

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

**ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ**

**Μπακαγιάννης Λεωνίδα  
Μεταπτυχιακός Φοιτητής**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**Επόπτης Μεταπτυχιακής Εργασίας: Καθηγητής, Ι. Στυλιανού**

**Τετάρτη , 25 Νοεμβρίου 2020 , ώρα 11:00 π.μ.**

**Κανάλι YouTube του Τμήματος**

**[https://www.youtube.com/channel/UC7uE3QIMTQjkrpByB\\_Gnt6Q/live](https://www.youtube.com/channel/UC7uE3QIMTQjkrpByB_Gnt6Q/live)**

**“ Περί της χρήσης Νευρωνικών Δικτύων για Βελτίωση Σήματος Φωνής και της διαστολής σε ένα περιβάλλον Παραγωγικού Αντιπαραθετικού Δικτύου ”**

### **Περίληψη**

Η Βελτίωση Σήματος Φωνής είναι ένα πεδίο της Επεξεργασίας Φωνής που στοχεύει στη βελτίωση της ποιότητας θορυβωδών σημάτων και προσπαθεί να αυξήσει την καταληπτότητά τους. Αρκετοί αλγόριθμοι προτάθηκαν για βελτίωση σήματος φωνής στον 20ο αιώνα. Οι περισσότεροι από αυτούς εκμεταλλεύοντουσαν κυρίως τα φασματικά χαρακτηριστικά του θορυβώδους σήματος. Η εντατικοποίηση, ωστόσο, της χρήσης των Νευρωνικών Δικτύων (ΝΔ) τα τελευταία χρόνια, έχει οδηγήσει στην ύπαρξη αρκετών συστημάτων βασισμένων σε νευρωνικά δίκτυα που χρησιμοποιούνται για να βελτιώσουν ένα σήμα και να αφαιρέσουν το θόρυβο. Μια σχετικά πρόσφατη εξέλιξη στον τομέα των νευρωνικών δικτύων είναι τα Παραγωγικά Αντιπαραθετικά Δίκτυα (Generative Adversarial Networks-GANs), τα οποία χρησιμοποιούν δύο νευρωνικά δίκτυα, το παραγωγικό και το διαχωριστικό, σε ανταγωνισμό μεταξύ τους για να

επιτύχουν τους σκοπούς του συστήματος. Τα δύο αυτά δίκτυα παίζουν ένα παίγνιο μηδενικού αθροίσματος, στο οποίο το παραγωγικό δίκτυο προσπαθεί να παράξει δείγματα που μοιάζουν αληθινά στο διαχωριστικό δίκτυο και ο απώτερος στόχος του παραγωγικού δικτύου είναι η παραγωγή δειγμάτων που το διαχωριστικό δίκτυο δεν μπορεί να ξεχωρίσει αν προκύπτουν από το παραγωγικό δίκτυο ή την πραγματική κατανομή. Σε αυτή την εργασία, τα κύρια συστήματα βελτίωσης φωνής βασισμένα σε νευρωνικά δίκτυα παρουσιάζονται μαζί με το κατά πόσο μια τεχνική που συναντάται στα νευρωνικά δίκτυα, η διαστολή, μπορεί να χρησιμοποιηθεί για να αυξήσει την αποδοτικότητα των συστημάτων βελτίωσης ποιότητας φωνής. Συγκεκριμένα, παρουσιάζεται μια συγκριτική μελέτη της αρχιτεκτονικής τριών συστημάτων βελτίωσης σήματος φωνής (SE-WaveNet, SEGAN, SE-FFNet) και η συγκριτική τους αξιολόγηση με υποκειμενικές (μέση βαθμολογία γνώμης) και αντικειμενικές (PESQ,STOI,CSIG,CBAK,CVAL,SSNR) μετρικές. Επίσης, παρουσιάζονται τα πειράματα που έγιναν σχετικά με την εφαρμογή της διαστολής σε ένα περιβάλλον Παραγωγικού Αντιπαραθετικού Δικτύου με σκοπό τη μείωση του αριθμού των παραμέτρων σε ένα Παραγωγικό Αντιπαραθετικό Δίκτυο βελτίωσης ποιότητας φωνής .

**University of Crete**

**Computer Science Department**

**M.Sc. Thesis presentation / examination**

**Bakayannis Leonidas**

**Master's Thesis Supervisor: Professor, I. Stylianou**

**Wednesday, 25 November 2020, 11:00 a.m**

**YouTube channel :**

**[https://www.youtube.com/channel/UC7uE3QiMTQjkrpByB\\_Gnt6Q/live](https://www.youtube.com/channel/UC7uE3QiMTQjkrpByB_Gnt6Q/live)**

# **“On the use of neural networks and dilation for speech enhancement in a Generative Adversarial Network environment”**

## **Abstract**

Speech Enhancement(SE) is a Speech Processing field which aims to improve speech quality of noisy signals in an attempt to increase their intelligibility and,as a consequence, to reduce the amount of effort that someone has to make in order to listen to them. Several algorithms have been proposed for speech enhancement throughout the 20th century. Most of them mainly took advantage of the spectral characteristics of the noisy signal. But with the use of Neural Networks (NNs) escalating over the recent years, there have been several neural-network-based systems that are used to enhance a signal and remove noise. A relatively recent class of machine learning frameworks based on Neural Networks are Generative Adversarial Networks(GANs) which use two separate neural networks,the Generator and the Discriminator, that compete with each other in order to achieve the system’s goals. These two networks play a minimax zero-sum game, where the Generator tries to produce samples that seem real to the Discriminator with the ultimate goal of the generator being the production of samples that the discriminator cannot distinguish whether they erupt from the generator or from the real distribution. In this thesis, a study of the main neural-network-based systems for speech enhancement is presented alongside a study on how a neural network concept, dilation, can be used to boost speech enhancement performance. Specifically, a comparative evaluation of the architectures of three Speech Enhancement system (SE-WaveNet, SEGAN, SE-FFNet) is presented as well their comparative evaluation based on objective (PESQ,STOI,CSIG,CBAK,CVAL,SSNR) and subjective (Mean Opinion Score) metrics. Additionally, the experiments regarding the application of dilation in a Generative Adversarial Network environment in an effort to reduce the number of parameters required for a Speech Enhancement Generative Adversarial Network is presented.