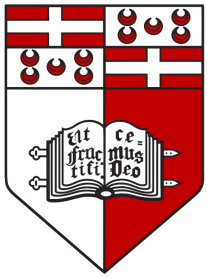
A neural network-based approach to accent conversion

Kenny W. Lino

Msc. Dissertation



Department of Intelligent Computer Systems
Faculty of Information and Communication Technology
University of Malta
2018

Supervisors:
Claudia Borg, Department of Artificial Intelligence, University of Malta
Andrea De Marco, Institute of Space Sciences and Astronomy, University of Malta
Eva Navas, Department of Communications Engineering, University of the Basque Country

Submitted in partial fulfilment of the requirements for the Degree of
European Master of Science in Human Language Science and Technology

Abstract

With the emergence of the use of technology in language learning through tools like Rosetta Stone and Duolingo, learners have slowly been given more autonomy of their language learning projection. Although these tools have allowed learners to tailor their learning to their own liking, there is a gap between the available resources to assist those that would like to improve their pronunciation. Previous research in the intersection of language learning and speech technology has made efforts to develop pronunciation training systems to address this problem, but the systems themselves tend to have gaps due to the lack of appropriate support for the users, especially in appropriately identifying errors and providing sufficient feedback to help them correct their errors.

Some researchers have purported that alongside other forms of feedback such as a visual articulatory representation, a voice conversion system could serve as a potential feedback mechanism by helping learners understand what their voice could sound like given the appropriate changes. However, like pronunciation training systems, voice conversion systems also faced many limitations especially in terms of the quality which made them unrenderable as useful tools. With that said, recent advances in speech technology using deep neural networks have become increasingly successful in achieving better accuracy and quality in a variety of tasks, allowing for the potential to return and address these said gaps in quality and performance for voice conversion.

In this thesis, I aim to investigate these advancements in applying deep neural networks to develop a voice conversion system that could potentially serve as a feedback mechanism as a part of a larger computer-based pronunciation training system. Specifically, I intend to adapt the methodologies of Aryal and Gutizerrez-Osuna (2014) to set forth an accent conversion system that strives to convert a source voice into a target accent, leveraging neural network architectures in place of Gaussian Mixture Models for conversion.