير في السنوات الاخيرة، حيث تم تطوير العديد من أنظمة CAPT لمتعلمي العديد من اللغات من الناطقين بغيرها. وبالرغم من أن اللغة العربية واحدة من اكثر اللغات انتشاراً في العالم، إذ تُعد الخامسة عالمياً من حيث عدد المتحدثين، إلا أنه يلاحظ أن هناك قليل من الاهتمام بتطوير الأنظمة الحاسوبية لاكتشاف أخطاء النطق لمتعلمي اللغة العربية من الناطقين بغيرها. وحيث أن المملكة العربية السعودية تولي خدمة اللغة العربية أهمية كبيرة، فإن نظام CAPT لتعليم اللغة العربية للناطقين بغيرها يعد امرا مهماً للمملكة. علاوة على ذلك، سيمكن نظام CAPT المملكة من تلبية احتياج عدد كبير من المسلمين حول العالم إلى تعلم اللغة العربية ليتمكنوا من قراءة القرآن الكريم.

تعد وحدة اكتشاف الأخطاء وتصحيحها (Mispronunciation Detection and Diagnosis - MDD) مكوناً أساسياً لإنظمة CAPT، لانها تستكشف الأصوات المنطوقة بشكل خاطئ وتُقدم أنواع مختلفة من التغذية الراجعة للمتعلمين. بالمقارنة مع اللغات الأخرى، تطوير نظام MDD للغة العربية يحتاج إلى مزيد من الدراسة والبحث، وذلك لندرة الأبحاث في نظام MDD للغة العربية بشكل عام، وندرة استخدام تقنيات التعلم العميق في بناء أنظمة MDD لتعليم اللغة العربية للناطقين بغيرها بشكل خاص، يُضاف الى ذلك عدم وجود قاعدة بيانات غنية ومُرمزة لمتحدثي اللغة العربية من الناطقين بغيرها. لذلك نهدف في هذه الأطروحة إلى دراسة وتطبيق أحدث تقنيات التعلم العميق لبناء نظام MDD يحاكي القدرة البشرية ويكون ذو كفاءة عالية مع توليد تغذية راجعة.

عالجنا مشكلة البحث من خلال عدة مسارات، في المسار الأول قمنا بصياغة مشكلة التعرف على الكلام في نظام MDD على أنها مهمة اكتشاف الأشياء (object detection) في الصور بحيث تم تمثيل الصوت بصور طيفية (image of spectrogram) واعتبار الفونيمات كائنات في هذه الصور. في المسار الثاني، صممنا نظام لاكتشاف الخصائص الصوتية (Articulatory Features) حيث مثلنا الخصائص الصوتية على أنها كائنات متعددة التسميات (multi-label) في الصور الطيفية.

بناءً على النتائج المتميزة في المسارين السابقين، قمنا في المسار الثالث بتطوير نظام MDD لمتحدثي اللغة العربية من الناطقين بغيرها، ويتمتع النظام المقترح بالقدرة على اكتشاف الصوتيات واكتشاف الخصائص الصوتية لكل صوت في آن واحد. باكتشافنا للصوتيات وللخصائص الصوتية للصوت في آن واحد، سنتمكن من مهمة توليد التغذية الراجعة لمساعدة المتعلم على تحسين نطقة. علاوة على ذلك استخدمنا الخوارزمية الجينية للعثور على أفضل المعاملات لشبكات التعلم العميق في النظام المقترح. وقد قمنا بمقارنة أداء عدة نماذج من النظام المقترح مع احدث نُظم ال (MDD) والتى تعمل بشكل نهاية-الى-نهاية (end-to-end) وحصل نظامنا على نتائج أفضل. بالإضافة الى ذلك، اقترحنا استخدام تقنيات الدمج (fusion techniques) بين النظام المقترح ونظام نهاية-الى-نهاية وحصلنا على أداء أفضل.

لمعالجة مشكلة ندرة قواعد بيانات الكلام العربي للناطقين بغير اللغة العربية، قمنا باستخدام تقنيات نقل التعلم (transfer learning). قمنا ايضاً بتطوير قاعدة بيانات متنوعة ومُرمزة تحتوى على متحدثين من جنسيات مختلفة

وأطلقنا عليها اسم (Arabic-CAPT). أخيرا، قمنا بإستخدام تقنيات توليد الكلام من النصوص (Text to speech) لبناء قاعدة بيانات جديدة تحتوي على مجموعة جديدة من الاصوات تم توليدها تقنياً.